

## Modelling the random effects covariance matrix in longitudinal data

Michael J. Daniels<sup>1,\*</sup>,<sup>†</sup> and Yan D. Zhao<sup>2</sup>

<sup>1</sup>*Department of Statistics, University of Florida, Gainesville, FL 32611, U.S.A.*

<sup>2</sup>*Eli Lilly & Company, Lilly Corporate Center, Faris II, Indianapolis, IN 46285, U.S.A.*

### SUMMARY

A common class of models for longitudinal data are random effects (mixed) models. In these models, the random effects covariance matrix is typically assumed constant across subject. However, in many situations this matrix may differ by measured covariates. In this paper, we propose an approach to model the random effects covariance matrix by using a special Cholesky decomposition of the matrix. In particular, we will allow the parameters that result from this decomposition to depend on subject-specific covariates and also explore ways to parsimoniously model these parameters. An advantage of this parameterization is that there is no concern about the positive definiteness of the resulting estimator of the covariance matrix. In addition, the parameters resulting from this decomposition have a sensible interpretation. We propose fully Bayesian modelling for which a simple Gibbs sampler can be implemented to sample from the posterior distribution of the parameters. We illustrate these models on data from depression studies and examine the impact of heterogeneity in the covariance matrix on estimation of both fixed and random effects. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: Cholesky decomposition; heterogeneity; mixed models

### 1. INTRODUCTION

Random effects (mixed) models are a class of models used frequently to model longitudinal data. They offer many advantages including the ability to handle different observation times across subjects and to allow non-stationary covariance structures. In these models, little time is typically spent on modelling the random effects covariance matrix. In particular, examining whether this matrix is the same for all subjects or whether it differs depending on subject-specific characteristics is often neglected in the modelling. For discrete longitudinal data modelled using generalized linear mixed models, ignoring this heterogeneity can result in biased estimates of the fixed effects [1]. For continuous longitudinal data, which will be our

---

\* Correspondence to: Michael J. Daniels, Department of Statistics, University of Florida, Gainesville, FL 32606, U.S.A.

<sup>†</sup> E-mail: mdaniels@stat.ufl.edu

focus here, the standard errors for the fixed and random effects, and consequently, inferences, will be incorrect, the random effects will not be shrunk correctly, and prediction of subject-specific trajectories can be poor. In addition, incorrectly modelling the covariance structure in the presence of missing data can result in biased estimates of fixed effects.

Accounting for heterogeneity in covariance matrices has recently been discussed by several authors. In marginal models, Chiu *et al.*, [2] model the covariance matrix using a log matrix parameterization and obtain estimates using estimating equations. In non-linear mixed models, heterogeneous covariance structures are frequently used, but typically with variance a function of the mean and constant correlations across subjects [3]. Pourahmadi and Daniels [4] develop a class of models they call dynamic conditionally linear mixed models in which the marginal covariance matrix is allowed to vary across individuals, but they consider the random effects covariance matrix to be constant across subjects. In the context of linear mixed models, Lin *et al.*, [5] examined heterogeneity in the within-individual variances in linear mixed models and Zhang and Weiss [6] discussed heterogeneity in the random effects covariance matrix but mainly consider models that allow the entire matrix to differ by a multiplicative factor. Little work has been done on modelling the entire random effects covariance matrix.

Here, we propose an approach that allows all the parameters of the random effects covariance matrix to be modelled, not just the variances. Specifically, we will model the parameters that result from a modified Cholesky decomposition of the covariance matrix (this decomposition has also been called a modified Gaussian [7]). This decomposition has been used previously to model the marginal covariance matrix of longitudinal observations on a subject [4, 8–10]. This decomposition results in parameters that can be easily modelled without concern of the resulting estimator being positive definite, have nice interpretations, and allow for relatively easy model fitting (sampling from the posterior distribution). Some discussion of interpretation in the context of modelling the random effects covariance matrix will be given in Section 2.

We reanalyse a data set previously discussed in Pourahmadi and Daniels [4]. The data was from a series of five depression studies conducted in Pittsburgh from 1982–1992 [11]. Patients were assigned one of two active treatments and measured at baseline and then weekly for 16-weeks. Here, we examine the rate of improvement of, and dependence in, weekly depression scores over this 16-week period for the 549 subjects with no missing baseline covariates. Previous work investigated the time to recovery from depression [11] and examined the rate of improvement in depression scores with a different class of models than those proposed here [4].

Preliminary analyses suggested that the random effects covariance matrix was not the same for subjects with different combinations of treatment (two levels: drug and psychotherapy versus psychotherapy only) and initial severity of depression (two levels: high and low). This motivated our modelling framework, in which we will model the parameters resulting from the modified Cholesky decomposition as a function of both drug (treatment) and severity.

In Section 2, we discuss the modified Cholesky decomposition of the covariance matrix, the interpretation of the parameters in this setting, and its use in modelling the random effects covariance matrix. We build models for the random effects covariance matrix on the depression data and discuss our results in Section 3. Conclusions and discussion comprise Section 4.

## 2. A MODEL FOR THE RANDOM EFFECTS COVARIANCE MATRIX

### 2.1. Model

For each subject  $i$ , we observe an  $n_i \times 1$  vector of responses,  $Y_i$ . We assume the  $Y_i$  follow a random effects model

$$\begin{aligned} Y_i &= X_i\beta + Z_ib_i + \varepsilon_i \\ b_i &\sim N(0, \Sigma_i) \\ \varepsilon_i &\sim N(0, \tau^2 I) \end{aligned} \quad (1)$$

where  $\beta$  is a  $p \times 1$  vector of fixed effects,  $X_i$ ,  $b_i$  is a  $q \times 1$  dimensional vector of random effects with corresponding design matrix  $Z_i$ , and the random effects covariance matrix  $\Sigma_i$  is indexed by subject,  $i$ . The  $b_i$  represent subject-specific deviations from an overall curve. These curves may be modelled using orthogonal polynomials or splines [12, 13]. We order the components of  $b_i$  according to their corresponding basis functions. For example, for orthogonal polynomials, a natural ordering corresponds to overall average, linear, quadratic, cubic, etc. We can similarly order the basis functions for cubic splines. To model  $\Sigma_i$ , we use the modified Cholesky decomposition [8, 14]. We discuss this in detail next.

We decompose the random effects covariance matrix using the approach described in Pourahmadi [8]. This decomposition results in a set of dependence parameters, generalized autoregressive parameters (GARP) and a set of variance parameters, the innovation variances (IV). These parameters can be understood in the context of the dynamic model which is described below. First, we define  $\tilde{b}_{ik}$  to be the linear least-squares predictor of  $b_{ik}$  based on its predecessors  $b_{i,k-1}, \dots, b_{i1}$ , and  $e_{ik} = b_{ik} - \tilde{b}_{ik}$  to be its prediction error with variance  $\sigma_{ik}^2 = \text{var}(e_{ik})$ . Thus, we have

$$b_{ik} = \sum_{j=1}^{k-1} \phi_{i,kj} b_{ij} + e_{ik}, \quad k=1, \dots, q \quad (2)$$

The special Cholesky decomposition of  $\Sigma_i$  is defined as  $T_i \Sigma_i T_i' = D_i$  where  $D_i = \text{diag}(\sigma_{i1}^2, \dots, \sigma_{iq}^2)$  and  $T_i$  is the unit lower triangular matrix with  $-\phi_{i,kj}$  as its  $(k, j)$ th entry. We refer to the  $\phi$  as the GARP and the  $\sigma^2$  as IV. The only constraint on these parameters for  $\Sigma_i$  to be positive definite is that the IV need to be positive.

The GARP parameters characterize the dependence of the random effects and the IV parameters the variance. For example, consider using orthogonal polynomials up to the second order (quadratic) as we use in the example in Section 3. Define  $b_{i1}$  to be the random effect corresponding to the subject-specific deviation from the overall average,  $b_{i2}$  to be the random effect corresponding to the deviation from the overall linear trend, and  $b_{i3}$  to be the random effect corresponding to the deviation from the overall quadratic trend. Then (2) can be written in three parts:

$$\begin{aligned} b_{i1} &= e_{i1} \\ b_{i2} &= \phi_{i,(2,1)} b_{i1} + e_{i2} \\ b_{i3} &= \phi_{i,(3,1)} b_{i1} + \phi_{i,(3,2)} b_{i2} + e_{i3} \end{aligned} \quad (3)$$

where  $\text{var}(e_{ik}) = \sigma_{ik}^2$ ,  $k=1, \dots, 3$ . The first equation corresponds to the marginal distribution of the random effect corresponding to the deviation from the overall average for subject  $i$ , the

second to the conditional distribution of the linear random effect given the overall average; and the third to the conditional distribution of the quadratic random effect, given the linear random effect and the overall average random effect. The GARP are clearly modelling the dependence between the different components of the orthogonal polynomial as we add higher order components one at a time, and the IV are modelling their variability as we add more complexity (higher order terms) to the model ( $\sigma_{i3}^2$  quantifies how well can we predict the deviation from the overall quadratic trend, given the deviations from the overall linear trend and the overall average).

We model the GARP/IV parameters using linear and log link functions

$$\phi_{i,kj} = W_{i,kj}\gamma \quad (4)$$

$$\log(\sigma_{ik}^2) = H_{ik}\lambda \quad (5)$$

where  $W_{i,kj}$  and  $H_{ik}$  are  $q_2$  and  $q_1$  dimensional vectors, respectively. These design vectors are used to model the GARP/IV parameters as functions of subject-specific covariates and/or to model structure on the  $\phi_{i,kj}$  and  $\sigma_{ik}^2$  within a subject. For example, within subject  $i$ ,  $\log(\sigma_{ik}^2)$  might be linear in  $k$ ,  $H_{ik} = (1, k)$ . For some examples of these structured GARP/IV models, see Pourahmadi and Daniels [4]. Under this formulation, it is easy to allow only a subset of the parameters of  $\Sigma_i$  to depend on covariates since there are no concerns about positive definiteness.

## 2.2. Sampling from the posterior distribution

We use a Gibbs sampler with Metropolis–Hastings steps to sample from the posterior distribution of  $(\beta, b_i, \tau^2, \gamma, \lambda)$ . The full conditional distributions of  $\beta, b_i, \gamma, \tau^2$  are all known forms, normal for  $\beta, b_i, \gamma$  and inverse gamma for  $\tau^2$ . To sample from the full conditional distribution of  $\lambda$  we use a Metropolis–Hastings step and use the normal approximation to the full conditional distribution as the candidate distribution. Details on the forms of the full conditionals and the Metropolis–Hastings sampler can be found in the Appendix. Sampling the parameters of the random effects covariance matrix is quite simple under this parameterization as there is no concern about positive definiteness and the full conditional distribution for the dependence parameters is a closed form. Computations are often a major impediment to modelling the covariance matrix in non-standard ways [15].

## 2.3. Other parameterizations of the covariance matrix

The GARP/IV parameterization provides parameters that can easily be modelled without the concern of the estimator being positive definite, that have a sensible interpretation, and that allow for simple computations. However, there are other parameterizations that can be considered. The log matrix [16] and the eigenvalue/Givens angle [17] parameterizations also pose no problems in relation to positive definiteness, but the resulting parameters can be difficult to interpret and model fitting (sampling from the posterior distribution) can be challenging. Parameterizing the matrix using the variances and correlations [18] provides easily interpretable parameters, but is impeded by the restrictions on the variances and correlations for the matrix to be positive definite. This can be troublesome when trying to model these parameters parsimoniously and/or as a function of covariates.

### 3. MODEL BUILDING AND RESULTS

#### 3.1. Depression data

There were a total of 549 subjects in the five depression studies. About 30 per cent (2840) of the possible measurements for the subjects (baseline measure + 16 weeks per subject) were missing, mostly intermittently. This can be seen in Table I, where the sample sizes at each time (out of 549) are given. As in Pourahmadi and Daniels [4], we assume the data are missing at random (MAR). Previous analyses of similar studies and communication with the investigators involved in the study provided support for a MAR assumption. This assumption implies that missingness is explained by measured covariates in the model and/or observed responses on the subject prior to and/or after the missing visit. In addition, some of the missingness was by design as some of the studies only measured patients every other week.

The scale used to quantify depression in these studies was the Hamilton Rating Scale for Depression (HRSD), which was measured at baseline and then weekly for 16 weeks; higher scores on this scale indicate more severe depression. For a detailed description of the psychotherapy treatment and the drugs used in these studies, we refer the reader to Thase *et al.* [1]. The sample means in Table I supported a quadratic trend for the HRSD trajectory over time.

The main questions of interest in this analysis were:

1. Do patients show more improvement on the combination drug/psychotherapy treatment as opposed to the only psychotherapy treatment?
2. How well does the initial severity of the patient predict improvement?

Table I. Weekly sample means (sample sizes) of HRSD broken down by treatment status (drug) and severity (high versus low).

Weeks	Drug		Severity	
	Y	N	High	Low
0	22.17(351)	20.29(198)	23.85(350)	17.34(199)
1	21.03(276)	17.86(93)	21.76(239)	17.43(130)
2	18.55(301)	16.99(92)	19.32(253)	16.14(140)
3	16.34(309)	16.02(92)	17.28(264)	14.31(137)
4	14.51(303)	14.31(104)	15.16(264)	13.16(143)
5	12.70(309)	13.42(106)	13.58(273)	11.56(142)
6	11.94(295)	12.38(104)	12.93(252)	10.55(147)
7	11.14(304)	12.39(109)	12.24(267)	10.06(146)
8	10.01(283)	12.13(104)	11.21(253)	9.40(134)
9	9.73(273)	10.73(121)	10.72(253)	8.80(141)
10	9.19(268)	10.01(115)	10.10(247)	8.23(136)
11	8.77(256)	10.69(103)	9.68(235)	8.65(124)
12	8.64(247)	9.31(104)	9.25(218)	8.17(133)
13	7.86(256)	9.11(97)	8.68(223)	7.38(130)
14	8.23(230)	10.22(93)	8.87(216)	8.67(107)
15	8.38(209)	9.74(96)	9.52(192)	7.60(113)
16	7.40(209)	9.01(83)	8.06(191)	7.49(101)

Table II. Table of models fit with  $W_{i,kj}$  (design vector for GARP) and  $H_{ik}$  (design vector for IV) defined as in (4) and (5).

Simple	$W_{i,kj} = 1$	$H_{ik} = 1$
Drug	$W_{i,kj} = (1, \text{Drug}_i)$	$H_{ik} = (1, \text{Drug}_i)$
Sev	$W_{i,kj} = (1, \text{Sev}_i)$	$H_{ik} = (1, \text{Sev}_i)$
Drug + Sev <sup>★</sup>	$W_{i,kj} = (1, \text{Drug}_i)$	$H_{ik} = (1, \text{Drug}_i, \text{Sev}_i)$
Drug + Sev	$W_{i,kj} = (1, \text{Drug}_i, \text{Sev}_i)$	$H_{ik} = (1, \text{Drug}_i, \text{Sev}_i)$
Drug $\times$ Sev	$W_{i,kj} = (1, \text{Drug}_i, \text{Sev}_i, \text{Drug}_i \times \text{Sev}_i)$	$H_{ik} = (1, \text{Drug}_i, \text{Sev}_i, \text{Drug}_i \times \text{Sev}_i)$

### 3. Is there an interaction between treatment and initial severity as measured by the rate of improvement?

The components of the fixed effects design vector were motivated by these questions and the apparent quadratic trend over time. For the  $i$ th patient at time  $j$ ,  $x_{ij}$  is the  $14 \times 1$  design vector, given by  $x_{ij} = (\text{poly}_j(2), \text{drug}_i \times \text{poly}_j(2), \text{severity}_i \times \text{poly}_j(2), \text{drug}_i \times \text{severity}_i \times \text{poly}_j(2), \text{gender}_i, \text{age}_i)$ , where  $\text{poly}_a(k)$  is a  $k$ th order orthogonal polynomial in  $a$ . Drug is coded as 1 if the subject was on the drug/psychotherapy treatment and 0 if the subject was on the only psychotherapy treatment; severity was coded as 1 for high severity and 0 for low severity; and gender was coded as 1 for male and 0 for female. The random effects design vector is set to  $z_{ij} = (\text{poly}_j(2))$ ; this implies each subject has their own HRSD trajectory.

#### 3.2. Models fit

Based on preliminary exploration of the data, we considered models for  $\Sigma_i$  which included drug and severity. In particular, we considered the models specified in Table II. The simple model is a standard random effects model. The Drug and Sev models allow the entire random effects covariance matrix to differ by drug and severity, respectively. The Drug  $\times$  Sev model includes four separate random effects covariance matrices for each combination of the binary covariates drug and severity. These four models could be fit in SAS proc mixed using likelihood methods [20]; see the Appendix for the syntax. The Drug+Sev and Drug + Sev<sup>★</sup> models are more parsimonious models that involve modelling some or all of the parameters of the modified Cholesky decomposition of the covariance matrix as an additive function of both drug and severity and cannot be fit in SAS.

#### 3.3. Model selection

To select the best models, we use a combination of examining 95 per cent credible intervals for the relevant parameters and computation of an overall statistic of model fit. For the latter, we compute the deviance information criterion (DIC) [19]. The DIC has a form similar to the AIC: a goodness-of-fit term, the deviance evaluated at the posterior mean of the parameters, and a penalty term, two times the effective number of parameters, computed as the mean deviance minus the deviance evaluated at the posterior mean. Thus

$$\text{DIC} = \text{Dev}(\bar{\theta}) + 2p_D \quad (6)$$

where  $\bar{\theta}$  is the posterior mean of  $\theta$  and  $p_D = \bar{\text{Dev}} - \text{Dev}(\bar{\theta})$ , where  $\bar{\text{Dev}}$  is the posterior mean of the deviance. We define the deviance and use the same parameterization as in Pourahmadi

and Daniels [4]

$$\text{Dev}(\theta) = -2 \times \log \text{lik}(\theta|y)$$

where  $\theta = (\beta, A_i^{-1}(\gamma, \lambda, \tau^2))$ , where  $A_i^{-1} = (\tau^2 I + Z_i \Sigma_i(\gamma, \lambda) Z_i^T)^{-1}$ . Based on the form of the DIC, we prefer models with small values. The DIC has the advantage of being easy to compute using output from a Gibbs sampler. We refer the reader to the paper by Spiegelhalter *et al.* [19] for a rigorous justification of this criterion.

Posterior means and 95 per cent credible intervals for  $(\gamma, \lambda)$  appear in Tables III and IV. It is clear that drug and severity are important explanatory factors for the covariance structure. In particular, the additive model for drug and severity appears to fit best. This is confirmed by the table of DIC (Table V) and examination of the 95 per cent credible intervals in Tables III and IV (that is, whether the parameters corresponding to the effects of drug and severity cover zero). The model with the smallest DIC was the one in which drug and severity are entered into the model additively. In addition, from the credible intervals for this model, it appears that the GARP did not depend on severity. We fit a simplified version of the additive model which now had the smallest DIC; this was the model Drug + Sev<sup>★</sup> given in Table II that had the GARP only depend on drug. In addition, given that  $\phi_i$  and  $\sigma_i^2$  were each only three-dimensional here, and there was no apparent structure, we did not pursue more parsimonious models.

The effective number of parameters,  $p_D$ , was roughly what we would expect for all the models. For example, for the simple random effects model,  $p_D = 21$ , the 14 parameters in  $\beta$  plus the one  $\tau^2$  plus six parameters in  $\Sigma$ . However, in some models the effective number was less than expected. This is a general phenomenon observed with the DIC [19] and motivates the terminology ‘effective’ number of parameters.

### 3.4. Estimation and Interpretation of GARP/IV parameters

The estimates of the GARP/IV parameters appear in Tables III and IV. To examine the covariance structure of the random effects, we first focus on the simple model. The log(IV) decrease as we add more complexity to the model (linear, quadratic), but not in a linear fashion. The three GARP (dependence parameters) correspond to the regression parameters from regressing the linear random effect on the overall average random effect ( $\phi_{21}$ ) and from regressing the quadratic random effects on the linear and overall average ( $\phi_{31}$  and  $\phi_{32}$ , respectively) (see the sequence of equations in (3)). The posterior mean of  $\phi_{21}$  is positive and the credible interval is above zero which implies a significant positive relationship (as the overall average increases, the linear trend increases (gets less steep since the slopes are negative)). We see a similar relationship between the quadratic trend and the overall average, after adjusting for the linear trend.

For the final model, the marginal variance of the deviation from the overall average significantly increases for an individual on drug and psychotherapy and for individuals with high initial severity. In addition, the conditional variances  $\sigma_2^2$  and  $\sigma_3^2$  significantly increase for those with high initial severity. For the GARP, the effect of being on drug and psychotherapy versus only psychotherapy indicates a marginally significant weaker association between the deviations from the linear trend and the overall average and between the deviations from the quadratic and the overall average when adjusting for the deviation from the linear trend.

Table III. Posterior means and 95 per cent credible intervals for the random effects covariance matrix parameters  $\gamma$  for the models for the random effects covariance matrix in Table II.

Parameter	Simple	Drug	Sev	Drug + Sev*	Drug + Sev	Drug $\times$ Sev
$\gamma$						
$\phi_{21}$ Int	0.039 (0.031, 0.047)	0.053 (0.035, 0.070)	0.041 (0.026, 0.055)	0.052 (0.034, 0.070)	0.050 (0.029, 0.070)	0.054 (0.026, 0.082)
$\phi_{21}$ Drug	—	−0.018 (−0.038, 0.003)	—	−0.017 (−0.038, 0.003)	−0.017 (−0.037, 0.003)	−0.023 (−0.054, 0.010)
$\phi_{21}$ Sev	—	—	−0.0015 (−0.020, 0.017)	—	0.0033 (−0.015, 0.021)	−0.0025 (−0.039, 0.035)
$\phi_{21}$ Drug $\times$ Sev	—	—	—	—	—	0.0076 (−0.034, 0.049)
$\phi_{31}$ Int	−0.0043 (−0.0049, −0.0037)	−0.0058 (−0.0071, −0.0046)	−0.0053 (−0.0058, −0.0044)	−0.0054 (−0.0065, −0.0046)	−0.0055 (−0.0070, −0.0043)	−0.0052 (−0.0072, −0.0038)
$\phi_{31}$ Drug	—	0.0020 (0.0006, 0.0034)	—	0.0015 (0.0005, 0.0027)	0.0013 (0.0000, 0.0029)	−0.0004 (−0.0022, 0.0030)
$\phi_{31}$ Sev	—	—	0.0012 (0.0000, 0.0022)	—	0.0003 (−0.0008, 0.0016)	0.00047 (−0.0028, 0.0022)
$\phi_{31}$ Drug $\times$ Sev	—	—	—	—	—	0.0015 (−0.0060, 0.0044)
$\phi_{32}$ Int	0.020 (0.012, 0.028)	0.039 (0.026, 0.052)	0.032 (0.024, 0.038)	0.034 (0.024, 0.045)	0.035 (0.021, 0.052)	0.024 (0.011, 0.042)
$\phi_{32}$ Drug	—	−0.027 (−0.044, −0.011)	—	−0.022 (−0.036, −0.007)	−0.018 (−0.035, −0.0020)	0.016 (−0.0060, 0.044)
$\phi_{32}$ Sev	—	—	−0.016 (−0.027, −0.004)	—	−0.004 (−0.022, 0.013)	0.018 (−0.0075, 0.041)
$\phi_{32}$ Drug $\times$ Sev	—	—	—	—	—	−0.050 (−0.088, −0.018)



Table IV. Posterior means and 95 per cent credible intervals for the random effects covariance matrix parameters  $\lambda$  for the models for the random effects covariance matrix in Table II.

Parameter	Simple	Drug	Sev	Drug + Sev*	Drug $\times$ Sev	Drug $\times$ Sev
$\sigma_{\text{Int}}^2$	2.74 (2.61, 2.88)	2.48 (2.26, 2.71)	2.33 (2.11, 2.56)	2.18 (1.90, 2.46)	2.17 (1.89, 2.46)	2.07 (1.72, 2.42)
$\sigma_{\text{Drug}}^2$	—	0.38 (0.10, 0.66)	—	0.30 (0.01, 0.59)	0.30 (0.01, 0.58)	0.49 (0.01, 0.97)
$\sigma_{\text{Sev}}^2$	—	—	0.58 (0.31, 0.86)	0.52 (0.23, 0.81)	0.53 (0.23, 0.81)	0.70 (0.24, 1.17)
$\sigma_{\text{Drug} \times \text{Sev}}^2$	—	—	—	—	—	—0.30 (—0.89, 0.30)
$\sigma_{\text{Int}}^2$	—2.50 (—2.71, —2.31)	—2.43 (—2.81, —2.08)	—2.99 (—3.44, —2.60)	—2.86 (—3.33, —2.43)	—2.87 (—3.33, —2.41)	—2.81 (—3.46, —2.27)
$\sigma_{\text{Drug}}^2$	—	—0.13 (—0.55, 0.31)	—	—0.31 (—0.75, 0.13)	—0.29 (—0.73, 0.15)	—0.50 (—1.37, 0.37)
$\sigma_{\text{Sev}}^2$	—	—	0.68 (0.22, 1.17)	0.79 (0.33, 1.28)	0.78 (0.30, 1.27)	0.67 (—0.07, 1.46)
$\sigma_{\text{Drug} \times \text{Sev}}^2$	—	—	—	—	—	0.25 (—0.80, 1.32)
$\sigma_{\text{Int}}^2$	—8.72 (—9.08, —8.39)	—11.5 (—18.0, —8.91)	—17.7 (—25.2, —12.4)	—16.7 (—28.9, —10.4)	—12.6 (—16.5, —10.2)	—15.4 (—22.0, —10.3)
$\sigma_{\text{Drug}}^2$	—	2.92 (0.29, 9.37)	—	1.87 (—0.36, 9.67)	0.46 (—0.51, 2.03)	—1.95 (—9.68, 6.41)
$\sigma_{\text{Sev}}^2$	—	—	9.48 (4.22, 16.9)	6.62 (1.53, 13.6)	3.86 (1.54, 7.75)	6.65 (1.33, 13.3)
$\sigma_{\text{Drug} \times \text{Sev}}^2$	—	—	—	—	—	2.47 (—6.03, 10.4)

Table V. DIC for the models in Table II.

Model	DIC	$\tilde{D}$	$D(\tilde{\theta})$	$p_D$
Simple	25330	25309	25288	21
Drug	25321	25294	25267	27
Severity	25284	25259	25234	25
Drug + Sev <sup>★</sup>	25279	25250	25220	29
Drug+Sev	25283	25251	25219	32
Drug × Sev	25288	25250	25213	37

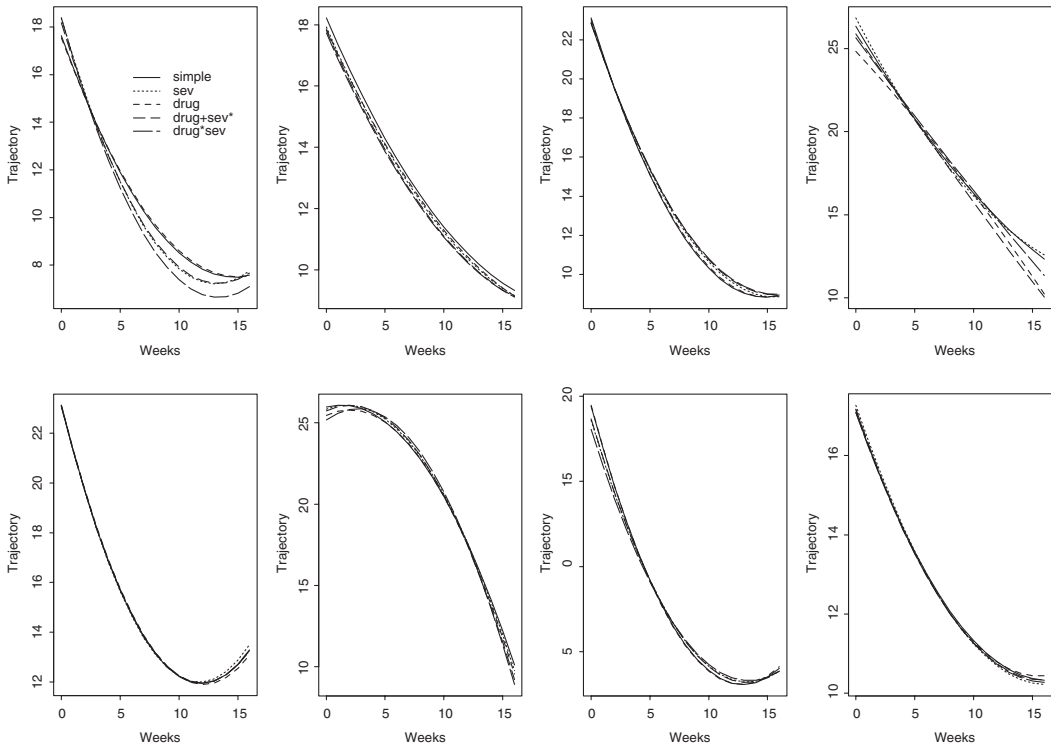


Figure 1. Subject specific trajectories for eight subjects representing large and small  $n_i$  and all four combinations of drug and severity for all the models in Table II. In particular, the first row represents small  $n_i$  and the second row large  $n_i$ . The columns represent no drug/low severity, no drug/high severity, drug/low severity, drug/high severity, respectively.

3.5. Estimation of subject specific trajectories

Figure 1 contains the posterior mean of the subject specific trajectories for eight subjects in the study for all the models in Table II. The first four are subjects with small  $n_i$  (three or four weekly observations) and the last four are subjects with large  $n_i$  (16 or 17 weekly observations) with each set of four representing all combinations of drug and severity. It is

Table VI. Estimated random effects and 95 per cent credible intervals for subjects 1 and 4 from Figure 1.

	Simple	Drug + Sev <sup>★</sup>	Drug × Sev (SAS)
$b_{1,1}$	0.120 (−5.3, 5.7)	−0.23 (−5.1, 4.6)	−0.47 (−5.2, 4.3)
$b_{1,2}$	0.084 (−0.53, 0.70)	0.056 (−0.40, 0.52)	0.023 (−0.43, 0.47)
$b_{1,3}$	−0.0035 (−0.036, 0.029)	0.0013 (−0.016, 0.018)	0.0037 (−0.011, 0.018)
$b_{4,1}$	4.15 (−1.44, 9.65)	3.52 (−2.32, 9.37)	4.09 (−1.77, 9.94)
$b_{4,2}$	−0.103 (−0.73, 0.52)	−0.21 (−1.01, 0.56)	−0.11 (−0.88, 0.66)
$b_{4,3}$	−0.012 (−0.044, 0.020)	−0.018 (−0.051, 0.021)	−0.017 (−0.054, 0.020)

Table VII. Posterior means and 95 per cent credible intervals for the fixed effects coefficients,  $\beta$ .

Covariate	Simple	Drug + Sev <sup>★</sup>	Drug × Sev (SAS)
Intercept	10.9 (10.0, 11.8)	11.0 (10.3, 11.7)	11.0 (10.3, 11.7)
Linear	−0.59 (−0.67, −0.51)	−0.59 (−0.67, −0.51)	−0.59 (−0.67, −0.51)
Quadratic	0.014 (0.009, 0.019)	0.013 (0.009, 0.018)	0.013 (0.009, 0.017)
Severity	3.30 (2.11, 4.49)	3.25 (2.23, 4.30)	3.30 (2.26, 4.34)
Drug	−0.41 (−1.61, 0.79)	−0.49 (−1.46, 0.49)	−0.46 (−1.45, 0.53)
Gender	0.26 (−0.30, 0.84)	0.23 (−0.33, 0.76)	0.21 (−0.33, 0.75)
Age	0.014 (−0.005, 0.032)	0.019 (0.006, 0.037)	0.017 (−0.001, 0.035)
Sev × Drug	−1.43 (−3.00, 0.06)	−1.39 (−2.72, 0.03)	−1.44 (−2.83, −0.047)
Drug × Linear	−0.12 (−0.23, −0.006)	−0.12 (−0.23, −0.022)	−0.12 (−0.22, −0.020)
Drug × Quad	0.0054 (−0.0019, 0.013)	0.0058 (−0.0002, 0.012)	0.0059 (−0.0001, 0.012)
Sev × Linear	−0.18 (−0.30, −0.063)	−0.19 (−0.32, −0.061)	−0.19 (−0.31, −0.065)
Sev × Quad	0.0057 (−0.0020, 0.013)	0.0058 (−0.0017, 0.013)	0.0055 (−0.0020, 0.013)
Sev × Drug × Linear	−0.069 (−0.22, 0.081)	−0.062 (−0.21, 0.091)	−0.062 (−0.21, 0.087)
Sev × Drug × Quad	0.0042 (−0.0058, 0.014)	0.0042 (−0.0047, 0.013)	0.0044 (−0.0048, 0.014)

clear from the figure that subject-specific trajectories can be sensitive to the model for the random effects covariance matrix. In particular subject 1 and 4 have considerable variability in their mean trajectories across models. In addition, the credible intervals for the random effects are sensitive to the model for the random effects covariance matrix. Table VI shows the posterior mean of the random effects and 95 per cent credible intervals. We observe differences in both the posterior means and the widths of the credible intervals when comparing the simple model to the final model. In particular, for subject 1 the credible intervals are considerably tighter for the final model and for subject 4 they are considerably wider (relative to the simple model). We would argue that since the random effects matrix clearly depends on both drug and severity that the confidence intervals for the Drug + Sev<sup>★</sup> model are more accurate than the ones based on the simple random effects model.

### 3.6. Estimation of fixed effects

Posterior means and 95 per cent credible intervals for  $\beta$  appear in Table VII for the simple random effects model and the best fitting random effects model. Given the large sample size,  $n=549$ , it is not too surprising that there are only small differences in the posterior means. However, the width of the 95 per cent credible intervals for most of the  $\beta$  in the Drug + Sev<sup>★</sup>

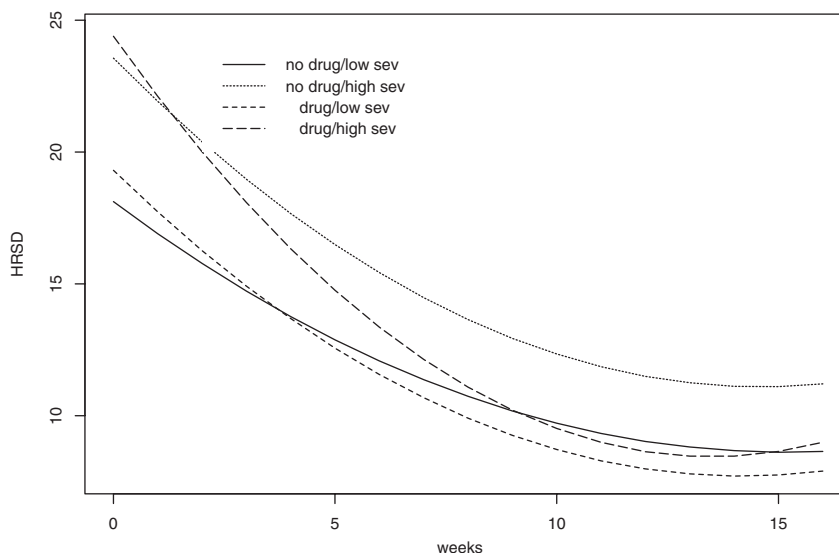


Figure 2. Population averaged trajectories for the best model.

are tighter (narrower). Thus, we can see the potential increase in efficiency from properly modelling the covariance structure here. In particular, we observe the decreased width of the credible intervals for the main effects of drug and severity. In addition, age is significant (credible interval does not cover 0) in the final model, but not in the sample random effects model. (The Gibbs sampler was run for enough iterations so that differences in credible intervals are not due to Monte Carlo error.)

### 3.7. Comparison with SAS

In this section, we provide a comparison of the final model chosen by the DIC criterion with the Drug  $\times$  Sev model which can be fit in SAS proc mixed [20]. The third column of Tables VI and VII contain the estimated random effects and fixed effects under the Drug  $\times$  Sev model fit in SAS. In Table VI, there are some differences in the point estimates of the random effects and slight differences in the confidence intervals. For the estimated fixed effects, given in Table VII, the effect of age goes from significant to marginally significant when using the model fit in SAS and severity by drug interaction becomes significant in the SAS model. Clearly, the overall conclusions changed very little, but specific inferences, as mentioned above, were affected.

### 3.8. Conclusions

There was a significant non-linear decrease in depression scores over the 16 week treatment period (Figure 2). We will first address the first question of Section 3.1. Table VIII contains the posterior means for the week 16 score and the change from baseline for the four combinations of drug and severity. The worst group at week 16 (highest HRSD) was observed in the psychotherapy only/high severity arm and the best in the drug/low severity arm and these

Table VIII. Posterior means and 95 per cent credible intervals for depression score at week 16 and the change from baseline for the best model.

Group	Score at week 16	Change from baseline
No drug/low sev	7.90 (6.77,9.05)	-9.43 (-10.65, -8.19)
No drug/high sev	10.35 (8.85,11.81)	-12.45 (-14.07, -10.86)
Drug/low sev	7.12 (6.12,8.16)	-11.41 (-12.41, -10.39)
Drug/high sev	8.18 (7.30,9.08)	-15.41 (-16.22, -14.59)

were significantly different (95 per cent credible intervals for the difference excluded zero). In terms of change from baseline, the largest change was observed in the drug/high severity arm and the smallest change was seen in the psychotherapy only/low severity arm (and these were significantly different). If we focus on the effect of drug among low and high severity patients, for the outcome at week 16, those on drug were slightly better off (lower HRSD) and this difference was significant among high severity patients. For the change from baseline outcome, those on drug and psychotherapy had much larger improvements than those only on psychotherapy within both the high and low severity patients (and these were significant).

The second question of interest stated in Section 3.1 addresses the impact of initial severity of depression. Individuals with more severe depression at the start of the study did worse on average than those with less severe depression. However, they improved at a quicker rate; there was a larger change from baseline (Table VIII), which was significant, and there was a significant interaction with the linear trend (Table VII) representing a steeper slope for the high severity subjects (also, see Figure 2).

There was not a significant interaction between drug and severity on the rate of improvement as the 95 per cent credible intervals for these two interactions covered zero (Table VII); this addresses the third question of interest given in Section 3.1.

#### 4. DISCUSSION

We have proposed a simple way to model the random effects covariance matrix via structure [4] and/or subject-specific covariates. This approach is computationally attractive and provides parameters to model which have a nice interpretation for modelling trajectories over time; in particular, the covariance parameters correspond to characterizing the dependence and variability of the random effects as we increase the complexity of the subject-specific curves. For the depression example, it is clear that this matrix differs by severity and drug and it affects inferences in terms of both the fixed effects (smaller credible intervals) and prediction of subject-specific trajectories (Figure 1). In our example, the random effects covariance matrix only depended on categorical covariates. If all the parameters in the covariance matrix differ on all possible combinations of these categorical covariates, then a separate matrix can be fit for each combination and the models can be fit in SAS or using a 'standard' Gibbs sampler for Bayesian random effects models. However, if this is not the case, or there are continuous covariates, the models developed here would be necessary and useful.

Clearly, when the main interest in a longitudinal data analysis is the covariance structure or subject-specific predictions, care must be taken in how the covariance structure is modelled.

However, even when these are not of direct interest, taking care is important. For example, when the true random effects matrix is a function of subject-specific covariates, but not modelled that way, the standard errors and confidence intervals for the fixed (and random) effects will be incorrect, which can result in incorrect inferences (for example, a parameter being significantly different from zero). In addition, in cases with unbalanced longitudinal data where the missing data is MAR, incorrectly modelling the covariance structure can even result in biased estimates of the fixed effects [21].

We can compare the DIC of the models in this paper to those in Pourahmadi and Daniels [4]. The best-fitting models in their paper were superior to the best ones here. The reason is that there is considerable non-i.i.d. variability after including the random effects; that is, the residuals in (1),  $\varepsilon_i$ , conditional on the random effects,  $b_i$ , were not independent with constant variance. The models in this paper could be extended to allow a more general form for the covariance of these residuals by replacing the first level covariance matrix  $\tau^2 I$  with a more general form. Non-linear models for the trend might also be considered to avoid the slight upturn in the quadratic curves near week 16, which is likely an artefact of the quadratic model.

Other parameterizations can be used to model the random effects covariance matrix. The GARP/IV parameterization has the advantages of computational attractiveness and a logical interpretation of parameters. This parameterization implies a non-linear model on the variances and correlations. For example, in the simple  $2 \times 2$  case, the covariance matrix expressed in terms of the model for the GARP/IV parameters given in (4) and (5) is

$$\begin{pmatrix} \exp(H_1 \lambda) & \exp(H_1 \lambda) W_{21} \gamma \\ \exp(H_2 \lambda) + \exp(H_1 \lambda) (W_{21} \gamma)^2 \end{pmatrix}$$

One issue with modelling the covariance matrix is choosing the link function; that is, the vector function which transforms the parameters of the original covariance matrix, the variances and covariances, to a new set of parameters on which the regression parameters are linear. For the specific form of the link function in our model, see Pourahmadi [8]. The DIC can be used to choose between various link functions for these models.

Finally, we are exploring extending these models to discrete data (for which bias in the fixed effects is a major issue) and to multivariate longitudinal data.

## APPENDIX

In this Appendix, we provide details on the forms of the prior distributions and full conditional distributions for the parameters used in the Gibbs sampling algorithm. We also provide the code to fit the Drug  $\times$  Sev model in SAS.

### A1. Prior distributions and full conditionals

We assume the prior distributions are

$$\begin{aligned} \lambda &\sim N(\lambda^*, \Omega_\lambda) \\ \beta &\sim N(\beta^*, \Omega_\beta) \end{aligned}$$

$$\begin{aligned}\gamma &\sim N(\gamma^*, \Omega_\gamma) \\ \tau^2 &\sim \text{IG}(a, b)\end{aligned}$$

where IG is an inverse gamma prior, parameterized to have mean  $1/(b(a-1))$ . In general, we recommend specifying these priors to be non-informative. In our example, we chose  $\lambda^*, \beta^*, \gamma^*$  to all be vector's of 0's with the respective  $\Omega$ 's identity matrices multiplied by a scalar which we set equal to 10000. Specifying non-informative priors for variance components can be tricky. Here, we specified  $a=3$  and  $b=9$  for the inverse gamma prior on  $\tau^2$ , based on some preliminary exploration of the data. Owing to the large sample size, however, the posterior distribution for  $\tau^2$  was insensitive to this choice.

The full conditional distributions for  $b_i$ ,  $\beta$ ,  $\tau^2$ , and  $\gamma$  have the following closed forms:

$$\begin{aligned}b_i|Y, \beta, \gamma, \tau^2, \lambda &\sim N\left(\left(\frac{Z_i^T Z_i}{\tau^2} + \Sigma_i^{-1}\right)^{-1} \frac{Z_i^T (Y_i - X_i \beta)}{\tau^2}, \left(\frac{Z_i^T Z_i}{\tau^2} + \Sigma_i^{-1}\right)^{-1}\right) \\ \beta|Y, b_i, \gamma, \tau^2, \lambda &\sim N\left(\left(\frac{\sum_i X_i^T X_i}{\tau^2} + \Omega_\beta^{-1}\right)^{-1} \left(\frac{\sum_i X_i^T (Y_i - Z_i b_i)}{\tau^2} + \Omega_\beta^{-1} \beta^*\right), \right. \\ &\quad \left. \left(\frac{\sum_i X_i^T X_i}{\tau^2} + \Omega_\beta^{-1}\right)^{-1}\right) \\ \tau^2|Y, \beta, b_i, \gamma, \lambda &\sim \text{IG}\left(\frac{\sum_i n_i}{2} + a, \left(2^{-1} \sum_i (Y_i - X_i \beta - Z_i b_i)^T (Y_i - X_i \beta - Z_i b_i) + b^{-1}\right)^{-1}\right) \\ \gamma|Y, \beta, b_i, \tau^2, \lambda &\sim N\left(\left(\sum_i G_i^T D_i^{-1} G_i + \Omega_\gamma^{-1}\right)^{-1} \left(\sum_i G_i^T D_i^{-1} b_i + \Omega_\gamma^{-1} \gamma^*\right), \right. \\ &\quad \left. \left(\sum_i G_i^T D_i^{-1} G_i + \Omega_\gamma^{-1}\right)^{-1}\right)\end{aligned}$$

where  $G_i$  is a  $q \times q_2$  matrix with the first row  $G_{i1}$  equal to 0 and the  $k$ th row  $G_{ik}$  equal to  $\sum_{j=1}^{k-1} W_{i,kj} b_{ij}$  when  $k > 1$ .

The full conditional distribution of  $\lambda$  does not have a standard form. It is proportional to the following:

$$\exp\left\{-(1/2) \times \left[\sum_{i,k} (H_{ik} \lambda + (b_{ik} - G_{ik} \gamma))^2 \exp(-H_{ik} \lambda) + (\lambda - \lambda^*)^T \Omega_\lambda^{-1} (\lambda - \lambda^*)\right]\right\}$$

To sample from this full conditional, we use a Metropolis–Hastings step with a normal approximation to the full conditional as the candidate distribution. For details, see Daniels and Pourahmadi [10]. In our example, the acceptance rate was very high.

The convergence of the Gibbs sampler was monitored by examining time series plots of the parameters over iteration and the Gelman and Rubin approach of using multiple chains [22]. For all the models fit, convergence to the posterior distribution was quick and the mixing was good.

## A2. SAS code for Drug $\times$ Sev model

The SAS code for fitting the Drug  $\times$  Sev model follows:

```
proc mixed data=dep method=reml;
class subject dsc;
model y=linear quad sev drug sex age drug*sev drug*linear drug*quad
      sev*linear sev*quad drug*sev*linear drug*sev*quad / s;
random intercept linear quad / type=un sub=subject grp = dsc;
run;
```

The 'grp=' statement allows the random effects covariance matrix to differ by a categorical covariate. Here, 'dsc' is a covariate that takes four levels, one for each of the drug by severity combinations.

## ACKNOWLEDGEMENTS

The work of the first author was supported by NIH grant CA85295-01A1. The authors would like to thank Zhongqi Zhang who did some of the initial work on deriving the full conditional distributions for the Gibbs sampling algorithm and Professors Joe Hogan and Hal Stern for helping to clarify the presentation of the example and the model. The authors would also like to thank two referees for comments which improved the manuscript.

## REFERENCES

1. Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimate and generalised linear mixed models. *Biometrika* 2001; **88**:973–986.
2. Chiu TYM, Leonard T, Tsui K-W. The matrix-logarithmic covariance model. *Journal of the American Statistical Association* 1996; **91**:198–210.
3. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall: 1995.
4. Pourahmadi M, Daniels MJ. Dynamic conditional linear mixed models for longitudinal data. *Biometrics* 2002; **58**:225–231.
5. Lin X, Raz J, Harlow SD. Linear mixed models with heterogeneous within-cluster variances. *Biometrics* 1997; **53**:910–923.
6. Zhang F, Weiss RE. Diagnosing explainable heterogeneity of variance in random effects models. *Canadian Journal of Statistics* 2000; **28**:3–18.
7. Bock RD. *Multivariate Statistical Methods in Behavioral Research*. McGraw Hill: New York, 1975; 54.
8. Pourahmadi M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterization. *Biometrika* 1999; **86**:677–690.
9. Pourahmadi M. Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika* 2000; **87**:425–435.
10. Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika* 2002; **89**:553–566.
11. Thase ME, Greenhouse JB, Frank E, Reynolds CF 3rd, Pilkonis PA, Hurley K, Grochocinski V, Kupfer DJ. Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. *Archives of General Psychiatry* 1997; **54**:1009–1015.
12. Shi M, Weiss RE, Taylor JMG. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics* 1996; **45**:151–163.
13. Rice JA, Wu CO. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 2001; **57**:253–259.
14. Hedeker D, Gibbons RD. MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine* 1996; **49**:229–252.
15. Daniels MJ, Kass RE. Shrinkage estimators for covariance matrices. *Biometrics* 2001; **57**:1173–1184.
16. Leonard T, Hsu JSJ. Bayesian inference for a covariance matrix. *Annals of Statistics* 1992; **20**:1669–1696.



17. Daniels MJ, Kass RE. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association* 1999; **94**:1254–1263.
18. Barnard J, McCulloch R, Meng X. A natural strategy for modelling covariance matrices with application to shrinkage. *Statistica Sinica* 2000; **10**:1281–311.
19. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 2002 **64**:583–639.
20. SAS Institute Inc. *SAS/STAT Software: Changes and Enhancements through Release. 6.12*. SAS Institute Inc.: Carey, NC, 1997.
21. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. Wiley: New York, 1987.
22. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with Discussion). *Statistical Science* 1992; **7**:457–511.