

Linear Mixed Models for Skew-Normal / Independent bivariate responses with an application to Periodontal Disease

Dipankar Bandyopadhyay^{a*}, Victor H. Lachos^b, Carlos A. Abanto-Valle^c and Pulak Ghosh^d

Bivariate clustered (correlated) data often encountered in epidemiological and clinical research are routinely analyzed under a linear mixed model framework with underlying normality assumptions of the random effects and within-subject errors. However, such normality assumptions might be questionable if the data-set particularly exhibit skewness and heavy tails. Using a Bayesian paradigm, we use the skew-normal/independent (SNI) distribution as a tool for modeling clustered data with bivariate non-normal responses in a linear mixed model framework. The SNI distribution is an attractive class of asymmetric thick-tailed parametric structure which includes the skew-normal distribution as a special case. We assume that the random effects follows multivariate skew-normal/independent distributions and the random errors follow symmetric normal/independent distributions which provides substantial robustness over the symmetric normal process in a linear mixed model framework. Specific distributions obtained as special cases, viz. the skew- t , the skew-slash and the skew-contaminated normal distributions are compared, along with the default skew-normal density. The methodology is illustrated through an application to a real data which records the periodontal health status of an interesting population using periodontal pocket depth (PPD) and clinical attachment level (CAL). Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: Bayesian; linear mixed model; MCMC; normal/independent distributions; skewness

1. Introduction

Periodontal disease usually refers to a collection of inflammatory disease affecting tissues called periodontium that surround and support the tooth and maintains them in the maxillary (upper jaw) and mandibular (lower jaw) bones. If left untreated, it can cause progressive bone loss around the tooth with loosening and eventual loss. It is well documented

^a Division of Biostatistics and Epidemiology, Medical University of South Carolina, USA

^b Departamento de Estatística, Universidade Estadual de Campinas, Campinas, Brazil

^c Department of Statistics, Federal University of Rio de Janeiro, Brazil

^d Department of Quantitative Methods and Information Sciences, Indian Institute of Management, Bangalore, India

*Correspondence to: Division of Biostatistics and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Charleston, SC 29425. E-mail: bandyopd@musc.edu

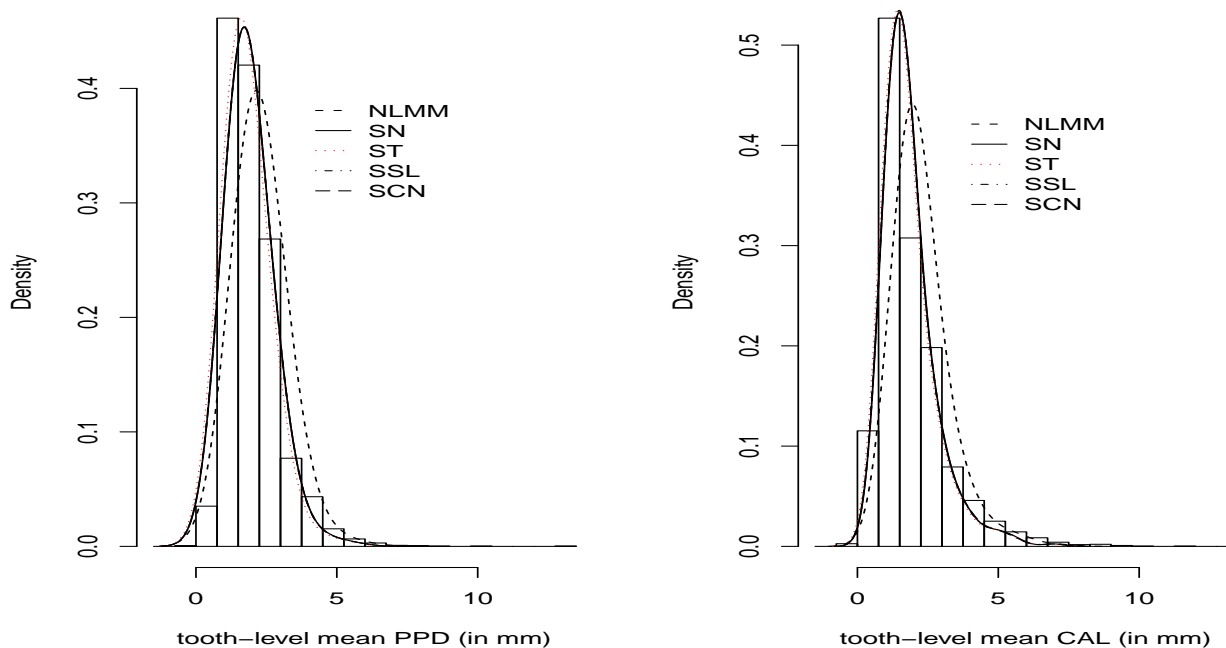


Figure 1. Histograms of tooth-level observed mean PPD and mean CAL scores overlaid with posterior predictive density estimates using normal linear mixed model (NLMM) and skew-normal (SN), skew- t (ST), skew-slash (SSL) and skew contaminated normal (SCN) densities as members of the SNLMM

that some 5% to 15% of any population is susceptible to severe generalized periodontitis worldwide [1]. Being the primary cause of adult tooth loss, it has been estimated that about 50% of U.S. adults over the age of 35 experience early stages of periodontal disease [2]. Periodontal progression is usually assessed by hygienists by measuring two correlated popular bio-markers, viz. periodontal pocket depth (PPD) and clinical attachment level (CAL) [3].

The motivating data example for this paper comes from a clinical study conducted at the Medical University of South Carolina (MUSC) to determine the periodontal disease status of Type-2 diabetic Gullah-speaking African-Americans. For an overall tooth level periodontal status, our bivariate response is the mean of the measurements, i.e. mean PPD and mean CAL for the 6 sites observed simultaneously for each tooth nested within a subject. More details on this data appears in Section 2. With this type of multiple outcome measures, the underlying statistical question is to estimate the functions that model their dependence on co-variates as well as to investigate the relationship between these functions. Similar clinical and epidemiological studies often generate clustered as well as longitudinal follow-up data with bivariate or multivariate outcomes as primary endpoints which are routinely analyzed using multivariate linear mixed models [4]. In this paper, we focus on a linear mixed effects model that accommodates the tooth level clustering within subjects as well as the correlation among bivariate (PPD and CAL) measures and facilitates borrowing of strength across all teeth when assessing the effects of co-variates, viz. age, gender, body mass index (obesity status), glycemic control (diabetic status), etc. on periodontal disease progression. We seek to address the pertinent question: ‘How do potential co-variates influence periodontal status of a particular tooth after accounting for subject-level clustering?’. In traditional linear mixed model (LMM) analysis [5], the correlation due to clustered/repeated measures on a subject are usually accounted for through the inclusion of random effects and within subject measurement-errors which are often assumed to be normally distributed. While such an assumption makes data analysis amenable to popular software like SAS, the usual fidelity to normality assumptions has been questionable [6, 7, 8] when data exhibits non-normal behavior. Figure 1 shows the raw density histogram plots of tooth-level mean PPD (0-14mm) and mean CAL (0-12mm) in our data. The density histograms demonstrate considerable

skewness along with (possible) thick tails. A common approach adopted for data analysis in these situations is reverting back to usual multivariate normality assumptions after suitable transformation of the response (viz. log transform) on a continuous scale. Although they may lead to reasonable empirical results, they may be avoided when a suitable alternative theoretical model is available because data transformation hinders underlying data generation mechanisms due to reduced information and often component-wise transformation does not lead to joint normality [9]. Besides, transformations are not universal, i.e. transforms used for one particular data may not be adapted for a different data. Moreover, mean PPD and mean CAL can have zero values and this hinders the applicability of popular log transformations. This motivates researchers to consider exploration of a more general mixed effects framework that takes into account the flexibility in distributional assumptions of random effects and measurement error to produce robust parameter estimates. The term ‘robustness’ is quite extensive; here robustness is achieved with respect to parameter estimation.

Considerable research has been done by introducing more flexible parametric families that can accommodate normality departures (skewness and heavy tails) and hence eliminate the need of ad-hoc data transformations [10]. In the context of LMM, the random effects distribution was relaxed using finite normal mixtures [6], smoothing [7], a semi-nonparametric (SNP) density [11] or a thick-tailed normal/independent (NI) densities [12]. Much of recent frequentist and Bayesian advances in regression problems revolve around the attractive and popular skew-normal (elliptical) distributions [10, 13, 14, 15]. Related literature in this context is very rich [8, 16, 17] and the curious reader might choose to venture an entire monograph [18] dedicated to discuss recent developments. Starting with the multivariate skew-normal (SN) density [15], SN linear mixed models (SNLMM) were proposed in [9, 19]. A common feature of these classes of SNLMM’s is that the normal linear mixed model (NLMM) is a special member in each class.

In this article, we propose a parametric modeling of LMM for robust estimation using skew-normal/independent (SNI) distributions under a Bayesian paradigm. We assume a SNI distribution for the random effects and a symmetric normal/independent (NI) distribution [27] for the within-subject errors, so that the skew-normal/independent linear mixed model (SNILMM) is defined. The multivariate SNI distributions used in this paper is developed primarily from the multivariate SN density proposed in [14] for Bayesian regression problems and is different from the multivariate SNI densities developed in [20] motivated from the SN version proposed in [15]. However, the differences are only due to the various parameterizations [21] used and an unification of all skew-normal (elliptical) variants is presented in [22]. Recent Bayesian implementation of multivariate SN distributions [9] involves skew-normal (SN) and skew- t (ST) densities using a conditional stochastic representation. Starting from a marginal stochastic representation as in [8, 19], our SNI distributions are amenable to Bayesian implementation and provides a unified class of skew-thick-tailed densities particularly attractive for robust parametric inference and contains as proper elements not only the SN, ST but also the skew-slash (SSL) and the skew-contaminated normal (SCN) densities.

The rest of the paper unfolds as follows. Section 2 illustrates the motivating data behind this research. In Section 3, we introduce the SNI distributions for our bivariate response setup and propose Bayesian inference and related model comparison techniques. We apply our SNILMM to the periodontal data in Section 4 and use model selection tools to determine the best model, comparing between the elements of the SNI class as well as the traditional NLMM. Conclusions and future developments are relegated to Section 5.

2. Periodontal disease data

The motivating data analyzed in this paper was collected from a clinical study [23] conducted at the Medical University of South Carolina (MUSC). The study was primarily aimed to explore the relationship between periodontal disease and diabetes level (determined by HbA_{1c}, or ‘glycosylated hemoglobin’) in Type-2 diabetic Gullah-speaking (or simply Gullah) African-Americans (13 years or older) residing in the coastal sea-islands of South Carolina. The substantial evidence of adverse effects of diabetes on periodontal health [24] has been extensively explored in dental research. The

2006 American Diabetes Association (ADA) Standards of Medical Care recommend diabetic patients strive to maintain the HbA1c < 7 , ideally between 4-6 [25]. Since this is part of an ongoing study, we selected 214 patients with complete covariate information.

To measure periodontal status/progression, dental hygienists usually record the periodontal pocket depth (PPD) and clinical attachment level (CAL), both measured in mm using a manual probe for 6 surfaces per tooth (disto-buccal, mid-buccal, mesio-buccal, disto-lingual, mid-lingual and mesio-lingual) for all 28 teeth per subject, except the third molars. Figure 2(a) provides a pictorial description of the two measures for a single tooth. PPD is defined as the distance (in mm) from the gingival margin to the base of the sulcus/pocket as measured by a periodontal probe. CEJ-GM, or gingival recession is the distance between the free gingival margin and the cemento-enamel junction [3]. The primary measure of perio progression, CAL is defined as $CAL = PPD - (CEJ-GM)$. [3] Clearly, site-level PPD and CAL are correlated. In

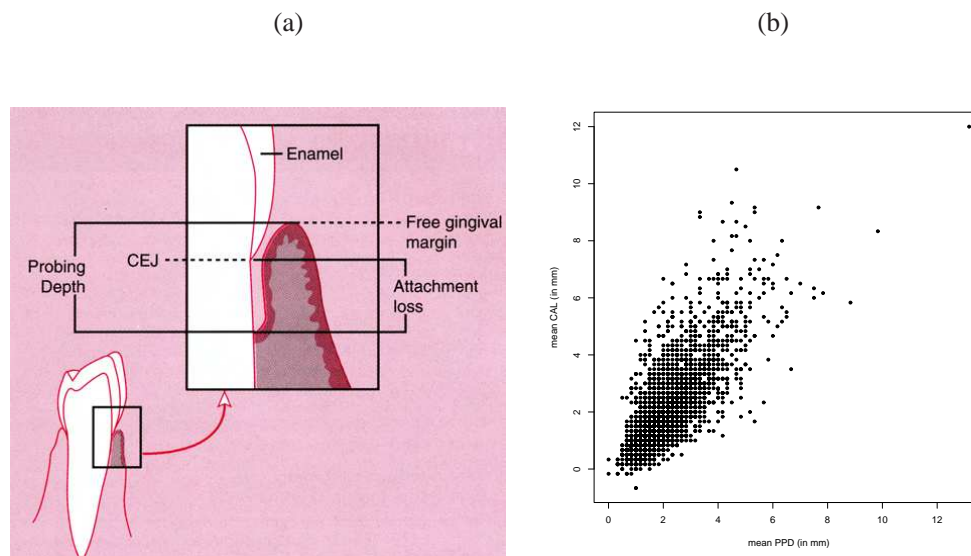


Figure 2. Panel (a): Graphical illustration of PPD and CAL measures for a tooth. This figure was published in ‘*Dental Hygiene: Theory and Practice*’, 1st edition, Michele L. Darby and Margaret M. Walsh, Chapter 17 Page 471, Copyright W.B.Saunders Company (1995); Panel (b): Scatter plot of mean PPD vs mean CAL values

our data, we take the mean PPD and mean CAL measures as representative tooth-level periodontal status clustered within a subject. Note that CEJ-GM was recorded as negative when the free gingival margin recessed below the cemento-enamel junction (CEJ). From the raw plot of mean PPD and mean CAL in Figure 2(b), we suspect some positive correlation between the two response measures. Additionally several subject level covariates were also collected in the study, viz. Age (in years), Gender (1=Female, 0=Male), Body Mass Index or BMI (in kg/m^2), glycemic status or HbA1c (1= High, 0 = controlled), etc. The mean age of the subjects is about 55 years with a range from 26-87 years. Female subjects seem to be predominant (about 75%) in our data, which is not uncommon in this population[26]. About 70% of subjects are obese (BMI ≥ 30) and 65% are with HbA1c ≥ 7 , an indicator of high glycemic level.

3. Statistical Model and Bayesian Inference

3.1. Skew-normal independent distributions

We start with the definition of the SN distribution proposed in [14] as an alternative to [15] for straightforward Bayesian inference. A $p \times 1$ random vector \mathbf{Y} follows a SN distribution with $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ positive definite dispersion matrix $\boldsymbol{\Sigma}$ and $p \times p$ asymmetry matrix $\boldsymbol{\Lambda} = \text{Diag}(\boldsymbol{\lambda})$ where $\text{Diag}(\cdot)$ is a diagonal matrix, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^\top$,

written as $\mathbf{Y} \sim SN_{p,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$, if its pdf is given by

$$f(\mathbf{y}) = 2^p \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Omega}) \Phi_p(\boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\Delta}), \quad (1)$$

where $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$, $\boldsymbol{\Delta} = (\mathbf{I}_p + \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda})^{-1} = \mathbf{I}_p - \boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Lambda}$, \mathbf{I}_p is the $p \times p$ identity matrix, $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_p(\cdot; \boldsymbol{\Sigma})$ are, respectively, the probability density function (pdf) of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and cumulative distribution function (cdf) of $N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Following [21], we use the notation $SN_{p,p}$ since both the symmetric kernel ϕ_p and the skewing function Φ_p are p-variates. Note that for $\boldsymbol{\Lambda} = \mathbf{0}_{p \times p}$ (or $\boldsymbol{\lambda} = \mathbf{0}_{p \times 1}$) where $\mathbf{0}_{p \times p}$ and $\mathbf{0}_{p \times 1}$ are respectively a $p \times p$ matrix and a $p \times 1$ vector of zeroes, (1) reduces to the symmetric $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -pdf, while for non zero values of $\boldsymbol{\Lambda}$, it produces a perturbed (asymmetric) family of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -pdf's.

Following [20], we define a SNI distribution as a process of the p -dimensional random vector

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2} \mathbf{Z}, \quad (2)$$

where U is a positive random variable with cdf $H(u|\boldsymbol{\nu})$ and pdf $h(u|\boldsymbol{\nu})$, which is independent of the $SN_{p,p}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ random vector \mathbf{Z} . Here the parameter $\boldsymbol{\nu}$ is a scalar or vector indexing the distribution of U . Given $U = u$, \mathbf{Y} follows a multivariate skew-normal distribution with location vector $\boldsymbol{\mu}$, scale matrix $u^{-1} \boldsymbol{\Sigma}$ and asymmetry matrix $u^{-1/2} \boldsymbol{\Lambda}$, i.e., $\mathbf{Y}|U = u \sim SN_{p,p}(\boldsymbol{\mu}, u^{-1} \boldsymbol{\Sigma}, u^{-1/2} \boldsymbol{\Lambda})$. Thus, U is affecting both $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$. From (1), the marginal pdf of \mathbf{Y} is:

$$f(\mathbf{y}) = 2^p \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1} \boldsymbol{\Omega}) \Phi_p(u^{1/2} \boldsymbol{\Lambda}^\top \boldsymbol{\Omega}^{-1}(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\Delta}) dH(u|\boldsymbol{\nu}). \quad (3)$$

The notation $\mathbf{Y} \sim SNI_{p,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, H)$ will be used when \mathbf{Y} has pdf (3). When $\boldsymbol{\Lambda} = \mathbf{0}$, the SNI distributions reduces to the respective normal-independent (NI) [27], represented by the pdf $f_0(\mathbf{y}) = \int_0^\infty \phi_p(\mathbf{y}; \boldsymbol{\mu}, u^{-1} \boldsymbol{\Sigma}) dH(u|\boldsymbol{\nu})$. We will use the notation $\mathbf{Y} \sim NI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, H)$ when \mathbf{Y} has distribution in the NI class. The asymmetrical class of SNI distributions includes the skew- t , the skew-slash and the skew contaminated normal distributions, all of which accommodates heavy tails than the SN and can be used for robust inference. For more technical details on skew-normal and skew-normal/independent distributions, we refer to the Supplement A available at <http://people.musc.edu/~bandyopd/SupplementSNIDental.pdf>

3.2. Skew-normal/independent linear mixed models

Now, we summarize the linear mixed model (LMM) for our periodontal progression data with bivariate correlated responses. Let $\mathbf{Y}_i^P = (Y_{i1}^P, Y_{i2}^P, \dots, Y_{im^P}^P)^T$ and $\mathbf{Y}_i^C = (Y_{i1}^C, Y_{i2}^C, \dots, Y_{im^C}^C)^T$ be the measurements (in mm) of tooth-level mean PPD and CAL, respectively, for subject i ($i = 1, \dots, n$). Here, m^P and m^C denotes the number of teeth accounted for within a particular mouth, without loss of generality which is 28 (excluding the third molars). Any missing tooth measurements were not included in the study. Let \mathbf{X}_i^P and \mathbf{X}_i^C be the $m^P \times p_1$ and $m^C \times p_2$ design matrix associated with the fixed effects $\boldsymbol{\beta}^P$ and $\boldsymbol{\beta}^C$ of the two markers respectively and \mathbf{Z}_i^P and \mathbf{Z}_i^C be the corresponding $m^P \times q_1$ and $m^C \times q_2$ design matrices associated with the random effects \mathbf{b}_i^P and \mathbf{b}_i^C respectively. Thus, we have our bivariate LMM:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_i^P \\ \mathbf{Y}_i^C \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i^P & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_i^C \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^P \\ \boldsymbol{\beta}^C \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_i^P & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_i^C \end{bmatrix} \begin{bmatrix} \mathbf{b}_i^P \\ \mathbf{b}_i^C \end{bmatrix} + \begin{bmatrix} \mathbf{e}_i^P \\ \mathbf{e}_i^C \end{bmatrix}, \quad (4)$$

where \mathbf{e}_i^P and \mathbf{e}_i^C are the within-subject residuals for the mean PPD and mean CAL, respectively. For robust estimation, we model \mathbf{b}_i and \mathbf{e}_i simultaneously as

$$(\mathbf{b}_i^\top, \mathbf{e}_i^\top)^\top \stackrel{ind.}{\sim} SNI_{n_i+q, n_i+q}(\mathbf{0}_{n_i+q}, \boldsymbol{\Sigma}_{\mathbf{b}\mathbf{e}_i}, \boldsymbol{\Lambda}_{\mathbf{b}\mathbf{e}_i}, H), i = 1, \dots, n, \quad (5)$$

where $\Sigma_{be_i} = \text{Diag}(\mathbf{D}, \Sigma_i)$ and $\Sigma_i = \text{Diag}(\sigma_1^2 \mathbf{I}_{m^P}, \sigma_2^2 \mathbf{I}_{m^C})$ with $\text{Diag}(\mathbf{A}, \mathbf{B})$ denoting a block diagonal matrix whose elements are the square matrices \mathbf{A} and \mathbf{B} ; $q = q_1 + q_2$, $n_i = m_i^P + m_i^C$, the matrix $\mathbf{D} = \mathbf{D}(\alpha)$ is the dispersion matrix corresponding to within-subject variability and depending on the unknown parameter α , $\Lambda_{be_i} = \text{diag}(\Lambda, \mathbf{0}_{n_i \times n_i})$, with $\Lambda = \text{Diag}(\lambda)$ and $\lambda = (\lambda_1, \dots, \lambda_q)^\top$. Thus the vector λ is the only parameter involved in the asymmetry matrix. Finally, $H = H(\cdot|\nu)$ is the cdf-generator that determines the specific SNI model that we are assuming. Integrating out the variable u_i , it follows from (2) and the stochastic representation A-1 in Supplement A, that

$$\mathbf{b}_i = (\mathbf{b}_i^P, \mathbf{b}_i^C) \stackrel{i.i.d.}{\sim} \text{SNI}_{q,q}(\mathbf{0}, \mathbf{D}, \Lambda, H), \quad \mathbf{e}_i = (\mathbf{e}_i^P, \mathbf{e}_i^C) \stackrel{ind.}{\sim} \text{NI}_{n_i, n_i}(\mathbf{0}, \Sigma_i, H), \quad i = 1, \dots, n.$$

In this modeling, we consider a bivariate generalization of the classical NLMM where the random errors are assumed to follow a NI distribution (with mean zero) and the random effects are assumed to follow a multivariate SNI distribution. As in [20], since for each $i = 1, \dots, n$, \mathbf{b}_i and \mathbf{e}_i are indexed by the same mixing factor U_i , they are not independent in general. The independence corresponds to the case when $U_i = 1$ ($i = 1, \dots, n$), so that the SNILMM reduces to the SNLMM as defined in [19]. However, from (2) and A-1 in Supplement A, conditional on U_i , \mathbf{b}_i and \mathbf{e}_i are uncorrelated, since $\text{Cov}(\mathbf{b}_i, \mathbf{e}_i) = E\{\mathbf{b}_i \mathbf{e}_i^\top\} = E\{E\{\mathbf{b}_i \mathbf{e}_i^\top | U_i\}\} = E\{E\{\mathbf{b}_i | U_i\} E\{\mathbf{e}_i | U_i\}\} = \mathbf{0}$.

Our main focus is to provide inference for $\theta = (\beta^P, \beta^C, \sigma_1^2, \sigma_2^2, \alpha^\top, \lambda^\top, \nu^\top)^\top$. Under the SNILMM proposition defined in (4)–(5), the marginal distribution of \mathbf{Y}_i is given by

$$f(\mathbf{y}_i | \theta) = 2^q \int_0^\infty \phi_{n_i}(\mathbf{y}_i; \mathbf{X}_i \beta, u_i^{-1} \Psi_i) \Phi_q \left(u_i^{1/2} \bar{\Lambda}_i^\top \Psi_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta); \Delta_{z_i} \right) dH(u_i | \nu), \quad (6)$$

where $\Psi_i = \Sigma_i + \mathbf{Z}_i \Omega \mathbf{Z}_i^\top$, $\Delta_{z_i} = \mathbf{I}_q - \bar{\Lambda}_i^\top \Psi_i^{-1} \bar{\Lambda}_i$, $\beta = (\beta^P, \beta^C)^\top$ and $\bar{\Lambda}_i = \mathbf{Z}_i \Lambda$, with $\Omega = \mathbf{D} + \Lambda \Lambda^\top$, $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_q)$, where $\mathbf{X}_i = \text{diag}(\mathbf{X}_{1i}, \mathbf{X}_{2i})$, $\mathbf{Z}_i = \text{diag}(\mathbf{Z}_i^P, \mathbf{Z}_i^C)$. The proof follows from Theorem 1 given in [19]. We call attention to the fact that the marginal distributions of the response vectors given in (6) does not belong to the SNI class introduced in (3), since in general $n_i \neq q$, i.e. the SNI is not closed under linear transformations. The specific forms of the marginal distributions of \mathbf{Y} for the sub-classes of our SNILMM are presented in Supplement B available in the link provided in Section 3.1.

3.3. Priors and joint posterior distributions

In this sub-section, we describe our choice of priors and associated posterior distributions of model parameters to implement Bayesian inference for our SNILMM. A key feature of this model is that it can be formulated in a flexible hierarchical representation. From (2) and the marginal stochastic representation of a SN random vector (see Appendix A-1), it follows that the SNILMM defined by (4) and (5) can be written hierarchically as:

$$\mathbf{Y}_i | \mathbf{b}_i, T_i = t_i, U_i = u_i \stackrel{ind.}{\sim} N_{n_i}(\mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i, u_i^{-1} \Sigma_i), \quad (7)$$

$$\mathbf{b}_i | T_i = t_i, U_i = u_i \stackrel{ind.}{\sim} N_q(\Lambda \mathbf{t}_i, u_i^{-1} \mathbf{D}), \quad (8)$$

$$\mathbf{T}_i | U_i = u_i \stackrel{ind.}{\sim} TN_q(\mathbf{0}, u_i^{-1} \mathbf{I}_q; \mathbb{R}_+^p), \quad (9)$$

$$U_i \stackrel{iid}{\sim} H(u_i | \nu), \quad (10)$$

$i = 1, \dots, n$, where $\mathbf{X}_i = \text{Diag}(\mathbf{X}_{1i}, \mathbf{X}_{2i})$, $\beta = (\beta_P, \beta_C)^\top$, $\mathbf{Z}_i = \text{Diag}(\mathbf{Z}_i^P, \mathbf{Z}_i^C)$, \mathbb{R}_+^p denotes the Euclidean vector space of all p -tuples of positive real numbers and $TN_p(\mu, \Sigma; \mathbf{A})$ denotes a p -variate truncated normal distribution for $N_p(\mu, \Sigma)$ lying within the hyperplane \mathbf{A} . Defining $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$, $\mathbf{t} = (t_1, \dots, t_n)^\top$, $\mathbf{u} = (u_1, \dots, u_n)^\top$ and $\mathbb{I}_{\{A\}}(\cdot)$ to be the indicator function of the set A , it follows that the complete likelihood function associated with

$(\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{t}^\top, \mathbf{u}^\top)^\top$ is given by

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}, \mathbf{u}, \mathbf{t}) \propto \prod_{i=1}^n [\phi_{n_i}(\mathbf{y}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\boldsymbol{\Sigma}_i)\phi_q(\mathbf{b}_i; \boldsymbol{\Lambda}\mathbf{t}_i, u_i^{-1}\mathbf{D}) \times \phi_q(\mathbf{t}_i; 0, u_i^{-1}\mathbf{I}_q)\mathbb{I}_{\{\mathbb{R}_p^+\}}(\mathbf{t}_i)h(u_i|\boldsymbol{\nu})]. \quad (11)$$

Now, to complete the Bayesian specification of the model, we need to put prior distribution on all the unknown parameters in $\boldsymbol{\theta}$. Since we have no prior information from historical data or from previous experiments, we assign conjugate but weakly informative priors to obtain well defined and proper posteriors. A popular choice to ensure posterior propriety in a LMM is to consider proper (but diffuse) conditionally conjugate priors like non-informative Normal priors (with large variance) for the fixed-effects, inverse gamma priors for a single variance components and inverse Wishart priors for the variance-covariance matrix, as suggested in [31, 32]. In general, we choose:

$$\begin{aligned} \boldsymbol{\beta} &\sim N(\boldsymbol{\beta}_0, \mathbf{S}_\beta), \\ \sigma_k^2 &\sim IG(\tau_k/2, T_k/2), \quad k = 1, 2, \\ \mathbf{D} &\sim IW_q(\mathbf{T}_b, \tau_b), \\ \boldsymbol{\lambda} &\sim N_q(\boldsymbol{\lambda}_0, \mathbf{S}_\lambda), \end{aligned}$$

where $N(\cdot, \cdot)$ is the multivariate normal density, $IG(a, b)$ is the inverse gamma density with parameters a and b and $IW_q(\mathbf{T}_b, \tau_b)$ is the q -variate inverse Wishart distribution where \mathbf{T}_b is a $q \times q$ positive definite scale matrix and hyperparameter τ_b is the degrees of freedom. The prior distribution of $\boldsymbol{\nu}$, with density $\pi(\boldsymbol{\nu})$, depends on the particular SNI distribution we use. The specific distributions, i.e. skew- t (ST), skew-slash (SSL) and skew contaminated normal (SCN) are discussed in Appendix A and related prior choices are described in more details in Section 4. Assuming elements of the parameter vector to be independent, the joint prior distribution of all unknown parameters is given by

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta}) \left(\prod_{k=1}^2 \pi(\sigma_k^2) \right) \pi(\boldsymbol{\lambda})\pi(\mathbf{D})\pi(\boldsymbol{\nu}). \quad (12)$$

Combining the likelihood function (11) and the prior distributions, the joint posterior distribution for $\boldsymbol{\theta}$ is now,

$$\pi(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \mathbf{D}, \boldsymbol{\lambda}, \mathbf{b}, \mathbf{u}, \mathbf{t}|\mathbf{y}) \propto \prod_{i=1}^n [\phi_{n_i}(\mathbf{y}_i; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\boldsymbol{\Sigma}_i)\phi_q(\mathbf{b}_i; \boldsymbol{\Lambda}\mathbf{t}_i, u_i^{-1}\mathbf{D}) \times \phi_q(\mathbf{t}_i; 0, u_i^{-1}\mathbf{I}_q)\mathbb{I}_{\{\mathbb{R}_p^+\}}(\mathbf{t}_i)h(u_i|\boldsymbol{\nu})]\pi(\boldsymbol{\theta}). \quad (13)$$

Distribution (13) is analytically intractable but MCMC methods such as the Gibbs sampler and Metropolis-Hastings algorithm can be used to draw samples, from which features of marginal posterior distribution of interest can be inferred. Given \mathbf{u} , all conditional posterior distributions are as in a standard SNLMM and have the same form for any element of the SNI class. An outline of the conditional posteriors of all model parameters including \mathbf{u} and $\boldsymbol{\nu}$ (for specific SNI distributions) are given in the Appendix.

3.4. Model selection and goodness-of-fit

To select our best fitting model and related goodness of fit assessments, we compare among the SN, ST, SSL and SCN models as well as the NLMM using Bayesian model selection tools. Specifically, we consider both deviance-based criterion [28] and measures based on posterior predictive performance [30].

The DIC [28] is a deviance-based measure appropriate for Bayesian model selection and is defined as $DIC = \bar{D} + p_D$, where \bar{D} is the posterior expectation of the deviance summarizing model-fit, p_D is a measure of model complexity defined

as $\bar{D} - D(\bar{\Theta})$, where $D(\bar{\Theta})$ is the deviance computed at the posterior mean of parameters and Θ denote the parameter space. Analogous to the AIC, the DIC summarizes the relative fit between a model and the ‘true model’ generating the data conditional on the data clusters, i.e., the study subjects using a single number summary with smaller values indicating better fit.

We also consider model selection based on predictive performance of competing models. If \mathbf{y}_{pr} denotes the predictive data vector, then the posterior predictive distribution is given by:

$$p(\mathbf{y}_{pr}|\mathbf{y}) = \int p(\mathbf{y}_{pr}|\Theta)p(\Theta|\mathbf{y})d\Theta. \tag{14}$$

One can obtain predictive data easily from a converged posterior sample and samples from the posterior predictive distribution are replicates of the observed model generated data. Using a squared-error loss function [29], we compare competing models based on the expected total predictive deviance (ETPD) defined by

$$ETPD = E\left\{\sum_{ij} (y_{ij,pr}^P - y_{ij}^P)^2 | y_{ij}^P\right\} + E\left\{\sum_{ij} (y_{ij,pr}^C - y_{ij}^C)^2 | y_{ij}^C\right\}$$

where, $y_{ij,pr}$ denote a replicate of the observed y_{ij} , the summations are taken over all observations and the expectations taken over the full posterior of all model parameters. Similar to DIC, this criterion chooses the model where the predictive values are centered near observed values, i.e. with the lowest predictive variation.

To determine model adequacy after selecting the best model, we use a discrepancy measure based on (14). If the observed value is extreme relative to the reference distribution (the posterior predictive distribution), there is some concern with respect to assessment of model-fit to the data. Define y_{ij}^C and y_{ij}^P to be the observed data on tooth-level mean CAL and PPD scores respectively. The discrepancy measure between model and data is computed as a summary statistic [30] using model parameters and data defined as

$$T^k(y, \Theta) = \sum_{i,j} \frac{(y_{ij}^k - E\{y_{ij}^k|\Theta\})^2}{Var\{y_{ij}^k|\Theta\}}, \quad k = P, C, \tag{15}$$

The Bayesian p-value/ posterior predictive p-value [30] p_B^k is defined as the number of times $T^k(y_{pr}, \Theta)$ exceeds $T^k(y, \Theta)$ out of L simulated draws i.e. $p_B^k = \Pr(T^k(y_{pr}, \Theta) \geq T^k(y, \Theta)|y)$, where y_{pr} denotes a simulated draw from (14). A very large p-value (> 0.95) or a very small (< 0.05) both signals model misspecification, i.e. the observed pattern would be unlikely to be seen in replications of the data under the true model [30]

4. Data analysis and findings

In this section, we apply our method to the periodontal data described in Section 2. Our SNI class of distribution allows continuous variation from symmetry to asymmetry and accommodates practical values of kurtosis. We posit 5 competing models with latent (unobserved) random effects and random errors from the SNI class. The models are as follows:

Model 1 (N): Normal distribution for the latent random effect and random error.

Model 2 (SN): Skew-normal distribution for the latent random effect and normal distribution for the random error (SN).

Model 3 (ST): Skew- t distribution for the latent random effect and Student- t distribution for the random error.

Model 4 (SSL): Skew-slash distribution for the latent random effect and slash distribution for the random error.

Model 5 (SCN): Skew contaminated normal distribution for the latent random effect and contaminated-normal distribution for the random error.

In the absence of historical data/experiment, we specify practical weakly-informative priors for all model parameters to obtain well-defined (proper) posteriors following the recommendations in [31, 32]. The components of β^P and β^C were

Table 1. Model comparison using DIC and ETPD criteria

Criterion	N	SN	ST	SSL	SCN
\bar{D}	20815.2	20805.7	17686.0	17732.2	18700.0
p_D	384.85	383.56	587.43	563.73	565.24
DIC	21200.05	21189.26	18273.43	18295.93	19265.24
ETPD	2.31	2.24	2.196	2.27	2.251

assigned independent Normal(0, Precision = 0.01) priors. For the scale parameter σ_k^2 , $k = 1, 2$, we assign a moderately diffuse IG(0.01, 0.01), so that the distribution has mean 1. The prior for the variance-covariance matrix \mathbf{D} is taken to be weakly-informative Inverse-Wishart with covariance $\mathbf{T}_b = \text{Diag}((0.01, 0.01)^\top)$ and τ_b (degrees of freedom) = 6. For the asymmetry parameters λ_1 and λ_2 , independent Normal(0, Precision = 0.01) are used to accommodate either positive or negative skewness and allow the data to determine it, although histogram plots reveal right skewness. Prior choice for ν follows exactly as in [33]. For the ST distribution, we choose ν as $\text{Exp}(0.1)\mathbb{I}_{\{(2, \infty)\}}$ (i.e. exponential density truncated at 2) to reflect a prior on ν (the t degrees of freedom parameter) with a well defined and finite variance of Y . For the SSL distribution, the prior for ν is a Gamma(a, b) with small positive values of a and b ($a = 0.01, b = 0.001$), primarily to ensure conjugacy. For the SCN distribution, $\nu = (\nu_1, \nu_2)^T$ and once again for posterior conjugacy, ν_1 is chosen as $U(0, 1)$ and ν_2 as $\text{Beta}(1, 1)$ (= $U(0, 1)$).

For each of the models, we ran 2 chains with widely dispersed initial values. For all the 4 models, viz. N, SN, ST and SSL, we used 80000 iterations with an initial burn-in of 30000. However for the SCN model, convergence was achieved at 60000 iterations. The complexity of the SNI structure is manifested in the relatively high burn-in size. However, this is straightforward to program within WinBUGS and associated code is available from the first author on request. Posterior convergence was assessed using trace plots, autocorrelation plots and the Gelman-Rubin scale-reduction factor \hat{R} [30]. To reduce autocorrelation among successive Markov draws, we used a spacing of 5. After discarding the initial burn-in samples, we used the remaining samples to compute the posterior parameter estimates.

Table 1 presents the comparison among the 5 competing models using Bayesian model choice criterion. Note that all the skewed versions produced better fit (in terms of DIC and ETPD) than the NLMM. In particular, the skew- t (ST) model (with the smallest \bar{D}) produces the best fit among the competing skew models. The SSL model is a close competitor with a lower p_D . Comparing ETPD values, the ST model outperforms all other models, though not substantially. We select Model 3 (ST) as our best fitting LMM. Figure 1 shows the smoothed posterior predictive densities overlaid on raw data histograms for our competing models. From there, it is also quite clear that our class of SNI densities provide a substantial better fit to our data over the NLMM. Also among the 4 competing models, the density for the ST model seems to produce a better fit to the raw data histogram for both the responses. The posterior mean of p_B^P and p_B^C for the ST model are 0.494 and 0.517 respectively, indicating no overall lack of fit using our omnibus statistic (15). Table 2 provides posterior estimates of asymmetry parameters (λ_1, λ_2), the variance components of random errors (σ_1^2, σ_2^2), the variance-covariance matrix (D) and ν specific to ST, SSL and SCN distributions. In particular, we provide estimates of posterior mean, standard deviation (SD) and 95% credible intervals (CI) for the ST (our best fitting model) and mean and SD for the other models. Interestingly, both λ_1 and λ_2 are significant and positive for all the 4 fitted models providing evidence of moderate right-skewness for our data. Figure 3 shows the box-plots for the parameters λ_1 and λ_2 for all the 4 models. Note that the CI does not include zero for all the models, confirming positive asymmetry of the bivariate responses with mean CAL to be more right-skewed than mean PPD. In particular, estimate (95% CI) of λ_1 and λ_2 for the ST model is 0.3965 (0.3359, 0.4981) and 0.5959 (0.5431, 0.6637) respectively. Figure 4 plots the marginal posterior densities of the parameter ν for ST and SSL densities and $\nu = (\nu_1, \nu_2)$ for the SCN density. All density plots shows some degrees of right asymmetry confirming non-normal nature. For the ST and SSL, as ν (the t degrees of freedom) $\rightarrow \infty$ (say 30), it approaches the normal density. However, the posterior mean estimate of ν for ST and SSL model are 3.82 and 1.04 respectively, which confirms its sufficient disparity from the normal framework. Also, the contaminated normal density \rightarrow normal density when the

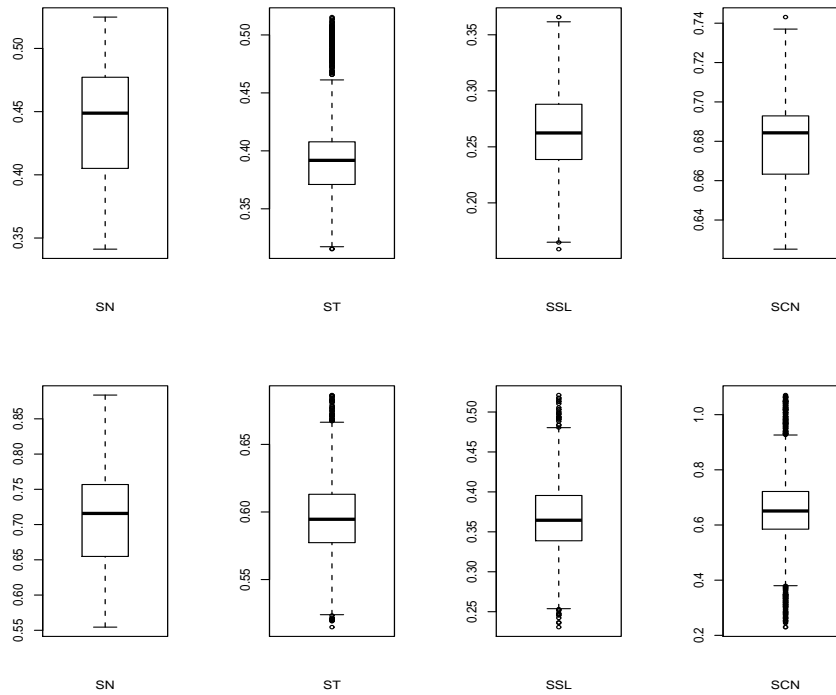


Figure 3. Box plot of asymmetry parameter for the 4 fitted models. The upper row is for λ_1 and the lower row for λ_2 .

Table 2. Posterior parameter estimates of fitting asymmetries and variance components to the periodontal progression data. SD, 2.5% and 97.5% represents standard deviation and percentiles from the posterior distributions of parameters, respectively. *s* denotes parameter significance.

Parameter	SN		ST		SSL		SCN			
	mean	SD	mean	SD	2.5%	97.5%	mean	SD		
λ_1	0.4862 ^s	0.0575	0.3965 ^s	0.0391	0.3359	0.4981	0.2635 ^s	0.0306	0.3701 ^s	0.0327
λ_2	0.7138 ^s	0.0765	0.5959 ^s	0.0286	0.5431	0.6637	0.3619 ^s	0.0373	0.5032 ^s	0.0416
σ_1^2	0.4353	0.0095	0.2464	0.0131	0.2221	0.2731	0.1612	0.0098	0.3323	0.0108
σ_2^2	0.6889	0.015	0.3546	0.0188	0.3204	0.3938	0.1123	0.0067	0.2286	0.0064
D_{11}	0.4177	0.0522	0.1477	0.023	0.109	0.1971	0.0698	0.0121	0.1549	0.0209
D_{12}	0.5273	0.0642	0.177	0.0263	0.1311	0.2322	0.0797	0.0126	0.1912	0.0246
D_{22}	0.6747	0.0987	0.2238	0.0442	0.1467	0.3193	0.098	0.0185	0.2382	0.0347
ν (ν_1 for SCN)			3.82	0.368	3.1171	4.552	1.041	0.0366	0.3388	0.0338
ν_2									0.2404	0.0079

proportion of contaminants (posterior estimate of ν_1) is 0. For the SCN model, the posterior mean of ν_1 is about 0.34 while the posterior density of ν_2 has a mode around 0.247. The estimate of the within-subject variances σ_1^2 and σ_2^2 for both mean PPD and mean CAL are smaller in the skewed class of models as compared to the NLMM (not shown here), primarily because of interrelation between high variability, heavy tails as well as skewness. Table 3 provides posterior estimates of the fixed-effects parameters obtained by fitting our SNI models. Similar to Table 2, we provide posterior mean, SD and 95% CI's for the ST model and the posterior mean and SD for the other models. For the ST model, the glycemic control (as determined by Hba1c) is positive and significant, indicating that mean PPD and mean CAL values seems to be higher for elevated levels of Hba1c controlling for the other covariates. Although Hba1c is not significant in all other competing models, it was significant in the mean PPD regression for the SL model. Estimates of 95% CI's for almost all fixed effects parameters (for both PPD and CAL regressions) in the ST model were tighter as compared to all other competing models. This is expected, because the ST density seems to provide the most precise fit to this data set. The estimated overall posterior correlation between mean PPD and mean CAL obtained from fitting the ST model is 0.378 which also confirms

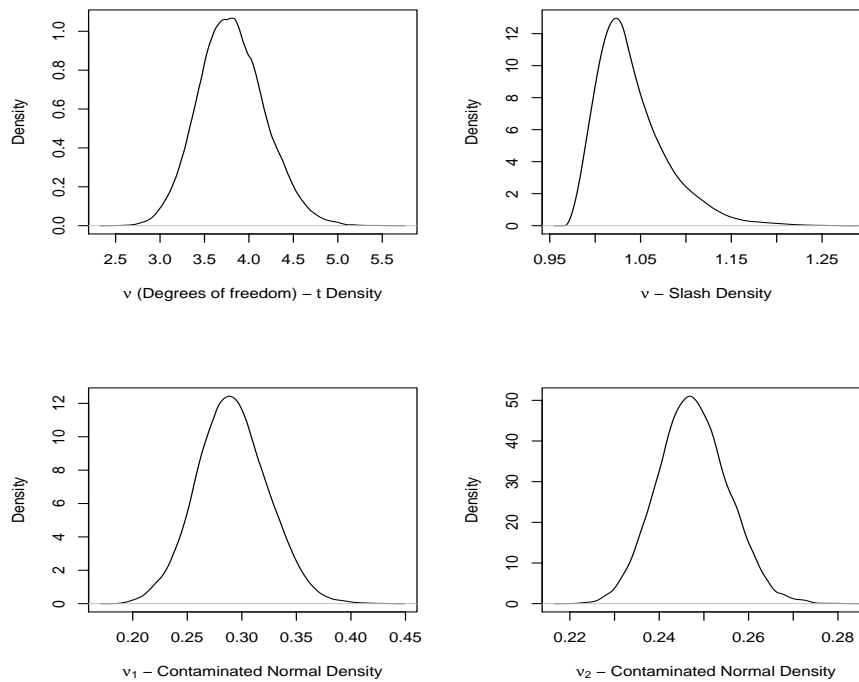


Figure 4. Marginal posterior densities estimates of parameter ν for all 4 distributions. The upper panel display plots for skew- t and skew-slash densities and the lower panel for skew contaminated normal density.

Table 3. Posterior estimates of fixed effects of fitting S-SNI models to the periodontal progression data. SD, 2.5% and 97.5% represents respectively the standard deviation and percentiles from the posterior distributions of parameters. s denotes parameter significance.

Parameter	SN		ST		SSL		SCN			
	mean	SD	mean	SD	2.5%	97.5%	mean	SD	mean	SD
Int ^P	1.665 ^s	0.3639	1.6231 ^s	0.1984	1.175	1.983	1.733 ^s	0.2427	1.744 ^s	0.2691
Gender ^P	-0.034	0.0978	-0.009	0.0861	-0.175	0.180	-0.0018	0.0791	-0.049	0.0889
Age ^P	-0.001	0.0048	-0.004	0.0025	-0.009	0.0001	-0.0048	0.0031	-0.0043	0.0035
BMI ^P	0.0012	0.0049	-0.001	0.0036	-0.007	0.007	-0.0025	0.0043	-0.0023	0.0048
HbA1c ^P	0.0730	0.1093	0.1514 ^s	0.0648	0.0169	0.2849	0.1604 ^s	0.0798	0.1594	0.071
Int ^C	1.025 ^s	0.4657	1.102 ^s	0.2184	0.651	1.48	1.138 ^s	0.2715	1.215 ^s	0.2937
Gender ^C	-0.0868	0.1359	-0.051	0.0858	-0.237	0.1598	-0.024	0.102	-0.073	0.1129
Age ^C	0.0031	0.0062	-0.001	0.003	-0.006	0.004	-0.0018	0.0032	-0.0009	0.0041
BMI ^C	0.0067	0.007	0.004	0.0042	-0.003	0.013	0.0019	0.0052	0.0019	0.0059
HbA1c ^C	0.1101	0.1432	0.1594 ^s	0.073	0.0221	0.3096	0.157	0.0891	0.1346	0.0904

some degree of positive association between the two measures. For hierarchical GLMMs, use of weakly-informative priors can lead to inference which are sensitive [32, 34] to the choice of priors on hyperparameters. To investigate this issue, we conducted sensitivity analysis on the routine use of inverse-gamma prior [35] on variance components as well as the inverse-Wishart prior of the variance-covariance matrix. In all the results, we focused our attention on the estimation of the fixed effects parameters β^P and β^C . We considered an array of weakly-informative to highly non-informative choice of priors. In particular, we took $\sigma_i^2 \sim \text{IG}(10^{\kappa_1}, 10^{\kappa_1})$, $i = 1, 2$, where $\kappa_1 \in \{-4, -3, -2, -1, 0, 1, 2\}$ and the prior choice on the scale matrix \mathbf{D} to be Inverse-Wishart with covariance $T = \text{Diag}((k, k)^\top)$, where $k = 0.001$ and 0.0001 . Although we notice slight changes in the values of fixed effects estimates as well as model comparison measures, results were quite robust on the overall and did not change any conclusions regarding our best fitted model, the posterior estimates of λ and ν and the strength of correlation between mean PPD and mean CAL.

5. Conclusions

Using a Bayesian linear mixed model framework, this paper considers a class of multivariate skew-normal/independent (SNI) regression models to jointly analyze mean PPD and mean CAL as determinants of periodontal progression. Our class of SNI models contains as a subclass some interesting family of models, viz. skew- t , skew-slash and skew contaminated normal densities. The nice hierarchical representation given in (7-10) provides easy model implementation using conventional Bayesian software like WinBUGS and thus might appeal to an applied researcher. Using suitable model choice criterion, the skew- t model provided the best fit to this data among other competing models. Since the data exhibit some degree of right-skewness in both components of its bivariate response, substantial improvement in model fit is observed by shifting away from the traditional normality assumptions.

Our approach motivated by a parametric class of skew-densities are relatively easy to implement and provides an interesting alternative to other computationally challenging semiparametric or fully nonparametric models [11, 36]. Although the wonderful memoir [18] provides many alternative expositions of skew-elliptical models, we specifically choose the skew-elliptical distributions starting from the skew-normal representation of [14] due to straightforward Bayesian analysis through hierarchical representations.

Our current analysis is focussed on exploring a 'clustered' cross-sectional periodontal progression data. However, often subjects who are brought in for periodontal assessments are subjected to randomized treatments and subsequent longitudinal followups. We plan to explore our class of SNI linear mixed models under these longitudinal framework. Also, periodontal progression is often believed to be associated with latent (within-mouth) spatial structures [37] and a diseased tooth seem to influence its neighboring tooth more than the non-neighbors while accounting periodontal progression. Methods to incorporate spatial dependencies in our class of SNI linear mixed models will be considered elsewhere.

Acknowledgements

The authors thank the Principal Investigator Dr. Jyotika Fernandez and the Center for Oral Health Research (COHR) at the Medical University of South Carolina for providing the data. The project described was supported by Award Numbers P20 RR017696 and R03 DE020114 from the US National Institute of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Dental & Craniofacial Research or the National Center for Research Resources, which are affiliated to the National Institutes of Health.

References

1. Position paper of the American Academy of Periodontology (2005). Epidemiology of Periodontal Disease, *Journal of Periodontology*, **76**, 1406-1419.
2. Oliver RC, Brown LJ and Loe H. Periodontal diseases in the United States population, *Journal of Periodontology* 1998; **69**: 269-278.
3. Darby ML and Walsh MM. *Dental Hygiene: Theory and Practice* (1st edn). W. B. Saunders Company, USA, 1995.
4. Matsuyama Y and Ohashi Y. Mixed models for bivariate response repeated measures data using Gibbs sampling, *Statistics in Medicine* 1997; **16**: 1587-1601.
5. Laird NM and Ware JH. Random effects models for longitudinal data. *Biometrics* 1982; **38**: 963-974.
6. Verbeke G and Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* 1996; **91**: 217-221.
7. Ghidry W, Lesaffre E and Eilers P. Smooth random effects distribution in a linear mixed model. *Biometrics* 2004; **60**: 945-953.
8. Lin TI. Robust mixture modeling using multivariate skew- t distributions. *Statistics and Computing* 2009; In Press, doi:10.1007/s11222-009-9128-9
9. Jara A, Quintana F and Martín E-S. Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Computational Statistics and Data Analysis* 2008; **52**: 5033-5045.
10. Azzalini A and Capitanio A. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B* 1999; **61**: 579-602.

11. Zhang D. and Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data, *Biometrics* 2001; **57**: 795-802.
12. Rosa GJM, Padovani CR and Gianola D. Robust linear mixed models with Normal/Independent distributions and Bayesian MCMC implementation. *Biometrical Journal* 2003; **45**: 573-590.
13. Azzalini A. and Capitanio A. Distributions generated by perturbation of symmetry with emphasis on the multivariate skew t-distribution. *Journal of the Royal Statistical Society Series B* 2003; **65**: 367-389.
14. Sahu SK, Dey DK and Branco MD. A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* 2003; **31**: 129-150.
15. Azzalini A. and Dalla-Valle, A. The multivariate skew-normal distribution. *Biometrika* 1996; **83**: 715-726.
16. Arellano-Valle RB, Branco MD and Genton MG. A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics* 2006; **34**: 581-601.
17. Azzalini A. The skew-normal distribution and related multivariate families (with discussion by Marc G. Genton and a rejoinder by the author). *Scandinavian Journal of Statistics* 2005; **32**: 159-200.
18. Genton M.G. *Skew-Elliptical Distributions and their Applications: A Journey Beyond Normality*, Chapman and Hall, Florida, USA, 2004.
19. Arellano-Valle RB, Bolfarine H and Lachos VH. Bayesian Inference for Skew-Normal Linear Mixed Models. *Journal of Applied Statistics* 2007; **34**: 663-682.
20. Lachos VH, Ghosh P and Arellano-Valle RB. Likelihood based inference for skew-normal/independent linear mixed model. *Statistica Sinica* 2010; **20**: 303-322.
21. Arellano-Valle RB and Azzalini A. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 2006; **33(3)**: 561-574.
22. Arellano-Valle RB and Genton MG. Fundamental skew distributions. *Journal of Multivariate Analysis* 2005; **96**: 93-116.
23. Fernandez JK, Salinas CF, London SD, Wiegand RE, Hill EG, Slate EH, Grewal JS, Werner P, Sanders JJ and Lopes-Virella MF. Prevalence of periodontal disease in Gullah African-American diabetics. *Journal of Dental Research* 2006; **85** (Special Issue A): 0997. URL (www.dentalresearch.org)
24. Taylor GW and Borgnakke WS. Periodontal disease: associations with diabetes, glycemic control and complications, *Oral Diseases* 2008; **14**: 191-203.
25. Reichard P, Nilsson BY and Rosenqvist U. The effect of long-term intensified insulin treatment on the development of microvascular complications of diabetes mellitus, *New England Journal of Medicine* 1993; **329**: 304-309.
26. Johnson-Spruill I, Hammond P, Davis B, McGee Z, Loudon D. Health of Gullah families in South Carolina with Type-2 diabetes, *The Diabeted Educator* 2009; **35**:117-123.
27. Lange KL and Sinsheimer JS. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* 1993; **2**: 175-198.
28. Spiegelhalter DJ, Best NG, Carlin BP and van der Linde A. Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society Ser. B* 2002; **64**: 583-639.
29. Gelfand AE, Ghosh SK. Model choice: A minimum posterior predictive loss approach, *Biometrika* 1998; **85**:1-11.
30. Gelman A., Carlin JB, Stern H and Rubin D. *Bayesian Data Analysis*, 2nd edition, Chapman & Hall/CRC, New York, 2004.
31. Hobert JP and Casella G. The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, *Journal of the American Statistical Association* 1996; **91**: 1461-1473.
32. Zhao Y, Staudenmayer J, Coull BA and Wand MP. General design Bayesian generalized linear mixed models, *Statistical Science* 2006; **21(1)**: 35-51.
33. Ghosh P, Bayes CL and Lachos VH. A robust Bayesian approach to null intercept measurement error model with application to dental data, *Computational Statistics and Data Analysis* 2009; **53(4)**: 1066-1079.
34. Natarajan R and Kass RE. Reference Bayesian methods for generalized linear mixed models, *Journal of the American Statistical Association* 2000; **95**: 227-237.
35. Gelman A. Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis* 2006; **1**: 515-533.
36. Müller P. and Rosner GL. A Bayesian population model with hierarchical mixture priors applied to blood count data, *Journal of the American Statistical Association* 1997; **92**: 1279-1292.
37. Reich BJ, Hodges JS and Carlin BP. Spatial analysis of periodontal data using conditionally autoregressive priors having two types of neighbor relations, *Journal of the American Statistical Association* 2007; **102**: 44-55.

APPENDIX: Outline of conditional posterior distributions

Under the full model as described in (13), given \mathbf{u} , the full conditional distribution of $\beta, \sigma_1^2, \sigma_2^2, \mathbf{D}, \lambda, \mathbf{b}_i, \mathbf{t}_i, i = 1, \dots, n$, are given as

$$\beta | \mathbf{b}, \mathbf{u}, \mathbf{t}, \sigma_1^2, \sigma_2^2, \mathbf{D}, \lambda \sim N_p(\mathbf{A}_\beta^{-1} \mathbf{a}_\beta, \mathbf{A}_\beta^{-1}), \quad (1)$$

where $\mathbf{A}_\beta = \mathbf{S}_\beta^{-1} + \sum_{i=1}^n u_i \mathbf{X}_i^\top \Sigma_i^{-1} \mathbf{X}_i$ and $\mathbf{a}_\beta = \mathbf{S}_\beta^{-1} \beta_0 + \sum_{i=1}^n u_i \mathbf{X}_i^\top \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{Z}_i \mathbf{b}_i)$;

$$\sigma_k^2 | \mathbf{b}, \mathbf{u}, \mathbf{t}, \beta, \mathbf{D}, \lambda \sim IG\left(\frac{N_k + \tau_k}{2}, \frac{T_k + \sum_{i=1}^n u_i \boldsymbol{\mu}_{ki}^\top \boldsymbol{\mu}_{ki}}{2}\right), \quad k = 1, 2, \quad (2)$$

where $N_1 = \sum_{i=1}^n m_i^P, N_2 = \sum_{i=1}^n m_i^C, \boldsymbol{\mu}_{1i} = \mathbf{y}_i^P - \mathbf{X}_i^P \beta^P - \mathbf{Z}_i^P \mathbf{b}_i^P$ and $\boldsymbol{\mu}_{2i} = \mathbf{y}_i^C - \mathbf{X}_i^C \beta^C - \mathbf{Z}_i^C \mathbf{b}_i^C$;

$$\lambda | \mathbf{b}, \mathbf{u}, \mathbf{t}, \beta, \sigma_1^2, \sigma_2^2, \mathbf{D} \sim N(\Sigma_\Delta^{-1} \boldsymbol{\mu}_\Delta, \Sigma_\Delta^{-1}), \quad (3)$$

where $\boldsymbol{\mu}_\Delta = \mathbf{S}_\lambda^{-1} \lambda_0 + \mathbf{D}^{-1} \sum_{i=1}^n u_i \mathbf{b}_i, \Sigma_\Delta = \mathbf{D}^{-1} \sum_{i=1}^n u_i \mathbf{t}_i^2 + \mathbf{S}_\lambda^{-1}$;

$$\mathbf{D} | \mathbf{b}, \mathbf{u}, \mathbf{t}, \beta, \sigma_1^2, \sigma_2^2, \lambda \sim IW_q\left(\left(\mathbf{T}_b^{-1} + \sum_{i=1}^n u_i (\mathbf{b}_i - \mathbf{A} \mathbf{t}_i)(\mathbf{b}_i - \mathbf{A} \mathbf{t}_i)^\top\right)^{-1}, \tau_b + n\right); \quad (4)$$

$$\mathbf{b}_i | u_i, \mathbf{t}_i, \beta, \sigma_1^2, \sigma_2^2, \mathbf{D}, \lambda \sim N_q(\mathbf{A}_{bi}^{-1} \mathbf{a}_i, u_i^{-1} \mathbf{A}_{bi}^{-1}), \quad (5)$$

where $\mathbf{A}_{bi} = (\mathbf{Z}_i^\top \Sigma_i^{-1} \mathbf{Z}_i + \mathbf{D}^{-1})$ and $\mathbf{a}_i = \mathbf{Z}_i^\top \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta) + \mathbf{t}_i \mathbf{D}^{-1} \mathbf{A}, i = 1, \dots, n$;

$$\mathbf{t}_i | \beta, \sigma_1^2, \sigma_2^2, \mathbf{D}, \lambda, \mathbf{b}_i, u_i \sim TN_q(\mathbf{A}_t^{-1} \mathbf{a}_{ti}, u_i^{-1} \mathbf{A}_t^{-1}; \mathbb{R}_+^q), \quad (6)$$

where $\mathbf{A}_t = (\mathbf{I}_q + \mathbf{A}^\top \mathbf{D}^{-1} \mathbf{A}), \mathbf{a}_{ti} = \mathbf{A}^\top \mathbf{D}^{-1} \mathbf{b}_i$.

Conditional posteriors for specific SNI cases follows.

- *The ST distribution*

Here ν is a scalar and when $\nu = 1$, the model is skew-Cauchy. We adopt a truncated exponential prior for ν of the form $E(\frac{\rho}{2}) \mathbb{I}_{(2, \infty)}$. The density of the conditional posterior distribution takes the form:

$$u_i | \boldsymbol{\theta}, \mathbf{b}, \mathbf{t}, \mathbf{y} \sim \text{Gamma}((n_i + 2q + \nu)/2; \nu/2 + C_i/2),$$

where $C_i = (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{b}_i)^\top \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{b}_i) + (\mathbf{b}_i - \mathbf{A} \mathbf{t}_i)^\top \mathbf{D}^{-1} (\mathbf{b}_i - \mathbf{A} \mathbf{t}_i) + \mathbf{t}_i^\top \mathbf{t}_i$.

The full conditional posterior density of ν is:

$$\pi(\nu | \boldsymbol{\theta}_{(-\nu)}, \mathbf{b}, \mathbf{u}, \mathbf{t}, \mathbf{y}) \propto \pi_1(\nu) \times \text{Gamma}\left(\frac{n\nu}{2} + 1, \frac{1}{2} \sum_{i=1}^n (u_i - \log u_i) + \rho\right) \mathbb{I}_{\{(2, \infty)\}},$$

where $\pi_1(\nu) = (2^{\nu/2} \Gamma(\nu/2))^{-n}$, which does not have a closed form but a Metropolis-Hastings or rejection sampling step can be utilized to obtain draws from ν .

- *The SSL distribution*

To ensure conjugacy, a $\text{Gamma}(a, b)$ distribution with small positive values of a and b ($b \ll a$) can be adopted as a prior for ν . The full conditional posterior density of each u_i is:

$$u_i | \boldsymbol{\theta}, \mathbf{b}, \mathbf{t}, \mathbf{y} \sim \text{Gamma}((n_i + 2q + 2\nu)/2; C_i/2) \mathbb{I}_{\{(0,1)\}}(\nu)$$

and

$$\nu | \boldsymbol{\theta}_{(-\nu)}, \mathbf{t}, \mathbf{b}, \mathbf{u}, \mathbf{y} \sim \text{Gamma}(n + a, b - \sum_{i=1}^n \log u_i) \mathbb{I}_{\{(0, \infty)\}}(\nu).$$

• *The SCN distribution*

A $U(0, 1)$ distribution is used as a prior for ν_1 , and an independent $\text{Beta}(a, b)$ is adopted as prior for ν_2 . The conditional distribution of each u_i is proportional to

$$\begin{cases} \nu_1 \nu_2^{\frac{n_i}{2} + q} \exp\{-\frac{1}{2} \nu_2 C_i\}, & \text{if } u_i = \nu_2 \\ (1 - \nu_1) \exp\{-\frac{1}{2} C_i\}, & \text{if } u_i = 1, \end{cases} \quad (7)$$

and the conditional probabilities are obtained readily by suitable normalization. The full conditional posterior of the proportion of outliers ν_1 is:

$$\nu_1 | \boldsymbol{\theta}_{(-\nu_1)}, \mathbf{b}, \mathbf{u}, \mathbf{t}, \mathbf{y}, \nu_2 \sim \text{Beta}\left(a + \frac{n - \sum_{i=1}^n u_i}{1 - \nu_2}; b + \frac{\sum_{i=1}^n u_i - n\nu_2}{1 - \nu_2}\right)$$

and the conditional posterior density of ν_2 is:

$$\pi(\nu_2 | \boldsymbol{\theta}, \mathbf{b}, \mathbf{u}, \mathbf{t}, \mathbf{y}, \nu_1) \propto \nu_1^{\left(\frac{n - \sum_{i=1}^n u_i}{1 - \nu_2}\right)} \times (1 - \nu_1)^{\left(\frac{\sum_{i=1}^n u_i - n\nu_2}{1 - \nu_2}\right)}.$$

Similarly, a Metropolis–Hastings proposal [12] can be used to update ν_2 .