The Canadian Journal of Statistics
Vol. 30, No. ?, 2002, Pages ???-???
La revue canadienne de statistique

Bayesian estimation of cognitive decline in patients with Alzheimer's disease

Patrick BÉLISLE, Lawrence JOSEPH, David B. WOLFSON and Xiaojie ZHOU

Key words and phrases: Alzheimer's disease; Bayes factors; Biphasic regression; Change-point; Gibbs sampler; Mini-Mental State Exam.
MSC 2000: Primary 62P10; secondary 62F15.

Abstract: Recently, there has been great interest in estimating the decline in cognitive ability in patients with Alzheimer's disease. Measuring decline is not straightforward, since one must consider the choice of scale to measure cognitive ability, possible floor and ceiling effects, between-patient variability, and the unobserved age-of-onset. The authors demonstrate how to account for the above features by modeling decline in scores on the Mini-Mental State Exam in two different data sets. To this end, they use hierarchical Bayesian models with change-points, for which posterior distributions are calculated using the Gibbs sampler. They make comparisons between several such models using both prior and posterior Bayes factors, and compare the results from the models suggested by these two model selection criteria.

Estimation bayésienne du déclin cognitif de patients atteints de la maladie d'Alzheimer

Résumé : On s'est beaucoup intéressé ces derniers temps à l'estimation du déclin des fonctions cognitives des personnes atteintes de la maladie d'Alzheimer. Il n'est pas facile de quantifier ce déclin, qui dépend de l'échelle utilisée pour mesurer les fonctions cognitives, mais aussi de la variabilité entre les individus, de l'incertitude entourant le moment exact du début de leur maladie et d'éventuels effets plancher et plafond. Les auteurs montrent comment il est possible de tenir compte de ces différents éléments en modélisant le déclin observé dans les résultats obtenus par deux groupes de patients au mini-examen de l'état mental. Ils utilisent pour ce faire des modèles bayésiens hiérarchiques avec points de jonction, pour lesquels ils calculent les lois a posteriori au moyen de l'échantillonneur de Gibbs. Ils comparent plusieurs modèles de ce type au moyen de facteurs de Bayes a priori et a posteriori ; ils comparent ensuite les résultats des modèles suggérés par ces deux critères de sélection.

1. INTRODUCTION

Alzheimer's Disease (AD) is a senile dementia that is characterized by progressive cognitive impairment, culminating in death (CSHA 1994). Much of the epidemiologic literature has focused on the pattern of cognitive deterioration in patients with AD. One of the most commonly used scales of cognitive impairment is the Mini-Mental State Scale (Folstein, Folstein & McHugh 1975).

Assessment of the pattern of cognitive decline in AD patients is important for several reasons. First, the design and analysis of clinical trials of treatments for AD depend on correctly modeling how a cohort of patients can be expected to decline. Models of decline can also be used to examine the role played by covariates such as age-of-onset, gender, and years of education in influencing the rapidity of decline (Katzman, Brown, Thal, Fuld, Aronson, Butters, Klauber, Wiederholt, Pay & Xiong 1988; Teri, McCurry, Edland, Kukull & Larson 1995). Finally, by using a regression model,

it would be possible to predict the disease course of Alzheimer's patients, thereby anticipating future care that may be needed.

Here, a hierarchical modeling approach will be presented to describe longitudinal decline in Mini-Mental State Exam (MMSE) scores based on data collected from two separate groups of patients. The maximum score is 30 and the minimum score is 0. Scores below 24 are taken to be indicative of some type of cognitive impairment. We introduce a change-point regression model to account for commonly observed patterns of decline. Five random-effect regression models of varying complexity are compared using both prior and posterior Bayes factors. Included amongst these is a linear random-effect model, a special case of what is known as mixed models. A non-linear (biphasic) change-point regression model is assessed to be superior to a linear mixed model.

The preponderance of the literature on modeling decline in AD has focussed on fixed-effect models (Katzman *et al.* 1988; Teri *et al.* 1995, Yesavage, Poulsen, Sheikh & Tanke 1988; Mortimer, Ebbitt, Jun & Finch 1992; Haxby, Raffaele, Gillette, Schapiro & Rapoport 1992), including fixed-effect growth curves (Liu, Tsai & Stern 1996) and a tri-phasic regression model proposed by Brooks, Kraemer, Tanke & Yesavage (1993). More recently, the well recognized between patient variability has led to the consideration of the linear mixed models of Laird & Ware (1982) by researchers in the field of AD (Maltby Broe, Creasy, Jorm, Christensen & Brooks 1994; Jacqmin-Gadda, Fabrigoule, Commenges & Dartigues 1997). Their use stems, to some extent, from the availability of software packages such as "PROC MIXED" from SAS, that easily handle such models. As will be seen below, however, these models have serious drawbacks, as they fail to take into account simultaneously four important features: both inter- and intra-patient variability, and plateaus in the decline rates arising from insensitivity of the MMSE and other scales to cognitive changes late in the disease course. In particular, standard mixed effects models, which can be used when the point at which the slope changes is a fixed constant, are not applicable here. That is, we introduce the point of change as a random effect and thereby are led to a mixture model which is non-linear. In fact, Liu *et al.* (1996) state in their concluding remarks that it would be very difficult to allow their regression and curve parameters to vary from subject to subject.

We assume, realistically, that conditional on the subject specific parameters, within-patient scores are independent, being separated by relatively long periods of time. There seems to be little advantage in using a smooth curve in place of a simple linear biphasic regression. An important use for models of decline is in the comparison of the rates of decline between different groups of subjects with AD. For parametrically defined smooth curves, the rate of decline, defined as the derivative, will vary with time. A reporting problem then arises as to which time point to choose in order to best summarize the rate of decline, especially as each subject changes at a different rate. Moreover, it has been well documented in the AD literature that decline in MMSE scores is nearly linear (Mortimer *et al.* 1992; Haxby *et al.* 1992; Brooks *et al.* 1993; Ashford, Shan, Butler, Rajasekar & Schmidt 1995), apart from the plateaus.

The outline of the paper is as follows. The two data sets of MMSE scores collected longitudinally on different groups of patients are introduced in Section 2. In Section 3, we show that by using a hierarchical Bayesian regression model with a change in slope, all of the important features common to longitudinal AD data as described above can be elegantly accommodated. The model we use here is similar to that proposed by Lange, Carlin & Gelfand (1992) in the context of following CD4 counts over time in patients with HIV. Carlin, Gelfand & Smith (1992) also present a variety of examples where Bayesian hierarchical change-point models can be used. The various alternative models are fitted and compared using the two data sets of MMSE scores in Section 4. In Section 4.3, we show how our model can be extended in a straightforward fashion to allow the rates of decline to depend on one or more covariates, without compromising the random-effect component of the model. In Section 5 we conclude with a discussion.

2. DESCRIPTION OF THE DATA SETS

Two independently collected data sets from different parts of the United States were analysed in this study. These will be referred to as the Minneapolis data and Palo Alto data, respectively. Although these data were not collected with a comparative study in mind, as will be seen in Section 4, the parameter estimates from each data set seem to be similar.

2.1 The Minneapolis Data Set

The Minneapolis data set consists of subjects who were recruited by Mortimer *et al.* (1992) from the Minneapolis VA Medical Center and from the community, to participate in a prospective longitudinal study of primary degenerative dementia. Ninety three subjects were initially screened, and after the imposition of various inclusion/exclusion criteria, sixty five subjects with probable or definite AD remained for the analysis.

Following acceptance into the study, subjects were assessed at 6-month intervals until death or loss due to other reasons. No subject contributed more than 8 data points; that is, a baseline measurement and three and one half years of follow-up. Data collected consisted of the approximate age-of-onset, age at entry into the study, gender and years of education. Age-of-onset was provided by the family and as such is probably inaccurate, because of the vague initial symptoms of AD. We discuss this point further in Section 3.4.

2.2 The Palo Alto Data Set

Subjects for the Palo Alto data set were identified and recruited from the Stanford Alzheimer's Disease Diagnostic and Treatment Center to participate in a longitudinal study of dementia (Brooks *et al.* 1993). Patients with less than three MMSE data points were excluded. Data on 55 subjects, again assessed at 6-month intervals, were available after various inclusion/exclusion criteria were applied. The maximum follow-up time was 7.5 years.

Refer to Table 1 for a summary of the main features of these two data sets. Figure 4 includes the data from eight typical subjects, four from each data set. Time = 0 on these plots refers to the time each patient entered the study.

TABLE 1: Demographic characteristics for study subjects. * indicates six subjects with missing information. † indicates four subjects with missing information. ‡ indicates one subject with missing information. Family history data were not available for the Palo Alto data.

	Palo Alto	Minneapolis
Number of subjects	55	65
Mean (sd) age at entry into the study	64.1 (9.0)*	67.6 (8.7)
Mean (sd) age of AD onset	60.3 (9.7)*	63.7 (8.5)†
Number (%) of males	32 (65%)	47 (72%)
Family history of dementia	NA	26 (41%)‡
Mean (sd) follow-up years	5.1 (2.7)	2.3 (0.85)
Mean (sd) number of visits	6.7 (3.1)	5.5 (1.7)

3. THE MODEL

3.1 Notation and Likelihood Function

In general, one may expect that three distinct periods could be identified in MMSE trajectories for dementia patients from disease onset to the final stages of disease: an initial period of stability, followed by a period of relatively steady decline ending most often with a MMSE score below 10, and a final stable period, where either decline truly stops, or, more likely, the MMSE becomes insensitive to further decline. Because of limited data early on in the disease, modeling here will take into account the latter plateau and decline periods only.

Let N be the total number of subjects in each data set, so that $N = 65$ and 55 for the Minneapolis and Palo Alto data sets, respectively. Let M be the largest possible number of MMSE tests across all individuals in the study, where $M = 8$ (3.5 years) for Minneapolis, while $M = 16$ (7.5 years) for Palo Alto.

Let m_i be the total number of evaluations (decline and plateau periods combined) for subject i, and let $n_i(\tau_i) \leq m_i$ be the total number of MMSE scores for each subject during the decline period. While m_i is known, n_i is a function of the latent change-point, and must be estimated. Let τ_i be the unobserved change-point representing the index at which the decline period ends and the plateau period begins for each subject i. This latent change-point is permitted to occur after the period of observation for subject i has ended, up to a maximum of $2M$ time periods. The extension of τ_i to times past m_i avoids problems of truncation in estimating the time from disease onset to the end of decline. The vast majority of patients should have entered their final plateau phases by time period $2M$.

Let θ_i be the slope (rate of decline) during the decline period. In particular, θ_i is the expected decline in MMSE score per six-month period (the interval between MMSE tests), for subject i. Thus the annual expected decline for subject i is $2\theta_i$.

Let h_i be the mean level at the final plateau period for subject i. Let Y_{ij} be the MMSE score for subject i in the jth time period, $j = 1, \ldots, M$, measured from entry into the study. The value Y_{ij} is, therefore, the score of the test administered $6 \times (j-1)$ months after entry into the study. This value is considered missing if the test was not administered during a given period for a particular subject. Therefore, also denote

$$\delta_{ij} = \begin{cases} 1 & \text{if } Y_{ij} \text{ is not missing, and} \\ 0 & \text{if } Y_{ij} \text{ is missing} \end{cases}$$

for $i = 1, 2, \ldots, N$, and $j = 1, 2, \ldots, M$. Let Y_i be the vector of MMSE scores for subject i, $Y_i = (Y_{i1}, \ldots, Y_{iM})$. Note that $n_i = \sum_{j \leq \tau_i} \delta_{ij}$. Finally, let σ^2_{1i} and σ^2_{2i} denote the variance of the MMSE measurements during the decline and plateau periods, respectively, for subject i.

With this notation, conditional on τ_i, θ_i, h_i, σ^2_{1i}, and σ^2_{2i}, we assumed $Y_{ij} \sim N(h_i + \theta_i(j - \tau_i), \sigma^2_{1i})$ for $j \leq \tau_i$, and $Y_{ij} \sim N(h_i, \sigma^2_{2i})$ for $j > \tau_i$, for all subjects $i = 1, \ldots, N$. The conditional likelihood for patient i given $(\tau_i, \theta_i, h_i, \sigma^2_{1i}, \sigma^2_{2i})$ under Model 1 is therefore

$$\begin{aligned} f(Y_i \,|\, \tau_i, \theta_i, h_i, \sigma^2_{1i}, \sigma^2_{2i}) = & \\ & \prod_{j \leq \tau_i} \left(\frac{1}{\sqrt{2\pi}\sigma_{1i}} \right) e^{-\frac{1}{2\sigma^2_{1i}} \{Y_{ij} - h_i - \theta_i(j-\tau_i)\}^2} \prod_{j > \tau_i} \left(\frac{1}{\sqrt{2\pi}\sigma_{2i}} \right) e^{-\frac{1}{2\sigma^2_{2i}} (Y_{ij} - h_i)^2} \\ \propto & \left(\frac{1}{\sigma_{1i}} \right)^{n_i(\tau_i)} \exp\left[-\frac{1}{2\sigma^2_{1i}} \sum_{j=1}^{\tau_i} \delta_{ij} \left\{ Y_{ij} - h_i - \theta_i(j - \tau_i) \right\}^2 \right] \times \\ & \left(\frac{1}{\sigma_{2i}} \right)^{m_i(\tau_i) - n_i(\tau_i)} \exp\left\{ -\frac{1}{2\sigma^2_{2i}} \sum_{j=\tau_i+1}^{M} \delta_{ij} \left(Y_{ij} - h_i \right)^2 \right\}, \end{aligned} \tag{1}$$

where the second term in the last line of equation (1) is set to 1 if $\tau_i \geq M$, and $n_i(\tau_i)$ and $m_i(\tau_i) - n_i(\tau_i)$ are the numbers of observations before and after τ_i for patient i, respectively. The full likelihood is obtained by taking the product of these individual subject contributions over all subjects.

Conditional on the subject specific parameters, therefore, we assume independence for both between- and within-patient data. While it is still possible to have correlated errors within patients in a random-effect model, accurate estimation of the correlation structure would be difficult owing to the small number of observations in each of the decline and plateau phases from each subject. Furthermore, it is difficult to distinguish between non-linearities such as those induced by change-point models and correlated errors within a linear structure. This is because departures from linearity of the type seen in biphasic regression can, over short sequences of data, be incorrectly interpreted as linearity with positively correlated errors. We therefore assume a parsimonious model with independent errors within subjects.

The model we described above, which we label as Model 1, is the most general model we investigated, but various simplifications are also of interest. A preliminary examination of the data via scatter plots indicated that many subjects may not yet have reached their plateau; that is, they were still in their period of decline when the study ended. This suggested that a simple linear mixed model without change-points may fit the data almost as well as the biphasic model, but with many fewer parameters. The likelihood function here is similar to that described above, but with the change-point fixed at $\tau = 2M$, that is, no change. It is also possible that a model with a common variance for all subjects within the decline and/or plateaus periods would fit the data almost as well, but with considerably fewer parameters, that is, setting $\sigma_{i1}^2 = \sigma_1^2$ and/or $\sigma_{i2}^2 = \sigma_2^2$. Therefore, five different models were fit to each of the two data sets. The main features of these models, henceforth referred to as Model 1 through Model 5 in order of decreasing numbers of parameters, are summarized in Table 2.

TABLE 2: Summary of the features of the five models fitted and compared in this study. N/A indicates "non-applicable."

Model	Model type	Individual variances for decline period	Individual variances for plateau period
1	biphasic	yes	yes
2	biphasic	yes	no
3	biphasic	no	no
4	linear	yes	N/A
5	linear	no	N/A

The above likelihood specification assumes missingness at random for the missing data. Even patients with very low MMSE scores (below 10) tended to come in at regular intervals, and there was a roughly constant proportion of missing data items across times. It is unlikely that missing values "between" other observed values would introduce bias, as subsequent observed scores would continue along the same straight line.

Let α_i represent the true age of onset for patient i. Estimated times of onset, a_i, are typically provided by a family member, but are subject to error. We assumed that $a_i = \alpha_i + \epsilon_i$, where ϵ_i are independent and identically distributed $N(0, \eta^2)$ variables. Combining the information from our main model for decline described above with this "mini-model" for the times of AD onset, we are able to estimate the time from disease onset to the end of the decline period.

3.2 Prior Distributions

The overall strategy was to create diffuse but still informative prior distributions, in the sense that parameter values for the prior distributions were selected to cover a somewhat wider interval than

the available prior information might suggest. In this way, potential differences between this and past studies are accommodated.

We assumed the variates for the slope, the timing of the end of the decline period, and MMSE level of the post-decline plateau period to be *a priori* independent, both within and between subjects. While these parameters are related, our use of diffuse prior distributions, and the relation between these parameters built into the likelihood function, means that it is not important that this correlation be included in the prior distribution. Specifically, a uniform prior distribution was assumed for each τ_i, so that $\mathrm{R}(\tau_i = k) = 1/(2M)$, $k = 1, \ldots, 2M$. For the prior distribution on h_i, it was assumed that $\mathrm{P}(h_i = 0) = \cdots = \mathrm{P}(h_i = 8) = 0.1026$, $\mathrm{P}(h_i = 9) = 0.0773$, $\mathrm{P}(h_i = 10) = 0.0519$, $\mathrm{P}(h_i = 11) = 0.0265$, and $\mathrm{P}(h_i = 12) = \cdots = \mathrm{P}(h_i = 30) = 0.0012$. These probabilities were derived by assuming a "piecewise linear" prior on these discrete points, with $\mathrm{P}(h_i \leq 10) = 0.95$.

A normal hierarchical model was assumed for the prior distribution for the slopes, where the θ_i were independently and identically distributed as

$$\theta_i | \mu, \phi^2 \sim N(\mu, \phi^2), \quad i = 1, \ldots, N.$$

The normal assumption was visually checked via a histogram of slope estimates across individuals in a preliminary analysis, which appeared close to normal. To estimate the slopes, we first graphically inspected the scatter plots of individual patient MMSE data over time, and removed the points which were clearly from the plateau phase. We next fit a simple (non-hierarchical) regression model without a change-point to estimate the slopes for each patient.

We used the normal/chi-square family of conjugate prior distributions for μ and ϕ^2, so that *a priori*,

$$\phi^2 \sim \quad s_0 \chi^{-2}_{\nu_0} \quad \text{and} \quad \mu | \phi^2 \sim \quad N\left(\mu_0, \frac{\phi^2}{\ell_0}\right).$$

Similarly, the prior distributions for the variances of the observations in both the decline and plateau periods, for each patient, were assumed to be inverse chi-square, so that

$$\sigma^2_{1i} \sim s_{01} \chi^{-2}_{\nu_{01}} \quad \text{and} \quad \sigma^2_{2i} \sim s_{02} \chi^{-2}_{\nu_{02}}.$$

Previous literature (Katzman *et al.* 1988; Yesavage *et al.* 1988) suggests that the mean annual change is in the range from -2 to -7. In keeping with our prior distribution selection strategy, we desired the average annual rates of change to be in the slightly wider range from about 0 to -8. We chose $\mathrm{E}(\mu) = \mu_0 = -2$ so that the prior mean average annual rate of change would be -4. Since we would like the standard deviation about this mean to be approximately 2, and since the inverse chi-square distribution has mean $1/(\nu - 2)$ and variance $2/\{(\nu-2)^2(\nu-4)\}$, we set $\mathrm{E}(\phi^2) = 4$ and $\mathrm{var}(\phi^2) = 25$. We chose $\ell_0 = 3$, $\nu_0 = 5.28$ and $s_0 = 13.12$, so that 95% of the prior probability for the mean of the hierarchical distribution for the slopes was in the range $(-3.9, -0.08)$, calculated from the marginal $t_{\nu_0 = 5.28}$ distribution for μ. This corresponds to $(-7.8, -0.16)$ on an annual basis. Similarly, we selected $s_{01} = 8$, $\nu_{01} = 6$, $s_{02} = 2.5$ and $\nu_{01} = 4.5$, so that $\mathrm{E}(\sigma^2_{1i}) = 2$, $\mathrm{var}(\sigma^2_{1i}) = 4$, $\mathrm{E}(\sigma^2_{2i}) = 1$ and $\mathrm{var}(\sigma^2_{2i}) = 4$. This was based on our belief that the variability of the MMSE scores would be higher during the decline phase than in the final plateau phase. The prior 95% probability intervals for σ^2_{1i} and σ^2_{2i} were $(0.35, 4.9)$ and $(0.12, 2.7)$, respectively.

For the age-of-onset, we assumed *a priori* that $(\alpha_1, \ldots, \alpha_N)$ and η^2 are independent with respective marginal densities

$$f(\alpha_1, \ldots, \alpha_N) = \begin{cases} 1/(r_i - c_1), & \text{for } c_1 \leq \alpha_i \leq r_i \\ 0, & \text{otherwise} \end{cases}$$

and

$$f\left(\eta^2\right) = \begin{cases} 1/c_2, & \text{for } \eta^2 \in (0, c_2] \\ 0, & \text{otherwise.} \end{cases}$$

The lower limit for the true age at disease onset was chosen from the literature to be $c_1 = 40$ years, and we right-truncated the age of onset for each subject to be less than their age at entry into the study, denoted by r_i. Of course, we also have $a_i \leq r_i$ across all subjects. The prior upper bound for the standard deviation of onset time in the population was set to be $\sqrt{c_2} = 4$ years. Therefore, the *a priori* range of the observed age of onset around the true (latent) value is $(-8, +8)$. Since virtually all caregivers should report an observed onset time within 8 years of the true value, a uniform prior distribution for η^2 was preferred to the usual conjugate but unbounded non-informative prior distribution for variances. Since the date at time 0 (first MMSE measurement) for each patient is known, the α_i's are easily converted to onset times, to be used later for the estimation of the latent duration from onset to entry into the plateau phase.

The prior distributions we selected are somewhat arbitrary, in the sense that other AD researchers may have different appreciations of the past literature. Nevertheless, our prior distributions were expected to have little influence on the posterior distributions of the parameters, since the information in the data is much larger than that contained in the prior distribution. Nevertheless, a limited robustness study was carried out using sampling importance resampling (SIR) as discussed by Rubin (1987) and Smith & Roberts (1993). For example, doubling the mean of the prior distribution for mean slope (setting $\mu_0 = -4$ in Equation 2) had only a negligible effect on the posterior distributions of all parameters, as did reversing the direction of the slope (setting $\mu_0 = +2$ rather than $\mu_0 = -2$). With such extreme changes in prior parameter values for the slope having no impact on posterior densities, we concluded that our prior distributions were sufficiently diffuse so that an exhaustive robustness analysis was not necessary.

3.3 Implementation

The biphasic regression model described above is an example of a multi-path (panel data) change-point problem. The Gibbs sampler is ideally suited for estimation in the change-point model used here, since considering the latent data τ_i, $i = 1, \ldots, N$, to be an additional set of unknown parameters allows standard results on conjugate distributions to be used throughout, considerably simplifying the analysis. In order to use the Gibbs sampler, the "full conditional distribution" for each unknown parameter must be specified. These are given in the Appendix.

In order to estimate the required number of iterations (cycles) to carry out in the Gibbs sampler, we used the method suggested by Raftery & Lewis (1992). Using the Gibbsit program available from these authors, and specifying that the 95% posterior credible intervals should have actual posterior coverage between 0.925 and 0.975 with high probability, we decided to run 5000 iterations with a burn-in of 100 iterations. This choice was verified by following the procedure suggested by Raftery and Lewis. We also ran the sampler several times for each model by using different starting values, with similar results each time (Gelman & Rubin 1992 statistic near 1). The Gibbs sampler was implemented using Splus (version 3.3) functions written specifically for these data sets. Bayes factors were calculated using a different algorithm programmed in BUGS (version 0.6), and in fact, BUGS could have been used for the entire Gibbs sampler as well. The calculation of the Bayes factor was less stable than sampling from the posterior density, so that up to 100,000 iterations were required for accurate estimation.

3.4 Bayes Factors

Competing models, M_i and M_j, $i, j = 1, \ldots, 5$, $i \neq j$, were compared at data x by computing the Bayes factor (Kass & Raftery 1995),

$$B_{ij} = \frac{p(x|M_i)}{p(x|M_j)} = \frac{p(M_i|x)/p(M_j|x)}{p(M_i)/p(M_j)} \ ,$$

where $p(x|M_j) = \int p_j(x|\gamma_j)p_j(\gamma_j)d\gamma_j$, and γ_j represents the vector of all unknown parameters in Model M_j, $j = 1, \ldots, 5$. We assumed *a priori* that $p(M_i) = p(M_j)$. Since no improper prior

distributions were used, approximate Bayes factors were calculated from the output of the Gibbs sampler, following the method of Chib (1995), Section 2.1.3. We called this the prior Bayes factor, to distinguish it from the posterior Bayes factor (Aitkin 1991). Posterior Bayes factors are defined as above, except that now $p(x|M_j) = \int p_j(x|\gamma_j)p_j(\gamma_j|x)d\gamma_j$, so that the likelihood function of the data is mixed over the posterior distribution of the unknown parameters γ_j, rather than the prior distribution of γ_j. The posterior Bayes factor is computationally simpler to estimate than the prior Bayes factor, since one may more directly use the output of the Gibbs sampler. In particular, using the notation of Section 3, we calculated

$$p(Y|M_j) = \prod_{i=1}^{N} \sum_{z=1}^{T} p(Y_i|M_j, \gamma_{iz})/T \ ,$$

where T is the total number of iterations of the Gibbs sampler (after convergence), and γ_{iz} is the vector of parameter values for $(\tau_i, \theta_i, h_i, \sigma_{1i}^2, \sigma_{2i}^2)$ at the zth iteration for the ith patient. The densities $p(Y_i|M_j, \gamma_{iz})$ were calculated from the likelihood functions discussed in Section 3.

There has been considerable discussion in the literature about the relative merits of prior versus posterior Bayes factors. Aitkin (1991), Kass (1993), and Conigliani & O'Hagan (2000) all point out that the prior Bayes factor is more sensitive to the choice of prior distribution compared to the posterior Bayes factor. Kass (1993) provides illustrations of this fact, and also mentions that hypothesis testing through Bayes factors is more sensitive to the prior inputs than posterior estimation of parameters. In other words, it is often the case that parameter estimation remains robust across a range of prior specifications, while the Bayes factor substantially changes across these same prior choices. The posterior Bayes factor is robust to changes in the prior distribution except in small sample sizes, and also avoids the Lindley paradox (Aitkin 1991). On the other hand, posterior Bayes factors have been criticized for making double use of the data, and for not being a strictly Bayesian approach (see the discussions by Lindley and Cuzick in Aitkin 1991). While a full discussion of these issues is beyond the scope of this article, we calculate and compare results from both the prior and posterior Bayes factors here. See Key, Pericchi & Smith (1999), Bayarri & Berger (1999), and De Santis & Spezzaferri (1997) for other alternatives to prior and posterior Bayes factors in model selection.

4. RESULTS

The results from the model selection procedure described in Section 3.6 are first discussed, followed by detailed descriptions of the inferences from the best fitting model. In all cases, posterior means for a given parameter are calculated as the mean of the samples across iterations of the Gibbs sampler. Similarly, 95% credible intervals are calculated from the 2.5% and 97.5% quantiles of the Gibbs iterates.

4.1 Model Selection

Table 3 lists the prior and posterior Bayes factors for both data sets for all possible model comparisons. In both data sets, the best supported model using posterior Bayes factors was Model 2 (biphasic with individual variances during the decline period but a common variance across all subjects during the plateau period). For the data set from Palo Alto, support for this model was substantial, but in the Minneapolis data set, a close competitor was the biphasic model with common variances in both periods. The Bayes factor was 2.34. Converting the odds given by the Bayes factor to a probability gives $2.34/(2.34 + 1) = 0.7$, so that the posterior probability that Model 2 is superior to Model 3 is 0.70. For the Palo Alto data set, Model 2 had a Bayes factor of approximately 10^{27} compared to Model 1, indicating virtual certainty of the superiority of Model 2. While this and other Bayes F factors in Table 3 are large, these are not unexpected given the large

total number of data points. This superiority of Model 2 over Model 1 indicates that there is little evidence of individual variability post-decline period, so that a single variance parameter suffices here.

TABLE 3: Prior and Posterior Bayes factors for all model comparisons in both the Palo Alto and Minneapolis data sets. Row and column numbers refer to the model numbers from Table 2. The entry in row i, column j refers to the Bayes factor for Model i compared to Model j. For example, the $i = 1$ and $j = 2$ entry in the top table, $10^{29.7}$, represents the Bayes factor for Model 1 compared to Model 2, with Model 1 superior to Model 2.

Prior Bayes factors
Data set from Palo Alto

	1	2	3	4	5	Model rank
1	1	$10^{29.7}$	$10^{40.8}$	$10^{119.2}$	$10^{96.5}$	1
2	$10^{-29.7}$	1	$10^{11.1}$	$10^{89.4}$	$10^{66.7}$	2
3	$10^{40.8}$	$10^{-11.1}$	1	$10^{78.4}$	$10^{55.6}$	3
4	$10^{-119.2}$	$10^{-89.4}$	$10^{-78.4}$	1	$10^{-22.7}$	5
5	$10^{-96.5}$	$10^{-66.7}$	$10^{-55.6}$	$10^{22.7}$	1	4

Data set from Minneapolis

	1	2	3	4	5	Model rank
1	1	$10^{59.8}$	$10^{103.3}$	$10^{191.6}$	$10^{179.6}$	1
2	$10^{-59.8}$	1	$10^{43.6}$	$10^{131.9}$	$10^{119.9}$	2
3	$10^{-103.3}$	$10^{-43.6}$	1	$10^{88.3}$	$10^{76.3}$	3
4	$10^{-191.6}$	$10^{-131.9}$	$10^{-88.3}$	1	$10^{-12.0}$	5
5	$10^{-179.6}$	$10^{-119.9}$	$10^{-76.3}$	$10^{12.0}$	1	4

Posterior Bayes factors
Data set from Palo Alto

	1	2	3	4	5	Model rank
1	1	$10^{-15.8}$	$10^{11.3}$	$10^{78.1}$	$10^{51.6}$	2
2	$10^{15.8}$	1	$10^{27.1}$	$10^{93.8}$	$10^{67.3}$	1
3	$10^{-11.3}$	$10^{-27.1}$	1	$10^{66.8}$	$10^{40.3}$	3
4	$10^{-78.1}$	$10^{-93.8}$	$10^{-66.8}$	1	$10^{-26.5}$	5
5	$10^{-51.6}$	$10^{-67.3}$	$10^{-40.3}$	$10^{26.5}$	1	4

Data set from Minneapolis

	1	2	3	4	5	Model rank
1	1	$10^{-25.4}$	$10^{-25.0}$	$10^{16.6}$	$10^{0.3}$	3
2	$10^{25.4}$	1	$10^{0.37}$	$10^{42.0}$	$10^{25.7}$	1
3	$10^{25.0}$	$10^{-0.4}$	1	$10^{41.7}$	$10^{25.3}$	2
4	$10^{-16.6}$	$10^{-42.0}$	$10^{-41.7}$	1	$10^{-16.3}$	5
5	$10^{-0.3}$	$10^{-25.7}$	$10^{-25.3}$	$10^{16.3}$	1	4

The prior Bayes factor largely agrees with the posterior Bayes factor, in that the relative ordering of Models 3, 4, and 5 are similar, and both criteria agree that a change-point model is superior to a linear mixed model. In both data sets, however, the prior Bayes factor prefers Model 1, with individual variances in both the decline and plateau phases, over Model 2 which has a single variance parameter for the plateau phase. As discussed in Section 3.4, both the prior and posterior

Bayes factors calculate the average likelihood with respect to a distribution over the unknown parameters, the only difference being whether this distribution is the prior or the posterior. In this case it seems that averaging over the wider prior distribution leads to higher average likelihoods in the model with individual variances in the plateau phase. However, once more accurate estimates of the variance parameters are available, the posterior Bayes factor suggests that individual variance parameters are not needed.

While the debate over which Bayes factor to use (if any) continues, detailed discussion of this issue is beyond the scope of the present article. For both of our data sets, the posterior distributions of the main parameters were very similar for Model 1 and Model 2. Therefore, the decision as to whether to base our model choice on prior or posterior Bayes factors is of little practical importance, as far as main parameter estimation is concerned. This agrees with the results of Kass (1993). We have somewhat arbitrarily chosen to present below the results from Model 2, which was supported by the posterior Bayes factor for both data sets.

4.2 Main Results

Figure 1 displays the prior distribution for the mean slope, μ, and the posterior densities for μ from the Palo Alto and Minneapolis data sets. The posterior distributions displayed here are obtained from "Rao-Blackwell" estimates (see Gelfand & Smith 1990) of Equation (2), using the output of the Gibbs sampler. The prior mean for μ was -2.0, and the posterior mean of -1.90 from the Palo Alto data set agreed closely with this value. The posterior mean slope calculated from the Minneapolis data set was slightly steeper, at -2.4. While this could be indicative of a small difference in the decline rates in the two populations, it could also have arisen by chance, or as an artifact of the longer follow-up times in the Palo Alto data set. A slight decrease in mean decline rate could occur if, occasionally, subjects with observations in the plateau phase are estimated not to have had a change-point, and therefore have for that iteration a less negative slope estimate.

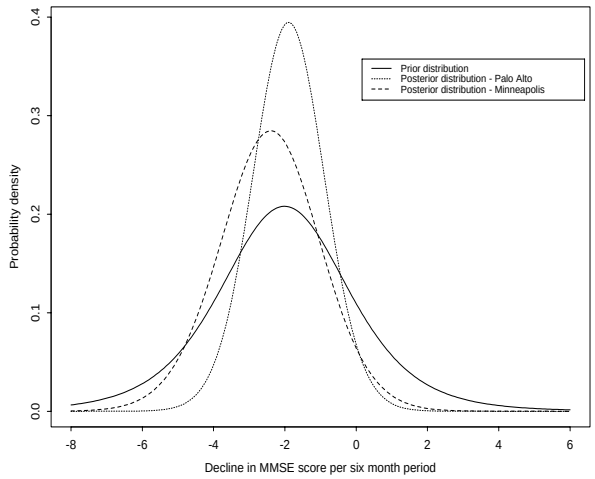

FIGURE 1: Prior distribution and posterior distributions for both Palo Alto and Minneapolis data sets for the mean decline slope μ.

Our hierarchical model assumes that the slope of a randomly selected individual in the population will follow a $N(\mu, \phi^2)$ distribution. This applies to patients selected from either Palo Alto or Minneapolis, although the two posterior distributions will be different since we analyse the two data sets separately. The prior distributions discussed in Section 3.4 are updated by the data

to posterior distributions, summarizing the information now known about μ and ϕ^2 at each site. For each site, we calculate marginal posterior distributions for these parameters obtained by "Rao-Blackwellization" (Gelfand & Smith 1990). Figure 2 displays each of these normal densities, and as such represents the densities of randomly selected slopes from each of these populations. Figure 2 can be interpreted as presenting the distribution of the slope for the "$(n+1)$th" subject drawn from each of the two updated distributions. The original prior density for this same quantity is also given. While Figure 1 displays prior and posterior distributions for the mean slope of a population of subjects, the more widely dispersed distributions of Figure 2 display the individual-to-individual variations about the mean slopes in Figure 1.

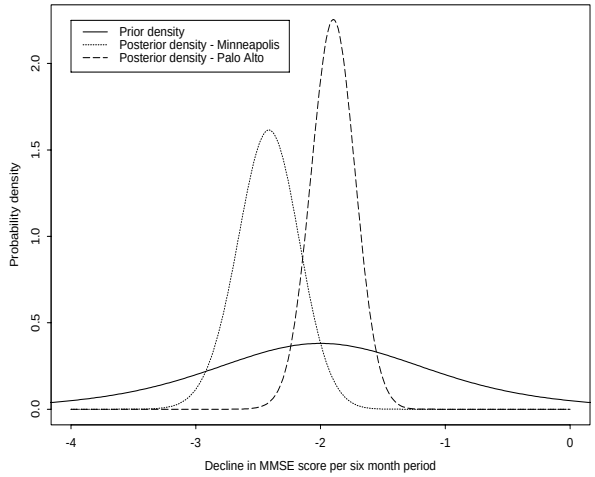

FIGURE 2: Prior distribution and posterior distributions for both Palo Alto and Minneapolis data sets for the decline slopes across individuals. Thus, this figure displays the posterior distributions of the slopes of the next randomly selected subject drawn from each of the two populations, along with this same distribution prior to the consideration of the current data sets.

While prior slopes ranged from approximately -7 to $+3$ points on the MMSE per six-month period, this narrowed to approximately -4 to 0 within the Palo Alto subjects. The range was similar but slightly wider for the Minneapolis data set. Again, this difference may be due to genuine differences between patients in the two different regions. It may also be due at least in part to the shorter follow-up times and to less information about the slopes (despite the slightly larger sample size), leading to a posterior density slightly closer to the prior density. Support for the first explanation comes from Figure 3, which displays histograms of the posterior means of the subject-specific marginal densities for the slopes during decline, plateau heights and duration from disease onset to end of the decline period. These were calculated as the sample means of the appropriate output from the Gibbs sampler. The top pair of histograms suggest that two subjects had evidence of a steep slope in the range of -5 to -6 MMSE points per six-month period. As expected, most plateau scores were below 10, with the majority below 5. Overall, the plateau period scores were lower in the Palo Alto data set, again likely due to the longer follow-up times, which also lead to the longer duration from onset to the end of the decline period.

Approximately 75% of the Palo Alto subjects were classified to have changed (probability that τ_i is less than the time of last follow-up exceeds 0.9) or not to have changed (probability that τ_i is less than the time of last follow-up is less than 0.1) from their decline period to the plateau period by the end of the study, while fewer than half could be as clearly classified in the Minneapolis data set. This is almost surely due to the difference in follow-up times between the two groups

of patients. The average true age-of-onset in the Palo Alto data was estimated to be 60.2 (95% credible interval = 57.5, 63.0) years, while it was 63.7 (95% credible interval = 61.5, 65.8) for the Minneapolis data set. The 95% range of age-of-onset was estimated to be 41.1 to 79.3 in the Palo Alto data set and 47.0 to 80.3 in the Minneapolis data set.

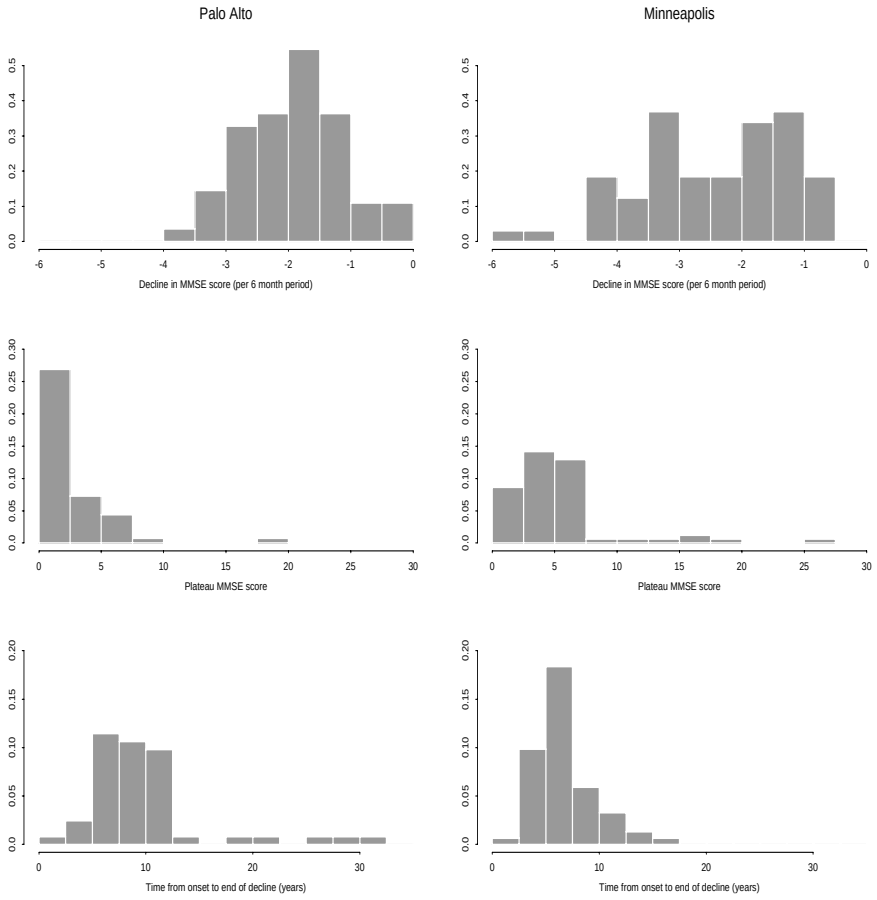

FIGURE 3: Histograms of the means of the subject specific posterior marginal densities for the decline slopes, plateau heights, and duration from disease onset to the end of the decline period, for both Palo Alto and Minneapolis data sets.

An exploratory analysis that plotted estimated slopes and decline durations versus age and gender in both data sets did not reveal any interesting patterns (data not shown). With relatively small data sets, however, we could not adequately investigate the effects of these and other potentially interesting covariates. This issue deserves further study in a larger data set.

4.3 Modeling Covariates

There is a very natural way to include covariates in our model. Referring to Section 3.4, suppose that subject i is associated with the covariate vector $\mathbf{z}_i' = (z_{i1}, z_{i2}, \ldots, z_{ir})$. Suppose that the decline slope, θ_i, for patient i is such that

$$\theta_i \mid \left[\mu\left(\boldsymbol{\beta}'\mathbf{z}_i\right), \phi^2\right] \sim N\left(\mu\left(\boldsymbol{\beta}'\mathbf{z}_i\right), \phi^2\right),$$

where $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_r)'$ is a vector of r regression coefficients. A standard choice for $\mu(\cdot)$ would be $\mu(\boldsymbol{\beta}'\mathbf{z}_i) = \boldsymbol{\beta}'\mathbf{z}_i$. Given a choice of reference prior, say $f(\boldsymbol{\beta}, \phi^2) \propto \phi^{-2}$, it then follows (Gelman, Carlin, Stern & Rubin 1995, p. 236) that

$$\boldsymbol{\beta} \mid [\mathbf{z}, \phi^2, \boldsymbol{\theta}] \sim \text{ Multivariate Normal},$$

with mean vector $\hat{\boldsymbol{\beta}} = (\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'\boldsymbol{\theta}$, and covariance matrix $\mathbf{z}'\mathbf{z}$, where $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_N)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N)$, and $\phi^2|\mathbf{z}, \boldsymbol{\theta} \sim s^2\chi^{-2}_{N-r}$, where

$$s^2 = \frac{1}{N-r} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\beta}}'\mathbf{z}\right)' \left(\boldsymbol{\theta} - \hat{\boldsymbol{\beta}}'\mathbf{z}\right).$$

Because these two distributions are standard, their inclusion in the Gibbs sampler is straightforward. In the case where a proper (conjugate) prior is used, one may apply the methods described in Gelman *et al.* (1995), Section 8.9.

The introduction of covariates allows the distribution from which the slopes θ_i are sampled to depend on these covariates. In this way, the realized individual slopes are tailored to their accompanying covariates. Moreover, because the slopes are sampled independently, even individuals with the same covariate values will have different slopes, thus reflecting the well documented variability in decline among ostensibly similar subjects. The random effects feature of the model is thus maintained.

5. DISCUSSION

The types of models and methods of inference employed here provide powerful alternatives to the limitations of linear mixed models in the analysis of clinical trials of drugs that may slow the progression of Alzheimer's Disease (AD). In particular, by including random change points that are allowed to differ from subject to subject, we allow for non-linearity in the slopes. Indeed, any model not specifically taking the plateau phase into account will lead to underestimation of the decline rate. The hierarchical parameters for the slopes means that decline rates in the treatment and control group may be compared. Thus, the models do not represent simple extensions of linear mixed models, but by incorporating features specific to AD (and certain other chronic diseases) one is led to non-linear mixture models.

The prior and posterior Bayes factors differed somewhat in their choice of final model. This is not too surprising, given that our two leading models, Models 1 and 2, were non-nested. Nevertheless, both criteria agreed that a change-point is needed. Furthermore, our results for the slope were robust to the choice of whether variance parameters were constant or not across individuals. Therefore, our results showed that the mean annual decline in MMSE scores in AD patients is in the range from 4 to 5 points, with most individual slopes in the range from 0 to 8 points per year. The time from disease onset to the end of the decline period, for the subjects of the two centres analyzed, is usually between 5 to 12 years. Hence an "average patient" will decline by about 4 points per year over a 5- or 6-year period, assuming an initial period of no measurable decline in the early stages of the disease, although there is great variability in these trajectories. Figure 4 presents a sample of trajectories.

The information provided by the model presented here could be helpful to clinicians advising patients and their families as to what to expect as the disease progresses, as well as helping in the planning of future clinical trials. In addition to posterior density plots like those in Figures 1 and 2, one can estimate the posterior probability that the mean slope in one group is greater than that in another group. This allows for simple quantitative between group comparisons, such as between two competing treatments in a clinical trial.

The two data sets showed remarkable similarity, for example in the choice of best fitting model and range of values for the slope. Hierarchical modeling was useful in obtaining more stable

estimates of the slopes for each subject, as well as for describing the distribution of the ensemble of slopes across subjects. We did not use hierarchical modeling for the variances or change-points across subjects, but this may be useful for some problems or data sets. Few data sets currently exist with sufficient follow-up time of AD patients for stable estimation of individual patient parameters based on the data from the individual alone. Our model may be applied to larger longitudinal data sets with data collected closer to AD onset times, which would allow for examination of the effects of covariates on decline rates as well as investigation of the early phase of the disease. This would also allow for better estimation of age-of-onset, since estimation of the first change-point in a tri-phasic model would add an earlier lower limit to the onset time. While we have assumed that decline levels off after a sufficiently low score is reached, it would be a simple matter to extend our model to allow for a non-zero slope beyond the change point. In all cases, estimation could proceed by Markov chain Monte Carlo methods similar to those presented here. With applicability in mind, we chose to use an easily interpreted, biphasic though non-standard mixture model for decline. Its use is clearly not restricted to the analysis of MMSE scores, as it may be applied to other measures of decline that depend on tests insensitive to change early or late in the disease, as well as to longitudinal data from other diseases.

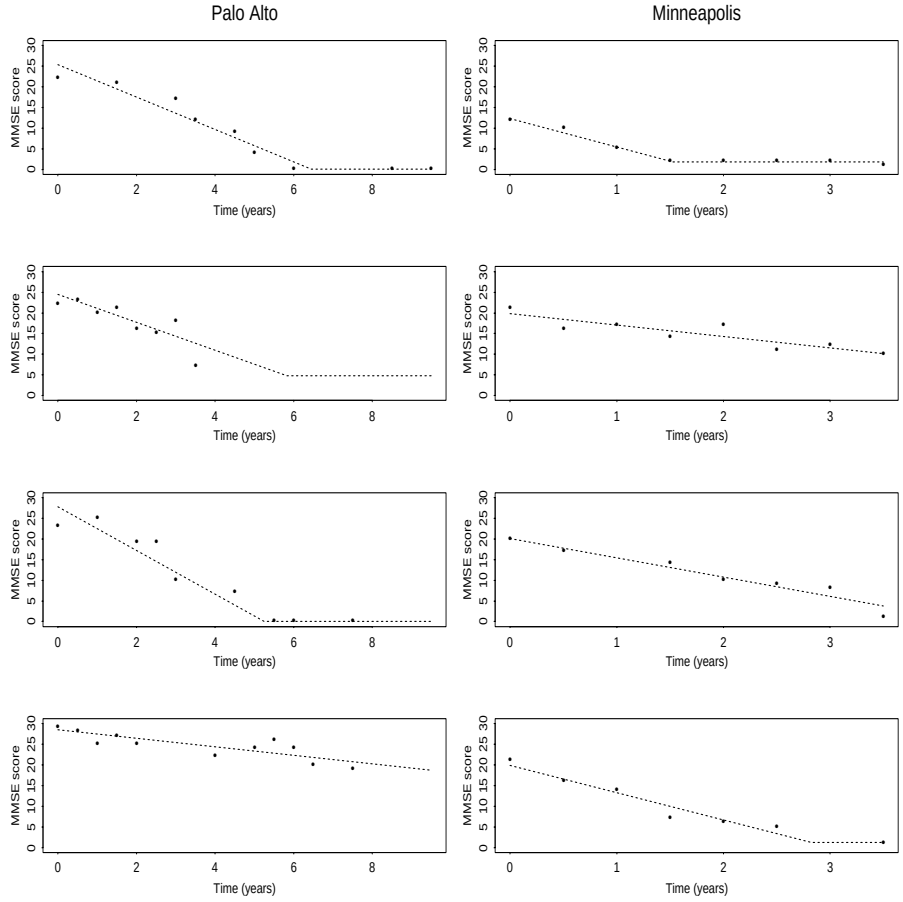

FIGURE 4: Scatter plots of the fitted Model 2 versus observed data for 4 representative subjects in each of the Palo Alto and Minneapolis data sets.

There have been attempts to use simple Markov models to describe decline in patients with AD (Jönsson, Lindgren, Wimo, Jönsson & Winblad 1999; Sonnenberg & Leventhal 1998), with the main goal of assessing the cost effectiveness of proposed treatments. In some cases, categories of cognitive function have been used to define the states of the Markov chain. While perhaps suitable as a tool for an economic analysis, this approach has the disadvantage that the collection of transition probabilities is not as easily interpreted as, say, the slope of a regression line. In addition, the hierarchical Bayesian approach that we believe succinctly accounts for the salient features of decline in AD, would be considerably more difficult to implement, particularly for a random change-point.

APPENDIX

Below we list the full conditional distributions required to implement the Gibbs sampler algorithm discussed in Section 3. We also provide the full conditional distributions needed to obtain the marginal posterior distribution of the ages-of-onset, α_i. Owing to the presence of change points, it is possible that some parameters are not present in each iteration of the Gibbs sampler. We used the pseudo prior method of Carlin & Chib (1995) to accommodate this feature of our model. The full conditional distributions are:

$$\mathrm{P}\left(\tau_i = k | Y_i, \theta_i, h_i, \mu, \phi^2, \sigma_{1i}^2, \sigma_{2i}^2\right) \quad \propto \quad f\left(Y_i | \tau_i = k, \theta_i, h_i, \sigma_{1i}^2, \sigma_{2i}^2\right),$$

where the right-hand side is given by (1) with k substituted for τ_i, for any $k \in \{1, \ldots, 2M\}$;

$$\mathrm{P}\left(h_i = h | Y_i, X_i, \theta_i, \tau_i, \mu, \phi^2, \sigma_{1i}^2, \sigma_{2i}^2\right) \quad \propto \quad f\left(Y_i | h_i = h, \theta_i, \tau_i, \sigma_{1i}^2, \sigma_{2i}^2\right) f(h)$$

for all $h \in \{0, 1, \ldots, 30\}$, and where $f(h)$ is the prior density for h given in Section 3.4;

$$f\left(\theta_i | Y_i, h_i, \tau_i, \mu, \phi^2, \sigma_{1i}^2, \sigma_{2i}^2\right) \quad \propto \quad N(\mu_i^*, \sigma_i^{2*}),$$

where

$$\mu_i^* = \frac{\sum_{j=1}^{\tau_i} \frac{\delta_{ij}(Y_{ij} - h_i)(j - \tau_i)}{\sigma_{1i}^2} + \frac{\mu}{\phi^2}}{\sum_{j=1}^{\tau_i} \frac{\delta_{ij}(j-\tau_i)^2}{\sigma_{1i}^2} + \frac{1}{\phi^2}}$$

and

$$\sigma_i^{2*} = \frac{1}{\sum_{j=1}^{\tau_i} \frac{\delta_{ij}(j-\tau_i)^2}{\sigma_{1i}^2} + \frac{1}{\phi^2}},$$

$$\sigma_{1i}^2 | Y_i, \theta_i, \tau_i, h_i, \mu, \phi^2, \sigma_{2i}^2 \quad \sim \quad s_{11i} \chi_{\nu_{11i}}^{-2},$$

where $s_{11i} = s_{01} + \sum_{j \in D_i} \delta_{ij} \{Y_{ij} - h_i - \theta_i(j - \tau_i)\}^2$, and $\nu_{11i} = n_i + \nu_{01}$;

$$\sigma_{2i}^2 | Y_i, \theta_i, \tau_i, h_i, \mu, \phi^2, \sigma_{1i}^2 \quad \sim \quad s_{12i} \chi_{\nu_{12i}}^{-2},$$

where $s_{12i} = s_{02} + \sum_{j \in P_i} \delta_{ij}(Y_{ij} - h_i)^2$, and $\nu_{12i} = m_i - n_i + \nu_{02}$;

$$\phi^2 \mid Y, \theta, \tau, h, \sigma_1^2, \sigma_2^2 \sim s_1 \chi_{\nu_1}^{-2} \tag{2}$$

$$\mu \mid \phi^2, Y, \theta, \tau, h, \sigma_1^2, \sigma_2^2 \sim N(\mu_1, \phi^2/n_1), \tag{3}$$

where $n_1 = \ell_0 + N$,

$$\begin{aligned} s_1 &= s_0 + \frac{\ell_0 N}{\ell_0 + N} (\mu_0 - \bar{\theta})^2 + \sum_{i=1}^{N} (\theta_i - \bar{\theta})^2 \\ &= s_0 + \ell_0 \mu_0^2 + N\bar{\theta}^2 - n_1 \mu_1^2 + \sum_{i=1}^{N} (\theta_i - \bar{\theta})^2, \end{aligned}$$

$\mu_1 = (\ell_0\mu_0 + N\bar{\theta})/n_1$, and $\nu_1 = \nu_0 + N$. The subscript i has been dropped in equations (2) and (3) to indicate conditioning on the values across all subjects. To obtain the posterior density of the age-of-onset,

$$\begin{aligned} \alpha_1, \ldots, \alpha_N \quad &| \quad a_1, \ldots, a_N, \eta^2 \sim \prod_{i=1}^{n} N\left(a_i, \eta^2\right) \delta_{\{[c_1, r_i)\}}(\alpha_i) \\ \eta^2 \quad &| \quad a_1, \ldots, a_N, \alpha_1, \ldots, \alpha_N \sim S\chi^{-2}_{N-2,T} \ , \end{aligned}$$

where r_i is the age at study onset for subject i, $S = \sum_{i=1}^{N}(a_i - \alpha_i)^2$, and $\chi^{-2}_{N-2,T}$ is an inverse chi-square random variable with $N-2$ degrees of freedom, and truncated to the interval $T = [0, c_2/S]$.

ACKNOWLEDGEMENTS

We would like to thank Dr John Brooks of the Stanford Alzheimer's Disease Diagnostic and Treatment Center for providing the data from Palo Alto, and Dr James Mortimer of the Minneapolis VA Medical Center for providing the data from Minneapolis. We thank the referees, the Associate Editor, and the Editor for their comments, which led to a much improved paper. This work was sponsored, in part, by the Natural Sciences and Engineering Research Council of Canada. Lawrence Joseph is a Senior Scientist funded by the Canadian Institute for Health Research.

REFERENCES

M. Aitkin (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society, Series B*, 53, 111–142.

J. Ashford, M. Shan, S. Butler, A. Rajasekar & F. Schmidt (1995). Temporal quantification of Alzheimer's disease severity: "Time Index" model. *Dementia*, 6, 269–280.

M. J. Bayarri & J. O. Berger (1999). Quantifying surprise in the data and model verification (with discussion). In *Bayesian Statistics 6*, J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, Eds. Oxford University Press, Oxford, pp. 53–82.

J. O. Brooks, H. C. Kraemer, E. D. Tanke & J. A Yesavage (1993). The methodology of studying decline in Alzheimer's disease. *Journal of the American Geriatric Society*, 41, 623–628.

B. P. Carlin & S. Chib (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 57, 473–484.

B. P. Carlin, A. E. Gelfand & A. F. M. Smith (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, 41, 389–405.

S. Chib (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90, 1313–1321.

C. Conigliani & A. O'Hagan (2000). Sensitivity of the fractional Bayes factor to prior distributions. *The Canadian Journal of Statistics*, 28, 343–352.

CSHA (1994). Canadian study of health and aging: Risk factors for Alzheimer's disease in Canada. *Neurology*, 44, 2073–2080.

F. De Santis & F. Spezzaferri (1997). Alternative Bayes factors for model selection. *The Canadian Journal of Statistics*, 25, 503–515.

M. F. Folstein, S. E. Folstein & P. R. McHugh (1975). "Mini-Mental State:" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

A. E. Gelfand & A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.

A. Gelman, J. B. Carlin, H. Stern & D. B. Rubin (1995). *Bayesian Data Analysis*. Chapman and Hall, New York.

A. Gelman & D. B. Rubin (1992). Inferences from iterative simulation and multiple sequences (with discussion). *Statistical Science*, 7, 457–511.

J. V. Haxby, K. Raffaele, J. Gillette, M. B. Schapiro & S. I. Rapoport (1992). Individual trajectories of cognitive decline in patients with dementia of the Alzheimer type. *Journal of Clinical and Experimental Neuropsychology*, 14, 575–592.

H. Jacqmin-Gadda, C. Fabrigoule, D. Commenges & J.-F. Dartigues (1997). A 5-year longitudinal study of the Mini-Mental State Examination in Normal Aging. *Americal Journal of Epidemiology*, 145, 498–506.

L. Jönsson, P. Lindgren, A. Wimo, B. Jönsson & B. Winblad (1999). The cost effectiveness of Donepezil therapy in Swedish patients with Alzheimer's disease: A Markov model. *Clinical Therapeutics*, 21, 1230–1240.

R. E. Kass (1993). Bayes factors in practice. *The Statistician*, 42, 551–560.

R. E. Kass & A. Raftery (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.

R. Katzman, T. Brown, L. Thal, P. Fuld, M. Aronson, N. Butters, M. Klauber, W. Wiederholt, M. Pay & R. Xiong (1988). Comparison of rate of annual change of mental status score in four independent studies of patients with Alzheimer's disease. *Annals of Neurology*, 24, 384–389.

J. Key, L. Pericchi & A. F. M. Smith (1999). Bayesian model choice: What and why? (with discussion). In *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith. Oxford University Press, Oxford, pp. 343–370.

N. M. Laird & J. H. Ware (1982). Random effects models for longitudinal data. *Biometrics*, 38, 963–974.

N. Lange, B. P. Carlin & A. E. Gelfand (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers (with discussion). *Journal of the American Statistical Association*, 87, 615–632.

X. Liu, W.-Y. Tsai & Y. Stern (1996). A functional decline model for prevalent cohort data. *Statistics in Medicine*, 15, 1023–1032.

N. Maltby, G. A. Broe, H. Creasy, A. F. Jorm, H. Christensen & W. S. Brooks (1994). Efficiency of tacrine and lecithin in mild to moderate Alzheimer's disease: Double blind trial. *British Medical Journal*, 308, 879–883.

J. A. Mortimer, B. Ebbitt, S. P. Jun & M. D. Finch (1992). Predictors of cognitive and functional progression in patients with probable Alzheimer's disease. *Neurology*, 42, 1689–1696.

A. Raftery & S. Lewis (1992). How many iteration in the Gibbs sampler? In *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, Eds. Oxford University Press, Oxford, pp. 763–773.

D. B. Rubin (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.

A. F. M. Smith & G. O. Roberts (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 55, 3–102.

F. Sonnenberg & E. Leventhal (1998). Modelling disease progression with Markov models. In *Health Economics of Dementia*, A. Wimo, B. Jönsson, G. Karlsson & B. Winblad, Eds. Wiley, New York, pp. 171–195.

L. Teri, S. M. McCurry, S. D. Edland, W. A. Kukull & E. B. Larson (1995). Cognitive decline in Alzheimer's disease: A longitudinal investigation of risk factors for accelerated decline. *Journals of Gerontology*, *Series A*, 50, M49–M55.

J. A. Yesavage, S. L. Poulsen, J. Sheikh & E. Tanke (1988). Rates of change of common measures of impairment in senile dementia of the Alzheimer type. *Psychopharmacology Bulletin*, 24, 531–534.

Received 6 July 2000
Accepted 22 December 2001

Patrick BÉLISLE: belisle@hirondeau.ri.mgh.mcgill.ca
Department of Epidemiology and Biostatistics, McGill University
Montréal (Québec) Canada H3G 1A4

David B. WOLFSON: david@math.mcgill.ca
Department of Mathematics and Statistics, McGill University
Montréal (Québec) Canada H3A 2K6

Lawrence JOSEPH: lawrence.joseph@mcgill.ca
Department of Epidemiology and Biostatistics, McGill University
Montréal (Québec) Canada H3G 1A4

Xiaojie ZHOU: zhou.x@pg.com
The Proctor and Gamble Company, Mason, OH 45040 USA