Published in final edited form as:
Stat Modelling. 2011 February ; 11(1): 25–47. doi:10.1177/1471082X1001100103.

Bayesian latent variable models for spatially correlated tooth-level binary data in caries research

Y Zhang[1], **D Todem**[1], **K Kim**[2], and **E Lesaffre**[3]

[1]Department of Epidemiology, Michigan State University, USA

[2]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, USA

[3]Department of Biostatistics, Erasmus Medical Centre, Erasmus University Rotterdam, the Netherlands and L-Biostat, Catholic University of Leuven, Belgium

Abstract

Analysis of dental caries is traditionally based on aggregated scores, which are summaries of caries experience for each individual. A well-known example of such scores is the decayed, missing and filled teeth or tooth surfaces index introduced in the 1930s. Although these scores have improved our understanding of the pattern of dental caries, there are still some fundamental questions that remain unanswered. As an example, it is well believed among dentists that there are spatial symmetries in the mouth with respect to caries, but this has never been evaluated in a statistical sense. An answer to this question requires the analysis to be performed at subunits within the mouth, which necessitates the use of methods for correlated data. We propose a Bayesian generalized latent variable model coupled with an undirected graphical model to investigate the unique spatial distribution of tooth-level caries outcomes in the mouth. Data from the Signal Tandmobiel® study in Flanders, a dental longitudinal survey, are used to illustrate the methodology.

Keywords

hierarchical structure; latent variable; multivariate conditional autoregressive models; spatial correlation; undirected graphical Gaussian model

1 Introduction

Oral health conditions include a number of congenital or developmental anomalies, such as clefts and tumours. But the most common oral health problem is tooth decay. Despite being largely preventable, dental caries remains one of the most chronic diseases among children and adults. It is estimated that tooth decay is four times more prevalent than asthma in childhood. More than half of all adults over age 18 present early signs of this disease and about three out of four adults will develop the disease at some point in life. Although many dental studies provide detailed tooth-level data on caries activity, most analyses still rely on aggregated scores such as the DMF (decayed/missing/filled) index developed in the 1930s by Klein *et al.* (see e.g., Klein and Palmer, 1938). These scores summarize at mouth level caries information for each individual typically recorded at the tooth level or tooth surface level. They have therefore been instrumental in evaluating and comparing the risks for

Address for correspondence: David Todem, Division of Biostatistics, Department of Epidemiology, Michigan State University, East Lansing, MI 48823, USA. dtodem@epi.msu.edu.

dental caries among population groups. Despite these advances in the etiology of dental caries, there are still some fundamental questions regarding the spatial distribution of dental caries in the mouth that remain unanswered. As an example, it is widely believed among dentists that there are spatial symmetries in the mouth with respect to caries, but this has never been evaluated in a statistical sense (Todem, 2008). Another interesting intra-mouth hypothesis that needs to be statistically evaluated is whether different types of teeth (incisors, canines and molars) and tooth surfaces (facial, lingual, occlusal, mesial, distal and incisal surfaces) are not equally susceptible to dental caries. An answer to these questions requires the analysis to be performed at the tooth or tooth surface level. This then necessitates the use of methods for correlated data. In dental research, intra-mouth data present a unique set of challenges to statistical analysis (Leroux, 2006). These challenges include, but are not limited to, large cluster sizes (large number of teeth within a mouth); informative cluster sizes and multilevel data structures (e.g., teeth within a quadrant and quadrants within the mouth) which generate a very complex correlation structure.

Vanobbergen *et al.* (2007) have proposed a generalized estimating equation (GEE)-based model to answer some of the questions related to the spatial association of dental caries in the mouth. Specifically, these authors have used so-called alternating logistic regression (ALR) to model simultaneously the marginal expectation of each binary outcome and the pairwise associations between outcomes (Liang *et al.*, 1992). This model was used to circumvent some of the numerical problems related to the second-order moments—GEE2 model due to larger cluster sizes. Although the ALR model is known to be numerically stable, its convergence for unstructured second-order moments remains an issue for moderately larger cluster sizes. Indeed, when the cluster size is relatively large, one may encounter a numerical difficulty when an unstructured odds ratio model is imposed. This limitation was the reason for these authors to perform their analysis on a small subset of teeth (8 per mouth). When the ALR with unstructured odds ratios is attempted on the full dataset (20 teeth per mouth), the algorithm did not converge. It did so only for very structured odds ratio models, which may limit our ability and flexibility to study the spatial association in the mouth. Most importantly, one of the drawbacks of ALR models is that they focus only on a few moments of the distribution, which is a great limitation if estimating the entire distribution (all moments) of dental caries in the mouth is the primary interest. As an alternative to these GEE-based models, Zhu *et al.* (2005) have proposed a likelihood-based generalized latent variable model to analyse spatially correlated multivariate data in time. This class of models is an extension of the multivariate spatial latent variable model proposed by Wang and Wall (2003) and could be used for complex spatially correlated data with large cluster sizes and multilevel data structures. Although their models have interesting applications in environmental and ecological studies, their applicability to intra-oral outcomes is limited. For example, the position of teeth in the mouth has a unique spatial structure which renders the use of the Euclidean distance inappropriate. From a computational standpoint, their expectation–maximization (EM)-based method is computationally very intensive as it uses a Markov chain Monte Carlo (MCMC) approach to integrate out latent variables from the score vector and the Hessian matrix in the M step of the algorithm. Moreover, the numerical difficulty quickly escalates as the dimension of the latent variables increases. Finally, our practical experience with these models shows that the final Fisher information matrix is often singular as the global identifiability of the model remains a great challenge.

In this paper, we propose a class of likelihood-based generalized models with multilevel spatial latent vectors. The joint distribution of the data is specified indirectly using undirected graphical Gaussian models (UGGM) (Dempster, 1972; Giudici and Green, 1999) coupled with the independence assumption (Dawid, 1979). The resulting model is used to study the spatial association of dental caries at the quadrant level. The connection to the

Gaussian distribution in our modelling framework allows for an easy and interpretable assessment of this association. A Bayesian approach is used for estimation and inference. Other formulations of latent variable models have been proposed in the literature. For example, Sammel *et al.* (1997) have proposed a joint model for different outcomes within the framework of a generalized linear model with normal latent variables. Moustaki and Knott (2000) then extended this modelling framework to a class of generalized latent trait models. Both these approaches are EM based and are known to be computationally challenging for higher dimensional latent variables. Another important work is that of Dunson (2000) who proposed a model that allows the observed and latent variables to have distributions in the exponential family. All these existing models have nice applications in other fields but little applicability to intra-oral data in caries research due to their unique spatial structure. Moreover, despite the long history and wide applicability of undirected graphical models, surprisingly, Bayesian treatments of large undirected graphical models are virtually non-existent. Our model then serves as a prototype for analysing intra-oral-level data in caries research, while the search for other alternatives continues.

In Section 2, we present and describe the generalized latent variable model. Section 3 describes the Bayesian approach for model estimation and inference. In Section 4, the proposed model is applied to the example dental dataset to study the spatial association and symmetry of dental caries in the mouth. Finally, the usefulness and future extensions of the model are discussed in Section 5.

2 A generalized latent variable model

2.1 The unique spatial structure of dental data and key notations

Figure 1 is an illustration of the primary dentition. In Europe, these deciduous teeth are numbered with two digits, where the first digit represents the quadrant and the second digit the position of the tooth within a quadrant. As an example, tooth 55 represents the last molar of the fifth quadrant. Quadrants V and VI form the maxilla or the upper jaw, whereas quadrants VII and VIII form the mandible or the lower jaw. This pictorial representation of the primary dentition exhibits a two-level spatial association structure. The first-level spatial association structure is that among quadrants (V)–(VIII), whereas the second-level spatial association structure is that of teeth nested within a quadrant.

We will denote by y_{ijk} the binary outcome variable for tooth $k = 1, \cdots, K$, in quadrant $j = 1, \cdots, J$, of subject $i = 1, \cdots, n$. The outcome y_{ijk} takes value 1 if the tooth is decayed and 0 otherwise. For the primary dentition, $K = 5$ is the number of teeth in a quadrant and $J = 4$ is the number of quadrants in the mouth. The number n is the sample size. To ease the notation, we will use $j = 1$ for quadrant V, $j = 2$ for quadrant VI, $j = 3$ for quadrant VII and $j = 4$ for quadrant VIII. We will denote by $y_{ij} = (y_{ij1}, y_{ij2}, \cdots, y_{ijK})'$ the outcome vector for quadrant j of subject i. We will also denote by $y_i = (y_{i1}', \cdots, y_{iJ})'$ the outcome vector for subject i and by $y = (y_1', \cdots, y_n')'$ the responses for all subjects.

2.2 The undirected graphical model as the basic modelling framework

Tooth-level data have a unique hierarchical association structure marked by the quadrant level as well as the tooth nested within quadrant-level associations. Due to the lack of a natural and flexible distribution for spatially correlated binary data, it is conceptually convenient to assume that the intra-oral hierarchical correlation can be captured marginally by two unobserved latent variables. Specifically, for the quadrant level, we introduce the random vector $Q_i = (Q_{i1}, \cdots, Q_{iJ})'$ to tie the four quadrants for subject i together, with Q_{ij} being the scalar latent variable for quadrant j. Similarly, at the tooth level nested within a quadrant, we introduce a latent random vector $T_{ij} = (T_{ij1}, \cdots, T_{ijK})'$ to capture the spatial

association among the teeth in the same quadrant, with T_{ijk} being the scalar latent variable of tooth k of quadrant j. Among other assumptions, we assume that the quadrant-level observation vectors $\{y_{ij}, j = 1, \cdots, J\}$ are assumed conditionally independent given Q_i, $i = 1, \cdots, n$, and tooth-level data $\{y_{ijk}, k = 1, \cdots, K\}$ are also conditionally independent given T_{ij} and Q_i. With these conditional independence assumptions, the latent processes generate a structured correlation among the dependent observations that can be specified by an undirected graphical model. At any level of the hierarchy, this undirected graphical model views the latent variables under investigation as nodes and the associations between these units as paths. In Figures 2 and 3, graphs that list the possible nodes and paths for both the quadrant and tooth nested within quadrant levels are constructed. These graphs are constructed under the condition that two nodes that are not directly connected are said to be conditionally independent given the other nodes in the graph. Thus, the quadrant level requires six pairwise paths, whereas the tooth nested within quadrant level requires 10 pairwise paths in the undirected graphical model. These pairwise paths measure the conditional undirected association between the nodes as described in the brief review below. A full review on the topic can be found elsewhere (see, e.g., Dempster, 1972).

Let $\mathcal{G} = (\mathcal{V}\ \mathcal{E})$ be an undirected graph with vertex set $\mathcal{V} = \{1, \cdots, V\}$ and edge set $\mathcal{E} = \{e_{\upsilon\upsilon'} : \upsilon \neq \upsilon' = 1, \cdots, V\}$, where $e_{\upsilon\upsilon'} = 1$ or 0 according to whether vertices υ and υ', $1 \leq \upsilon \neq \upsilon' \leq V$, are directly connected in $\mathcal{G}$or not. In the undirected graphical model, the edge set describes the association structures of the vertex set. Random vectors are assigned to the edge set to represent the association strength between corresponding vertices. The undirected graphical Gaussian model consists of all V-dimensional normal distributions, say $X = \{X_1, \cdots, X_V\}$, with $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and precision matrix $\Omega = \Sigma^{-1} = \{\omega_{\upsilon\upsilon'}\}$, where Σ is unknown but satisfies the restrictions that when two nodes are not directly connected, they are said to be conditionally independent given the other nodes in the graph. This pairwise conditional independence can then be formulated using the Markov properties (Drton and Perlman, 2004) as $e_{\upsilon\upsilon'} = 0 \Leftrightarrow \tilde{\rho}_{\upsilon\upsilon'} = 0 \Leftrightarrow X_\upsilon \perp X_{\upsilon'} | \{X_u, u \neq \upsilon, u \neq \upsilon'\}$, where $\{\tilde{\rho}_{\upsilon\upsilon'}\}$ is the so-called partial correlation between the υth and υ'th vertex in the graph, defined as $\tilde{\rho}_{\upsilon\upsilon'} = -\omega_{\upsilon\upsilon'} / \sqrt{\omega_{\upsilon\upsilon} * \omega_{\upsilon'\upsilon'}}$. This partial correlation is a measurement of association between two nodes conditionally on the remaining ones. We will use these partial correlations to examine the spatial association of dental caries in the mouth.

2.3 The model

Following the work of Dunson (2000, 2007), we model the binary response y_{ijk} for the tooth k nested within quadrant j of subject i, by an exponential family distribution (McCullagh and Nelder, 1989) with the probability density function of a general form,

$p(y_{ijk}|\eta_{ijk}, \varphi) = \exp\left\{\frac{\eta_{ijk} y_{ijk} - b(\eta_{ijk})}{a_i(\varphi)} + c(y_{ijk}, \varphi)\right\}$, where φ is a scalar dispersion parameter and η_{ijk} is the canonical parameter. The mean and the variance of y_{ijk} are related to the latter parameter by $E(y_{ijk}|\eta_{ijk}, \varphi) = \frac{db(\eta_{ijk})}{d\eta_{ijk}}$ and $var(y_{ijk}|\eta_{ijk}, \varphi) = a_i(\varphi)\frac{d^2 b(\eta_{ijk})}{d\eta_{ijk}^2}$. The function $db(\eta_{ijk})/d\eta_{ijk}$ represents the canonical or natural link function. Other link functions can be used to relate the mean to the canonical parameter as $E(y_{ijk}|\eta_{ijk}, \varphi) = g^{-1}(\eta_{ijk})$, where $g(\cdot)$ is a monotone, differentiable and invertible function. The canonical parameter η_{ijk} is related to the tooth location in the mouth and latent variables Q_{ij} and T_{ijk} through the linear model, $\eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{ijk}$, where the parameters α, $\beta = (\beta_1, \cdots, \beta_J)'$ and $\gamma = (\gamma_{1(1)}, \ldots, \gamma_{K(1)}, \ldots, \gamma_{K(J)})'$ represent the regression coefficients for the fixed effects. Specifically, β_j represents the fixed effect for quadrant j and $\gamma_{k(j)}$ represents the fixed effect for tooth k within quadrant j, where $j = 1, \ldots, J$ and $k = 1, \ldots, K$. For these coefficients to be identifiable, we impose the following constraints; $\beta_J = 0$ and $\gamma_{K(j)} = 0$, $j = 1, \ldots, J$, which set

the last quadrant as the reference quadrant and the last molar within each quadrant as the reference tooth. These fixed effect terms globally capture the effects of tooth location in the mouth on the caries prevalence, whereas the latent variables Q_{ij} and T_{ijk}, respectively, describe how these global effects are modified for each quadrant and each tooth within a quadrant. Under the undirected Gaussian graphical model, we assume $Q_i \sim \mathcal{N}(\mu_Q, \Sigma_Q)$ and $T_{ij} \sim \mathcal{N}(\mu_{T_j}, \Sigma_{T_j})$. To ensure identifiability, we set $\mu_Q = 0$ and $\mu_{T_j} = 0$, $j = 1, \ldots, J$. We also assume that the quadrant-level spatial latent vectors $\{Q_i : i = 1, \ldots, n\}$ and the tooth-level spatial latent vectors $\{T_{ij} : j = 1, \ldots, J, i = 1, \ldots, n\}$ are mutually independent. This last assumption obviously imposes a structured correlation model among tooth-level responses. Indeed, two teeth from different quadrants are only tied by the association between their respective quadrant latent variables. Unstructured covariance matrices are often advocated whenever possible to avoid any misspecification of the association structure at the expense of efficiency. This approach is motivated by the idea that any constraint structure may limit the flexibility of the model to assess the spatial association. Although this is true in principle for a covariance matrix with fewer parameters such as Σ_Q, matrices with a considerable number of parameters such as the within-quadrant associations can be structured to improve efficiency. Indeed, if left completely unrestricted, the covariance matrix Σ_{T_j} requires 10 correlation parameters for each quadrant (40 parameters for the whole mouth) to capture the within-quadrant association. A restricted model such as the conditional autoregressive (CAR) model on T_{ij} may then be used to improve efficiency at the cost of creating a potential bias. This model is a natural choice as it reflects the configuration of the tooth locations within a quadrant. It defines the spatial covariance structure of the underlying tooth latent vector $T_{ij} = (T_{ij1}, \ldots, T_{ijK})'$ based on neighbourhood adjacencies. Hence, following the work of Cressie (1991), the CAR model is given by $(T_{ijk} | T_{ijk'}, k' \neq k, \tau^2_{jk}) \sim \mathcal{N}(\rho_j \sum_{k' \neq k} b_{kk'} T_{ijk'}, \tau^2_{jk})$, $k = 1, \ldots, K$, where K is the number of spatial nodes. Following Brook's (1964) lemma, the resulting joint density function of T_{ijk}, $k = 1, \ldots, K$, takes the form

$$f(T_{ij} | \tau^2_j) \propto \exp\left\{-\frac{1}{2} T'_{ij} D^{-1}_{\tau^2_j} (I - \rho_j B) T_{ij}\right\}, \tag{2.1}$$

where $\tau^2_j = (\tau^2_{j1}, \cdots, \tau^2_{jK})'$, B is $K \times K$ matrix with $B = (b_{kk'})$ and $b_{kk} = 0$ and $D_{\tau^2_j} = \mathrm{Diag}(\tau^2_j)$ is a $K \times K$ diagonal matrix with non-zero entries $\{\tau^2_{jk}, k = 1, \ldots, K\}$. From the classical CAR theory, it is clear that if the precision matrix $D^{-1}_{\tau^2_j}(I - \rho_j B)$ is positive definite, then (2.1) is a proper distribution. For this matrix to be symmetric, it requires $b_{kk'} \tau^2_{jk'} = b_{k'k} \tau^2_{jk}$, $\forall k, k' = 1, \ldots, K$. In addition, the precision matrix $D^{-1}_{\tau^2_j}(I - \rho_j B)$ is non-singular if $\rho_j \in (\lambda^{-1}_{max}, \lambda^{-1}_{min})$, where λ_{min}, λ_{max} are the smallest and largest eigenvalues of B, respectively. One often assumes that $D_{\tau^2_j} = \sigma^2_j M$, where M is a diagonal matrix with diagonal elements m_{kk} proportional to the conditional variance τ^2_{jk}, i.e., $\tau^2_{jk} = \sigma^2_j m_{kk}$. The term σ^2_j controls the overall variability and ρ_j represents the overall spatial association. The weights matrix $B = (b_{kk'})$ captures the spatial association between nodes k and k'. *GoeBUGS* (2004) sets $b_{kk'} = 1/n_k$, for $k \neq k'$ and $m_{kk} = 1/n_k$ where n_k is the number of nodes adjacent to node k. Naturally, for complete data, the matrices B and M do not depend on the quadrant-level index j, due to the spatial structure of teeth within any of the four quadrants. Under the above settings, the spatial latent vector T_{ij} follows a proper zero-centered distribution given by $T_{ij} \sim \mathcal{N}(\mathbf{0}, \sigma^2_j (I - \rho_j B)^{-1} M)$. This CAR covariance matrix $\sum^{CAR}_{T_j} = \sigma^2_j (I - \rho_j B)^{-1} M$ only

requires 2 (8 overall) parameters to model the intra-quadrant variability as opposed to 15 (60 overall) parameters for the unstructured covariance matrix Σ_{T_j}.

3 Estimation and inferences

3.1 Prior distributions specifications

In this section, prior distributions are specified for the regression parameter vector $\theta = (\alpha, \beta', \gamma')'$, the dispersion parameter φ and the precision matrices $\Omega_Q = \sum_Q^{-1}$ and $\Omega_{T_j} = \sum_{T_j}^{-1}$, $j = 1, \ldots, J$. To simplify the sampling from the posterior distributions, we choose hierarchically independent priors for these model parameters, given by $\pi(\theta, \Omega_Q, \Omega_{T_1}, \cdots, \Omega_{T_J}, \varphi) = \pi(\theta) \times \pi(\Omega_Q) \times \pi(\Omega_{T_1}) \times \cdots \times \pi(\Omega_{T_J}) \times \pi(\varphi)$, where $\pi(.)$ stands for the prior distribution function. Note that for the CAR model, $\pi(\Omega_{T_j})$ is replaced by $\pi(\sigma_j^2) \times \pi(\rho_j)$ in the above expression. Zhao *et al.* (2006), Zeger and Karim (1991) and Dunson (2000) all suggested diffuse prior distributions for the model parameters to ensure that the posterior is driven by data.

Following Bedrick *et al.* (1996) and Dunson (2000), we assign a normal prior to the regression parameter vector, $\theta \sim \mathcal{N}(\mu, \mathbf{F})$, where μ is a vector of location parameters and $\mathbf{F}$ is a covariance matrix. We set $\mu = \mathbf{0}$ and $\mathbf{F}$ to a diagonal matrix with very large entries to achieve a diffuse prior (Dunson, 2000). Conjugate diffuse Wishart priors (Dunson, 2000; O'Malley *et al.*, 2008) are specified for the precision matrices $\Omega_Q = \sum_Q^{-1}$ and $\Omega_{T_j} = \sum_{T_j}^{-1}$, $j = 1, \ldots, J$, where these matrices are all left unstructured. Specifically, for the quadrant-level precision matrix, we set $\Omega_Q \sim \textit{Wishart}(\upsilon_Q, \Lambda_Q)$, with the degrees of freedom υ_Q and the precision Λ_Q. In practice, the common diffuse Wishart prior is chosen by specifying $\Lambda_Q = \mathbf{I}_{\upsilon_Q}$, the identity matrix of order $\upsilon_Q = \textit{rank}(\Omega_Q) + 1$. This choice leads to a proper prior under which the marginal distribution of each correlation parameter is $U(1, 1)$ (O'Malley *et al.*, 2008). Similarly, independent Wishart processes are assigned as priors to unstructured precision matrices $\{\Omega_{T_j}, j = 1, \ldots J\}$ as $\Omega_{T_j} \sim \textit{Wishart}(\upsilon_{T_j}, \Lambda_{T_j})$, $j = 1, \cdots, J$, with the degrees of freedom $\upsilon_{T_j} = \textit{rank}(\Omega_{T_j}) + 1$ and the precision $\Lambda_{T_j} = \mathbf{I}_{\upsilon_{T_j}}$, the identity matrix of order υ_{T_j}. Note here that proper priors are assigned to the precision matrices Ω_Q and Ω_{T_j} as $\upsilon_Q > \textit{rank}(\Omega_Q)$ and $\upsilon_{T_j} > \textit{rank}(\Omega_{T_j})$, $j = 1, \ldots, J$. For structured matrices $\sum_{T_j}^{CAR}$, the prior distribution of $\{\sigma_j^2, \rho_j, j=1, \ldots, J\}$ needs to be specified. Specifically, independent inverse gamma distributions as proper conjugate priors (Dunson, 2000) are assigned to the overall variation parameters $\{\sigma_j^2, j=1, \ldots, J\}$ as $\sigma_j^2 \sim IG(\varepsilon, \varepsilon)$, $j = 1, \ldots, J$, where ε is a small positive number. However, this prior distribution is sharply peaked near zero and might distort posterior inferences. The uniform density on σ_j, with $j = 1, \ldots, J$, is a simple and well-behaved prior, which is equivalent to $p(\sigma_j^2) \propto \sigma_j^{-1}$, an inverse χ^2 density with degrees of freedom −1 (Gelman, 2006). It results in an improper prior distribution. An alternative would be to define the prior distribution in the range $(0, A)$ with positive large value of A, i.e., $\sigma_j \; U(0, A)$, $j = 1, \ldots, J$. For the overall spatial association parameters $\{\rho_j, j = 1, \ldots, J\}$, diffuse uniform priors (*GeoBUGS:* Thomas *et al.*, 2004) are assigned as $\rho_j \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1})$, $j = 1, \ldots, J$, where λ_{min} and λ_{max} are the smallest and largest eigenvalues of the matrix B defined in Section 2.3. Finally, a gamma prior is assigned for the dispersion parameter φ as $\varphi \sim \text{gamma}(\nu, \phi)$.

3.2 Posterior distributions

MCMC techniques are used for the posterior computations of the model parameters which can be obtained in a standard way (Zeger and Karim, 1991; Dunson, 2000). We assume that

the outcome variables are conditionally independent given the latent vectors, $Q=(Q_1', \cdots, Q_n')'$ with $Q_i = (Q_{i1}, \cdots, Q_{iJ})'$ and $T=(T_1', \cdots, T_n')'$ with $T_i=(T_{i1}', \cdots, T_{iJ}')'$ and $T_{ij} = (T_{ij1}, \cdots, T_{ijK})'$. The joint probability density of y conditional on the set of latent vectors Q and T, the regression vector θ and the dispersion parameter φ is then given by $p(y|Q, T, \theta, \varphi)=\prod_{i=1}^{n}\prod_{j=1}^{J}\prod_{k=1}^{K} p(y_{ijk}|\eta_{ijk}, \varphi)$. Given the precision matrixes Ω_Q and $\{\Omega_{T_j}, j = 1, \ldots, J\}$, the joint posterior distribution for the regression parameter vector θ, the dispersion parameter φ and latent vectors Q and T, is given by

$$\begin{aligned} p(\theta, Q, T, \varphi|y) &\propto p(y|\theta, Q, T, \varphi)\pi(\theta, Q, T, \varphi) \\ &\propto \exp\left\{\sum_{ijk}\log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \frac{1}{2}\theta'\mathbf{F}^{-1}\theta - \frac{1}{2}\sum_{i=1}^{n} Q_i'\Omega_Q Q_i\right\} \\ &\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J} T_{ij}'\Omega_{T_i}T_{ij} + (\nu - 1)\log(\varphi) - \phi\varphi\right\}, \end{aligned} \tag{3.1}$$

where $\log\{p(y_{ijk}|\eta_{ijk}, \varphi)\}=\frac{\eta_{ijk}y_{ijk} - b(\eta_{ijk})}{a_i(\varphi)} + c(y_{ijk}, \varphi)$ and $\sum_{ijk}$ is shorthand notation for $\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K}$. If the MCMC algorithm is a Gibbs sampler, the full conditional distribution of each of the unknowns in (3.1) can be obtained in a standard way (see, e.g., Zeger and Karim, 1991; Dunson, 2000). For the fixed effect θ, the full conditional distribution is given by

$$p(\theta|Q, T, \varphi, y) \propto \exp\left\{\sum_{ijk}\log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \frac{1}{2}\theta'\mathbf{F}^{-1}\theta\right\}. \tag{3.2}$$

The full conditional distribution for the Gaussian spatial latent vector Q is

$$p(Q|\theta, T, \varphi, y) \propto \exp\left\{\sum_{ijk}\log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \frac{1}{2}\sum_{i=1}^{n} Q_i'\Omega_Q Q_i\right\}. \tag{3.3}$$

The full conditional distribution for the Gaussian spatial latent vectors T is

$$p(T|\theta, Q, \varphi, y) \propto \exp\left\{\sum_{ijk}\log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J} T_{ij}'\Omega_{T_j}T_{ij}\right\}. \tag{3.4}$$

The full conditional distribution of the precision matrix Ω_Q is given by

$$p(\Omega_Q|\theta, Q, T, \varphi, y)=Wishart\left(\upsilon_Q+n, \Lambda_Q+\sum_{i=1}^{n} Q_i Q_i'\right). \tag{3.5}$$

Similarly for the unstructured matrices $\{\Omega_{T_j}, j = 1, \ldots, J\}$, the posterior distributions are given as follows

$$p(\Omega_{T_i}|\theta, Q, T, \varphi, y) = Wishart\left(\upsilon_{T_j}+n, \Lambda_{T_j}+\sum_{i=1}^{n} T_{ij}T_{ij}'\right). \quad (3.6)$$

For the structured CAR model at the tooth level, the posterior distribution of the variances σ_j^2 and correlation ρ_j, $j = 1, \ldots, J$, are, respectively, given by

$$p(\sigma_j^2|\theta, Q, T, \varphi, y) \propto \sigma_j^{2(\varepsilon-1)} \exp\left\{\sum_{ijk} \log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \sigma_j^2\varepsilon\right\}, \quad (3.7)$$

and, if $\sigma_j \propto U(0, A)$, we have

$$p(\sigma_j|\theta, Q, T, \varphi, y) \propto \exp\left\{\sum_{ijk} \log\{p(y_{ijk}|\eta_{ijk}, \varphi)\}\right\} \delta_{(0,A)}(\sigma_j), \quad (3.8)$$

and,

$$p(\rho_j|\theta, Q, T, \varphi, y) \propto \exp\left\{\sum_{ijk} \log\{p(y_{ijk}|\eta_{ijk}, \varphi)\}\right\} \delta_{(\lambda_{max}^{-1}, \lambda_{min}^{-1})}(\rho_j), \quad (3.9)$$

where the indicator function $\delta_A(x) = 1$ if $x \in A$ and 0 otherwise. Finally, the full conditional distribution of the dispersion parameter is

$$p(\varphi|\theta, Q, T, y) \propto \varphi^{\nu-1} \exp\left\{\sum_{ijk} \log\{p(y_{ijk}|\eta_{ijk}, \varphi)\} - \varphi\phi\right\}. \quad (3.10)$$

The algorithm proceeds by sampling θ, Q, T, φ, Ω_Q and Ω_{T_j}, $j = 1, \cdots, J$, from the conditional distributions proportional to expressions (3.2)–(3.10). Note that all the posterior distributions, except those of $\{\rho_j, j = 1, \ldots, J\}$, are proper as they are based on proper conjugate priors. The uniform priors for the overall spatial parameters are not conjugate, which may lead to improper posterior distributions. A simple technique to verify that the posterior distribution of parameters is proper is to check that the posterior distribution is proper for reduced independent data (O'Brien and Dunson, 2004). These reduced data can be constructed by discarding all but a single outcome per subject, leading to independent outcomes. Since the covariance structures do not appear in the reduced data likelihood and also the support for the spatial association parameters is bounded, i.e., $\rho_j \in (\lambda_{max}^{-1}, \lambda_{min}^{-1})$, $j = 1, \ldots, J$, the posterior distributions of the spatial association parameters $\{\rho_j, j = 1, \ldots, J\}$ are thus proper.

3.3 Inferences on assessing spatial association at quadrant level

The spatial association that we are investigating in this paper refers to the association of dental caries outcomes at quadrant levels. Lesaffre and Bogaerts (2006) have shown using a GEE-based ALR model that each molar in the left quadrant has the strongest association at the quadrant level with the corresponding molar in the right quadrant. Our analysis goes beyond these two teeth to investigate whether this property applies to all teeth (incisors, canines and molars). To evaluate these hypotheses, we consider the matrix Σ_Q that captures

the spatial association among quadrants on the latent scale. No structure was imposed on this matrix to allow the flexibility of the model to assess this spatial association. For this, let $\tilde{\rho}_{jj'}$ be the partial correlation coefficient that captures the association between the latent variables of quadrants j and j', fixing the other quadrants. To asses the quadrant-level spatial association, we consider the equal weight average correlations for the 'left–right', 'up–down' and 'across' associations given, respectively, by

$\tilde{\rho}_{LR}=\frac{1}{2}(\tilde{\rho}_{12}+\tilde{\rho}_{34}), \tilde{\rho}_{UD}=\frac{1}{2}(\tilde{\rho}_{14}+\tilde{\rho}_{23})\ \tilde{\rho}_{A}=\frac{1}{2}(\tilde{\rho}_{13}+\tilde{\rho}_{24})$. The statistical hypothesis testing about the overall spatial association in terms of 'left–right' versus 'up–down', 'left–right' versus 'across' and 'across' versus 'up–down' can then be formulated as *(i)* left–right versus up–down, $H_0 : \tilde{\rho}_{LR} = \tilde{\rho}_{UD}$ *vs.* $H_a : \tilde{\rho}_{LR} \neq \tilde{\rho}_{UD}$; *(ii)* left–right versus across, $H_0 : \tilde{\rho}_{LR} = \tilde{\rho}_{A}$ *vs.* $H_a : \tilde{\rho}_{LR} \neq \tilde{\rho}_{A}$ and *(iii)* across versus up–down, $H_0 : \tilde{\rho}_{A} = \tilde{\rho}_{UD}$ *vs.* $H_a : \tilde{\rho}_{A} \neq \tilde{\rho}_{UD}$. Alternatively, tests for all 12 pairwise comparisons can also be performed. As an example, the left–right association for the upper jaw can be compared to the left–right association of the lower jaw. This is equivalent to the hypothesis $H_0 : \tilde{\rho}_{12} = \tilde{\rho}_{34}$ *vs.* $H_a : \tilde{\rho}_{12} \neq \tilde{\rho}_{34}$.

Since multiple tests are being performed, it is necessary to construct simultaneous credible regions (Besag *et al.*, 1995) to control the type S error rate, a similar concept to the type I error rate in the frequentist's framework (Gelman and Tuerlincks, 2000). The $100K_o/M$% simultaneous credible regions for overall association differences are then based on order statistics (Besag *et al.*, 1995) $\{[(\tilde{\rho}_I - \tilde{\rho}_{II})^{[M+1-t^*]}, (\tilde{\rho}_I - \tilde{\rho}_{II})^{[t^*]}] : (I, II) \in \{(\text{'LR','UD'}), (\text{'LR', 'A'}), (\text{'A','UD'})\}\}$, where t^* is given by $\min\{t : \#\{(\tilde{\rho}_I - \tilde{\rho}_{II})^{[M+1-t^*]} \leq (\tilde{\rho}_I - \tilde{\rho}_{II})^{(t)} \leq (\tilde{\rho}_I - \tilde{\rho}_{II})^{[t^*]}\} \geq K_o\}$ with $\{(\tilde{\rho}_I - \tilde{\rho}_{II})^{(t)}, t = 1, \ldots, M\}$ being the posterior samples of $\tilde{\rho}_I - \tilde{\rho}_{II}$ and $[x]$ being the integer part of a real number x. Similarly, the $100K_o/M$% simultaneous credible regions for specific pairwise spatial association differences are given by $\{[(\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{[M+1-t^*]}, (\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{[t^*]}] : u \neq u', \upsilon \neq \upsilon'\}$, where $t^* = \min\{t : \#\{(\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{[M+1-t^*]} \leq (\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{(t)} \leq (\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{[t^*]}\} \geq K_o\}$ and $\{(\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'})^{(t)}, t = 1, \ldots, M\}$ are the posterior samples of $\tilde{\rho}_{uu'} - \tilde{\rho}_{\upsilon\upsilon'}$.

4 A numerical example

4.1 The Tandmobiel® study

The Signal Tandmobiel® project is a longitudinal study conducted with the support of three Flemish universities (Catholic University of Leuven, University of Gent and Free University of Brussels), the Association of Flemish Dentists and the Flemish Scientific Association for Youth Health and industry (LeverElida). The project took place in a selected number of schools of the three different existing educational systems in the five Flemish provinces of Belgium. There were three distinct components in the study: *(i)* screening schoolchildren for oral health conditions, *(ii)* giving dental health information and education and *(iii)* making a scientific contribution. For the last component, the project collected epidemiological data on the oral health condition of 7- to 12-year-old children from Flanders longitudinally over a 6-year period (1996–2001). The study sample represented about 7% of the corresponding Flemish population. The target population consisted of children born in 1989 and attending primary education in Flanders. Children were followed during the period they spent in primary school, meaning that at the start of the study, the children were 6–7 years of age, and at the end they were 12–13 years of age. Equal numbers of boys and girls were examined. The project is unique according to the methodology used, the large size of the cohorts, the longitudinal design with the cross-sectional supplements and the high quality of the data.

The Signal Tandmobiel® study provides detailed tooth-level and tooth surface level data that can be used to assess the spatial distribution of dental caries in the mouth. The collected outcome data include information on *(i)* the caries experience of all deciduous and

permanent teeth, including the tooth surfaces affected by caries, *(ii)* the plaque index measured at six locations in the mouth and *(iii)* the gingival condition. For the present analysis, however, the outcome under investigation is the caries status of the primary dentition dichotomized as being decayed or not. Each child had at most 20 binary tooth-level outcomes. The covariate of interest in our analysis was the position of the tooth in the mouth for each child. In addition, our analysis was restricted to 4351 schoolchildren who attained their seventh birthday anniversary during the course of the study. A more detailed description of the Signal Tandmobiel® project can be found elsewhere (see, e.g., Vanobbergen *et al.*, 2000).

4.2 A primary analysis

The frequency table for the prevalence of caries experience in the deciduous dentition is shown in Table 1, for the 7-year-old children. The descriptive statistics suggest a left–right spatial symmetrical pattern in terms of caries frequencies. Specifically, the paired teeth $5j - 6j$, $j = 1, \cdots, 5$, have roughly the same caries prevalence estimates. Similarly, findings are observed for paired teeth $8j - 7j$, $j = 1, \cdots, 5$. One, however, should be careful in interpreting these results as they only represent population-averaged frequencies, which may not be relevant in understanding the spatial distribution in the mouth. Indeed, at the mouth level, molars 55 and 65, as an example, may still achieve the same caries prevalence estimates by being negatively correlated. An appropriate method to study the spatial distribution of dental caries in the mouth then needs additional parameters that capture the within-mouth associations, or more that characterize the full spatial distribution.

A sensible approach to study the spatial clustering in the mouth is to consider a GEE-based ALR model with an unstructured design matrix for the odds ratio model. Structuring the odds ratios may limit the flexibility of the model to study the spatial association of dental caries in the mouth. The complexity of this approach is characterized by the large number of odds ratio parameters to be estimated, however. Indeed, a tooth-level analysis of caries outcomes using a GEE-based ALR model with an unstructured odds ratio model requires $190 = 20 \times (20 - 1)/2$ pairwise odds ratios to be estimated for the primary dentition. (Note here that 20 is the maximum number of teeth per mouth.) This then forced Vanobbergen *et al.* (2007) to perform their analysis using the subset of teeth {54 55; 64 65; 74 75; 84 85}. This subset corresponds to choosing only the two molars for each quadrant in the primary dentition. The result of their analysis in terms of pairwise odds ratio of caries experience under an unstructured ALR model is presented in Table 2. The result shows that for the first and second molars, the left–right association appears to be the strongest association for both the mandible and the maxilla. This finding gives a strong indication of a left–right spatial association in the primary dentition. Further analysis by these authors suggest that decayed teeth of discordant contralateral pairs tend to aggregate on the right or the left side of the subject's mouth more than would be expected by chance alone (Vanobbergen *et al.*, 2007).

4.3 The Bayesian analysis

4.3.1 Model specification—Now we use the proposed Bayesian latent variable model to study the spatial association of dental caries at the quadrant level. We first specify all the model components and related notations. As the outcomes under investigation are binary, we assume $a_i(\varphi) = \varphi = 1$, $b(\eta_{ijk}) = \log\{1 + \exp(\eta_{ijk})\}$, $c(y_{ijk}, \varphi) = 0$, $E(y_{ijk}|\eta_{ijk}) = \{1 + \exp(-\eta_{ijk})\}^{-1}$ and $g(x) = \log\{x/(1 - x)\}$, for $k = 1, \ldots, 5$, $j = 1, \ldots, 4$, $i = 1, \ldots, n$. Hence, we have $\log p(y_{ijk}|\eta_{ijk}) = \eta_{ijk}y_{ijk} - \log\{1+\exp(\eta_{ijk})\}$. The canonical parameter $\{\eta_{ijk}, k = 1, \ldots, 5, j = 1, \ldots, 4, i = 1, \ldots, n\}$ is modelled as $\eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{ijk}$. Hence, the parameters of interest in the observation model are $\theta = (\alpha, \beta', \gamma')'$ and ξ the vector of unique entries in the matrices $\{\Sigma_Q, \Sigma_{T1}, \ldots, \Sigma_{TJ}\}$. Priors for parameters of interest are given by diffuse proper conjugate priors, which give comparable results to those of frequentist's as the sample size

increases. More specifically, the priors are given by $\alpha, \beta_j, \gamma_{k(j)} \sim \mathcal{N}(0, 1000)\ \forall k = 1, \ldots, K, j = 1, \ldots, J$, with constraints $\beta_J = 0$ and $\gamma_{K(j)} = 0$, $j = 1, \ldots, J$, for identifiability of the observation model. For the priors of the precision matrix Ω_Q, we propose a scaled Wishart distribution as conjugate proper priors (O'Malley *et al.*, 2008), $\Omega_Q \sim Wishart(4 + 1, \mathbf{I}_4)$, where $\mathbf{I}_4$ is the identity matrix of order 4. Similarly for the unstructured precision matrix Ω_{T_j}, we propose $\Omega_{T_j} \sim Wishart(5+1, \mathbf{I}_5)$, $\forall j = 1, \ldots, 4$, where $\mathbf{I}_5$ is the identity matrix of order 5. And for its CAR counterpart, we propose, for $j = 1, \ldots, 4$,

$\sigma_j^{-2} \sim \text{IG}(0.001, 0.001)$ or $\sigma_j^{-2} \sim \text{IG}(0.5, 0.005)$ or $\sigma_j \sim \text{U}(0, 1000)$ and $\rho_j \sim U(\lambda_{max}^{-1}, \lambda_{min}^{-1})$.

To construct 95% simultaneous credible regions, we use 11 000 MCMC iterations with 1000 burn-in samples, i.e., $M = 10\,000$ and $K_o = 9500$. The 95% simultaneous credible regions are more conservative simultaneous confidence regions than frequentist's for the multiple hypothesis statements since they have a type I error rate between 0% and 2.5% (Gelman and Tuerlincks, 2000). The generalized latent variable models are implemented in *WinBUGS* (Spiegelhalter *et al.*, 2002), using diffuse priors for the parameters of interest as described above. The DIC adjusted for missing data (Celeux *et al.*, 2006) was used for the choice of the appropriate covariance model.

4.3.2 Results—We recall that the proposed generalized latent variable model defines a hierarchical association structure, the quadrant-level association as well as the within-quadrant association. At the quadrant level, no structure was assumed on the precision matrix as any restriction may limit the flexibility of the model to assess the quadrant spatial association. At the tooth level, however, a CAR model was assumed and assessed against the unstructured model. To assess the sensitivity of our results to the choice of a prior for the jth quadrant overall spatial variability parameter σ_j, $j = 1, \ldots, 4$, three different prior specifications were considered: $\sigma_j^{-2} \sim IG(0.001, 0.001)$ (Spiegelhalter *et al.*, 2003), $\sigma_j^2 \sim IG(0.5, 0.005)$ (*GeoBUGS*: Thomas *et al.*, 2004) or $\sigma_j \sim U(0, 1000)$ (Gelman 2003, 2006). The posterior means and credible intervals for the fixed effects parameters appear to be non-sensitive to the choice of the prior imposed on the overall spatial variability parameters. Similar findings were obtained for the posterior means and credible intervals of spatial association parameters. We only present estimation and inference results based on the uniform prior $\sigma_j \sim U(0, 1000)$ and the *Wishart* prior for the unstructured Σ of the within-quadrant spatial association.

Based on the results from the unstructured model and CAR model at the tooth level, the posterior inferences about the spatial associations are similar. The quadrant-level spatial association was assessed through simultaneous credible regions for differences of interest. The results of the analysis are shown in Tables 3 and 4, respectively, for the overall and pairwise differences under different models with different priors for the overall spatial precision parameters. Using the DIC, our inference results suggest that the model with unstructured precision matrices at the tooth level appears to provide the best fit. The DIC model selection was based on the rule of thumb that if the difference between the DICs of two different models is more than 10, then the model with the smaller DIC provides the most parsimonious fit. Note, however, that the DIC is optimal only when the model selection is simultaneous through the reversible jump MCMC (Green, 1995) or the birth and death MCMC (Stephens, 2000). Hence, based on the chosen model, the conclusions of the hypothesis testing about both overall and specific spatial associations among quadrants are as follows: *(i)* left–right spatial association is the strongest, which is shown in terms of 95% simultaneous credible intervals of the differences between left–right and across and the differences between left–right and up–down having lower bounds that are all positive (see Table 4); *(ii)* the difference of spatial associations between across and up–down is not significant at type I error rate between 0% and 2.5% (Gelman and Tuerlincks, 2000), since

the 95% simultaneous credible intervals of the differences between across and up–down include zero. It is not surprising that the left and right quadrants have the strongest quadrant-level association as these quadrants constitute the same jaw, maxilla or mandible. At the tooth level, the last molars exhibit the highest rates of decay in primary dentition for the mandible. Specifically, all the 95% credible intervals of the parameters $\gamma_{1(3)}$, $\gamma_{2(3)}$ and $\gamma_{3(3)}$ for the right quadrant and $\gamma_{1(4)}$, $\gamma_{2(4)}$ and $\gamma_{3(4)}$ for the left quadrant do not include zero. For the maxilla, however, the incisors appear to have the same caries decay rates as the last molars. The 95% credible intervals of the parameters $\gamma_{1(1)}$ and $\gamma_{2(1)}$ include zero. These findings are consistent with existing literature.

5 Discussion

In this paper, we proposed a class of generalized latent variable models for spatially correlated binary data with a multilevel dependence structure. This regression model belongs to the class of cluster-specific approaches for modelling correlated data as opposed to population-averaged approaches of which the most common one is the GEE-based ALR model. The ALR models focus on the marginal expectation of the response and the second-order associations through pairwise odds ratios. In contrast, the cluster-specific approaches such as the proposed generalized latent variable model, focus on the conditional expectation given the quadrant- and tooth-specific latent variables. Most importantly, unlike ALR models which focus only on a few moments of the distribution, the model proposed in this paper attempts to specify the entire distribution of the data via undirected graphical Gaussian models coupled with the conditional independence. The connection of the latent variable model to the multivariate normal distribution then allows for a flexible modelling of the association structure both at the quadrant and at the tooth levels. Specifically, at the quadrant level, we propose a multivariate normal distribution with an unstructured variance–covariance matrix for which individual entries are used to assess the spatial association among quadrants. At the tooth level, unstructured and structured CAR variance–covariance matrices were considered and compared. The CAR model takes advantage of the natural configuration of teeth within a quadrant although it did not provide a good fit for our data.

The Bayesian framework is a natural approach for estimation and inferences about the proposed generalized latent variable model. Diffuse priors were considered for estimation to ensure that the posterior is driven by the data. We understand that prior distributions may be used to incorporate prior information into inference when this is available. To our knowledge, consistent and reliable prior information on the spatial association of tooth-level caries data is readily not available. We have used the MCMC method to implement the model using *WinBUGS*. The MCMC approach is particularly useful in models where some structure is imposed on the covariance matrix as the Gibbs sampler simulates conditional distributions sequentially and so does not have to adjust to the structure of the model. In addition, MCMC methods yield inferences based upon samples from the full posterior distribution. Therefore, they can produce exact inference in cases where the likelihood-based methods only yield approximations. Our model yields valid inferences under the missing at random (MAR) mechanism as defined by Little and Rubin (1987). Unlike most statistical packages where incomplete cases are removed from the analysis, *WinBUGS* generates a sample to replace missing responses from the posterior distribution of the response variable under the more relaxed MAR assumption.

Our model assumes that the two spatial latent vectors are normally distributed and independent. This assumption may limit the ability of the model to capture unobserved heterogeneities in the data. Most importantly, there is a risk that the assumed distribution may not represent the true underlying model, which obviously would lead to bias. It would be interesting to introduce non-Gaussian latent processes to model the underlying spatial

dependence among quadrants and teeth nested within the quadrant, which would lead to a richer class of models. It would also be interesting to use a spatiotemporal model to assess how the left–right association highlighted in this paper evolves over time in a longitudinal study. These and other generalizations of the model are the subject of further research.

Acknowledgments

This study was supported with funding by the small grant mechanism from the National Institute on Dental and Craniofacial Research grant 1R03 DE 016618. The authors thank collaborators at K.U. Leuven, Belgium for providing us with the Tandmobiel dataset that motivated and illustrated our methodology. We would like to thank Kris Bogaerts for providing us with the pictorial representation and the European numbering of the primary dentition. The fourth author acknowledges partial support from Research Grant OT/05/60, Catholic University of Leuven. Data collection was supported by Unilever, Belgium. The Signal Tandmobiel® project comprises following partners: D. Declerck (Dental School, Catholic University of Leuven), L. Martens (Dental School, University Ghent), J. Vanobbergen (Oral Health Promotion and Prevention, Flemish Dental Association), P. Bottenberg (Dental School, University Brussels), E. Lesaffre (Biostatistical Centre, Catholic University of Leuven) and K. Hoppenbrouwers (Youth Health Department, Catholic University of Leuven; Flemish Association for Youth Health Care).We also acknowledge support from the Interuniversity Attraction Poles Program P5/24—Belgian State—Federal Office for Scientific, Technical and Cultural Affairs.

Appendix: Bayesian model selection using the DIC

The deviance information criterion (DIC) can be used to compare models within the Bayesian framework. The criterion (Spiegelhalter *et al.*, 2002) is a hierarchical modelling selection criterion that can be viewed as a generalization of the Akaike's information criterion (Akaike, 1973) and the Bayesian information criterion (Kass and Raftery, 1995). It is particularly useful in selecting Bayesian models when the posterior distributions of parameters have been obtained by MCMC simulations. The DIC statistic as a measure of model complexity and goodness of fit is defined as $DIC=\overline{D(\psi)}+pD$, where ψ is the model parameter vector and $\overline{D(\psi)}$ is the posterior mean of the deviance $D(\psi) = -2\log(f(y|\psi))$. The latter term is a measurement of goodness of fit of the model under investigation to the observed data. Let $D(\bar{\psi})$ be the deviance evaluated at the posterior mean of ψ and let $pD=\overline{D(\psi)} - D(\bar{\psi})$ denote the effective number of parameters in the model, a penalty for model complexity. It is well established that the quantities $\overline{D(\psi)}$ and $D(\bar{\psi})$ can be obtained from an MCMC simulation chain.

In order to deal with the latent vectors, one needs to compute the complete DIC in Celeux *et al.* (2006). Let $E_\psi[\psi|y, q, t]$ denote the posterior mean of ψ, based on the complete data $(y', q', t')'$, where $(q', t')'$ is the realization of the spatial latent vectors $(Q', T')'$. The DIC for the complete data model is $DIC(y, q, t) = -4E_\psi[\log(f(y, q, t|\psi))|y, q, t] + 2\log(f(y, q, t|E_\psi[\psi|y, q, t]))$. As in the EM algorithm, we can then integrate Q and T out from the expression above to get $DIC = E_{Q,T}[DIC(y, Q, T)|y] = -4E_{\psi,Q,T}[\log(f(y, Q, T|\psi))|y] + 2E_{Q,T}[\log(f(y, Q, T|E_\psi[\psi|y, Q, T]))|y]$. All the integrations can be obtained using Monte Carlo integration approximation from the MCMC posterior samples in the code file of *WinBUGS*.

The above DIC can be modified to account for missing data (Celeux *et al.*, 2006). For this, we define by $r = (r_1, \cdots, r_n)'$ the missing data indicator vector for the whole dataset, where $r_i = (r_{i1}, \cdots, r_{iJ})'$ and $r_{ij} = (r_{ij1}, \cdots, r_{ijK})'$ with $r_{ijk} = 1$ if tooth-level data y_{ijk} is observed and 0 otherwise, $i = 1, \cdots n$, $j = 1, \cdots J$, $k = 1, \cdots K$. The deviance function based on the observed data y_{obs} and observed missing observation vector r are given as

$$D(o)=-2\log\{f(y_{obs}, r|\psi, \varrho)\}=-2\log\left\{\int f(y_{obs}, y_{mis}|r;\psi)f(r|y_{obs}, y_{mis}, \varrho)dy_{mis}\right\}, \quad \text{(A.1)}$$

where y_{mis} is the vector of missing data and $o = (\psi', \varrho')'$ with ϱ being the parameter for the distribution of $f(r|y_{obs}, y_{mis}, \varrho)$. Pettitt *et al.* (2006) gave an approximation to this deviance as $D(o) = -2 \log \{f(y_{obs}, \hat{y}_{mis}|r;\psi)\, f(r|y_{obs}, \hat{y}_{mis}, \varrho)\}$ where $\hat{y}_{mis}$ is the posterior predictor for the missing data y_{mis}. Combining the last two equations, we have an adjusted DIC for Bayesian generalized latent variable models for missing data in the form

$$\begin{aligned} DIC = E_{Q,T}[\, & DIC(y_{obs}, \\ & \hat{y}_{mis}, r, Q, T)|y_{obs}, \\ & r] = -4E_{\psi,\varrho,Q,T}[\log(f(y_{obs}, \\ & \hat{y}_{mis}, Q, T|\psi, \\ & r, \varrho) f(r|y_{obs}, \\ & \hat{y}_{mis}, \varrho))|y_{obs}, \\ & r] + 2E_{Q,T}[\log(f(y_{obs}, \hat{y}_{mis}, Q, T|E_{\psi,\varrho}[\psi, \\ & \varrho|y_{obs}, \\ & r, Q, T]) \times f(r|y_{obs}, \\ & \hat{y}_{mis}, E_{\varrho}[\varrho|y_{obs}, \\ & r]))|y_{obs}, r]. \end{aligned} \tag{A.2}$$

References

Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN.; Csáki, F., editors. Second international symposium on information theory; Akademiai Kiado; Budapest. 1973. p. 276-281.

Bedrick EJ, Christensen R, Jonson W. A new perspective on priors for generalized linear models. Journal of the American Statistical Association. 1996; 91:1450–1460.

Besag J, Green P, Higdon D, Mengersen K. Bayesian computation and stochastic systems. Statistical Science. 1995; 10:3–66.

Bogaerts K, Leroy R, Lesaffre E, Declerck D. Modeling tooth emergence data based on multivariate interval-censored data. Statistics in Medicine. 2002; 21:3775–3787. [PubMed: 12483766]

Brook D. On the distinction between the conditional probability and joint probability approaches in the specification of the nearest neighbour systems. Biometrika. 1964; 51:481–489.

Celeux G, Forbes F, Robert C, Titterington DM. Deviance information criteria for missing data models. Bayesian Analysis. 2006; 1:651–674.

Cressie, N. Statistics for spatial data. New York: Wiley; 1991. p. 61-63.

Dawid AP. Conditional independence in statistical theory (with discussion). Journal of the Royal Statistical Society, Series B. 1979; 41:1–31.

Dempster A. Covariance selection. Biometrics. 1972; 28:157–175.

Drton M, Perlman M. Model selection for Gaussian concentration graphs. Biometrika. 2004; 91:591–602.

Dunson DB. Bayesian latent variable models for clustered mixed outcomes. Journal of the Royal Statistical Society, Series B. 2000; 62:355–366.

Dunson DB. Bayesian methods for latent trait modeling of longitudinal data. Statistical Methods in Medical Research. 2007; 16:399–415. [PubMed: 17656454]

Gelman, A. Running WinBugs and OpenBugs from R. Appendix C. 2003. http://www.stat.columbia.edu/~gelman/bugsR/software.pdf

Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006; 1:515–533.

Gelman A, Tuerlincks F. Type S error rates for classical and Bayesian single and multiple comparison procedures. Computational Statistics. 2000; 15:373–390.

Green P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Giudici P, Green PJ. Decomposable graphical Gaussian model determination. Biometrika. 1999; 86:785–801.

Kass R, Raftery A. Bayes factors and model uncertainty. Journal of American Statistical Association. 1995; 90:773–795.

Klein H, Palmer C. Studies on dental caries vs. familial resemblance in the caries experience of siblings. Public Health Report. 1938; 53:1353–1364.

Leroux B. Analysis of correlated dental data: challenges and recent developments. Statistical Methods for Oral Health Research, JSM 2006. 2006

Lesaffre E, Bogaerts K. Spatial correlations in caries attack patterns in the deciduous dentition. Unpublished Biostatistics Report, K.U. Leuven. 2006

Liang KY, Zeger SL, Qaqish B. Multivariate regression analysis for categorical data. Journal of the Royal Statistical Society, Series B: Methodological. 1992; 54:3–24.

Little, R.; Rubin, D. Statistical analysis with missing data. New York: Wiley; 1987.

McCullagh, P.; Nelder, JA. Generalized linear models. 2nd edition. New York: Chapman & Hall; 1989.

Moustaki I, Knott M. Generalized latent trait models. Psychometrika. 2000; 49:115–132.

O'Brien SM, Dunson DB. Bayesian multivariate logistics regression. Biometrics. 2004; 60:739–746. [PubMed: 15339297]

O'Malley A, James, Zaslavsky AM. Domain-level covariance analysis for multilevel survey data with structured nonresponse. Working Paper. 2008

Pettitt N, Tran TV, Haynes MA, Hay L. A Bayesian hierarchical model for categorical longitudinal data from a social survey of immigrants. Journal of the Royal Statistical Society, Series A. 2006; 169:97–114.

Sammel MD, Ryan LM, Legler JM. Latent variable models for mixed discrete and continuous outcomes. Journal of the Royal Statistical Society, Series B. 1997; 59:667–678.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). Journal of Royal Statistical Society, Series B. 2002; 64:583–640.

Spiegelhalter, DJ.; Thomas, A.; Best, N. WinBUGS. Version 1.4 User Manual. 2003. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf

Stephens M. Bayesian analysis of mixtures with an unknown number of components an alternative to reversible jump methods. The Annals of Statistics. 2000; 28:40–74.

Thomas, A.; Best, N.; Lunn, D.; Arnold, R.; Spiegelhalter, D. GeoBUGS Version1.2 User Manual. 2004. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs12manual.pdf

Todem, D. Oral health. In: Boslaugh, S., editor. Encyclopedia of epidemiology. Vol. vol. 2. Los Angeles: Sage; 2008. p. 762-764.

Vanobbergen J, Lesaffre E, Garca-Zattera MJ, Zattera G, Jara A, Martens L, Declerck D. Caries patterns in primary dentition in 3-, 5- and 7-year-old children: spatial correlation and preventive consequences. Caries Research. 2007; 41:16–25. [PubMed: 17167255]

Vanobbergen J, Martens L, Lesaffre E, Declerck D. *The Signal-Tandmobiel® project*, a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. European Journal of Paediatric Dentistry. 2000; 2:87–96.

Wang FJ, Wall MM. Generalized common spatial factor model. Biostatistics. 2003; 4:569–582. [PubMed: 14557112]

Zeger SL, Karim MR. Generalized linear models with random effects a Gibbs sampling approach. Journal of the American Statistical Association, Theory and Methods. 1991; 86:79–86.

Zhao Y, Staudenmayer J, Coull BA, Wand MP. General design Bayesian generalized linear mixed models. Statistical Science. 2006; 21:35–51.

Zhu J, Erickhoff JC, Yan P. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. Biometrics. 2005; 61:674–683. [PubMed: 16135018]

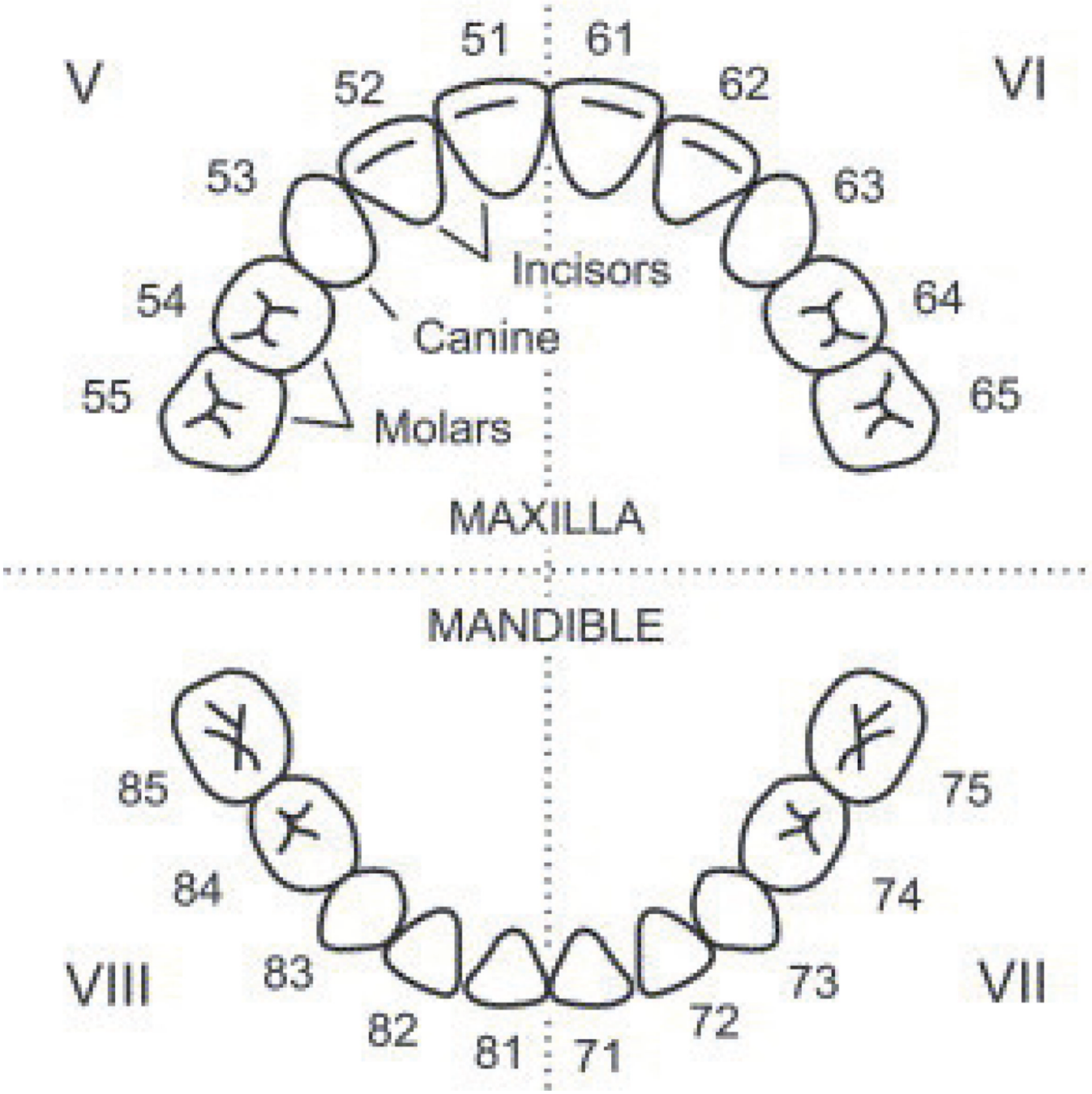

Figure 1.
European notation for the deciduous tooth locations in the mouth (adapted from Bogaerts *et al.*, 2002)

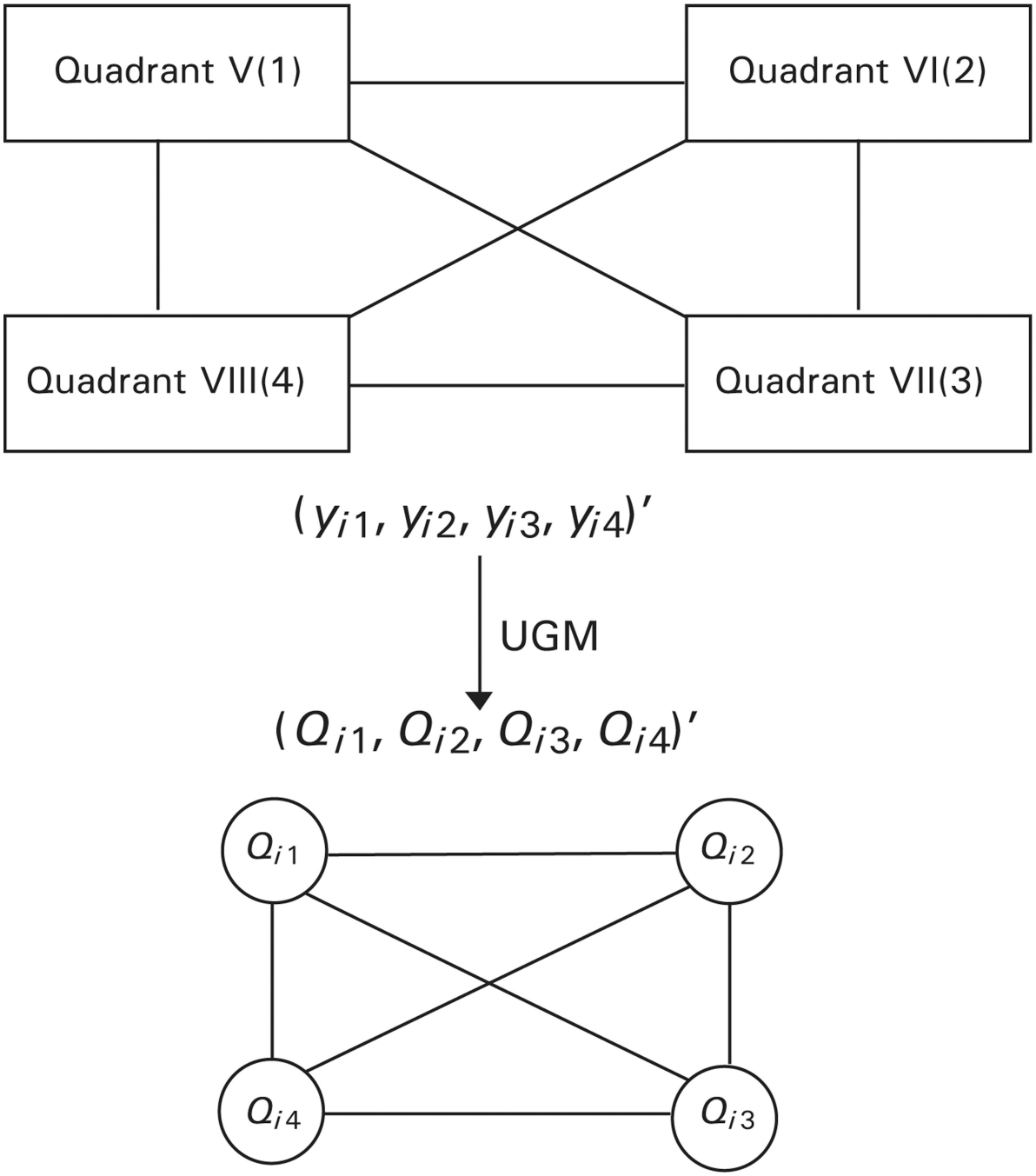

Figure 2.
Undirected graphical representation for quadrant-level latent vector

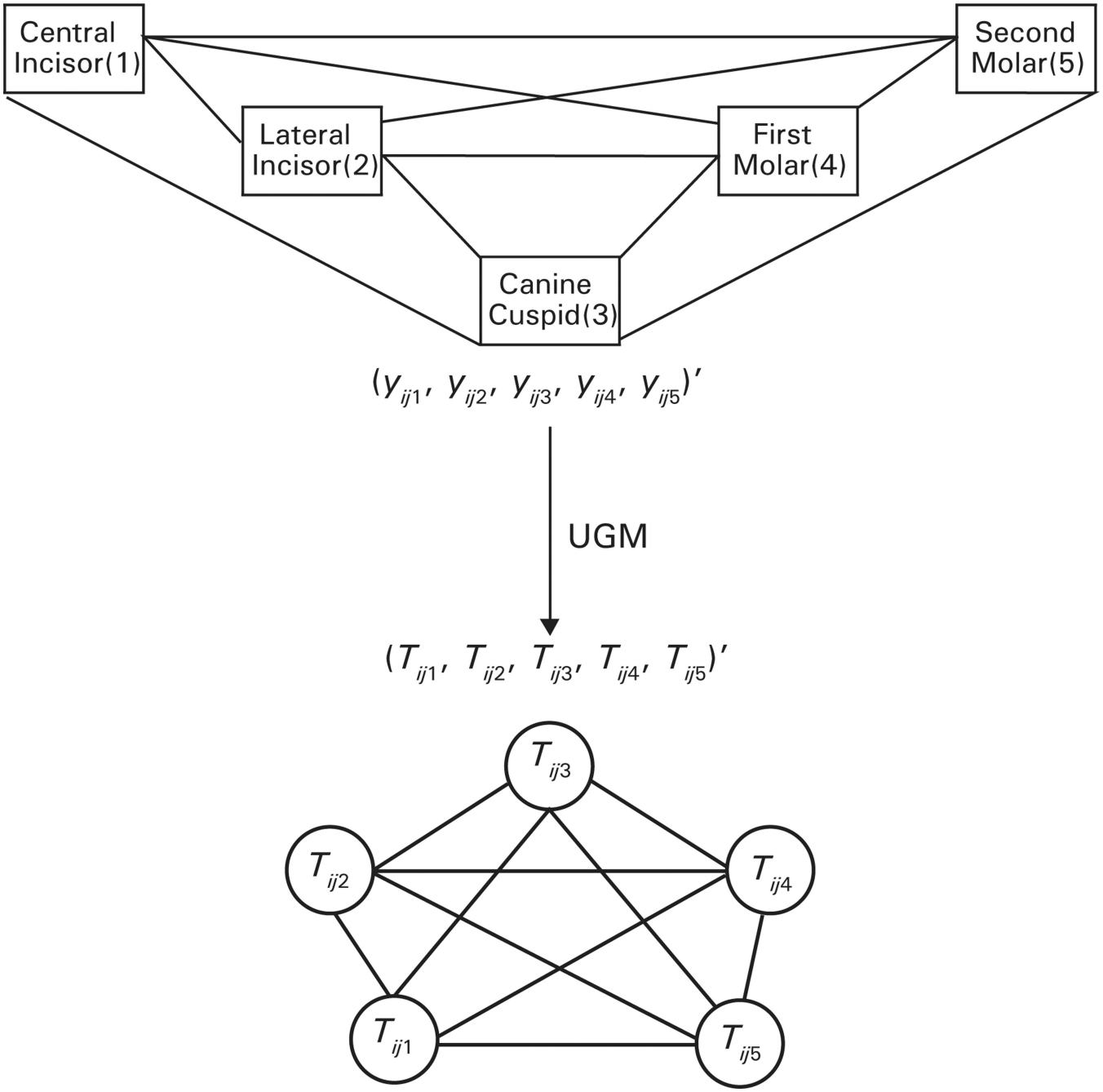

Figure 3.
Undirected graphical representation for the tooth-level latent vector

Table 1

Prevalence of caries experience (% affected) in the primary dentition of 7-year-old children $n = 4351$

Tooth	55	54	53	52	51	61	62	63	64	65
Prevalence	8.92	5.20	0.74	3.72	7.81	7.06	2.23	1.86	5.20	8.55
Tooth	85	84	83	82	81	71	72	73	74	75
Prevalence	10.78	13.75	1.12	0.74	0.37	0.37	0.37	0.37	11.15	9.67

Table 2

ALR estimates and 95% confidence intervals of the odds ratios for the first and second molars using the subset data {54 55; 64 65; 74 75; 84 85} of 7-year-old children

First molar (ALR model)			
	64	74	84
54	16.48 (13.75–19.74)	8.17 (6.91–9.64)	7.23 (6.13–8.53)
64		7.61 (6.47–8.97)	7.18 (6.10–8.44)
74			22.82 (19.28–27.00)
Second molar (ALR model)			
	65	75	85
55	15.47 (13.09–18.28)	8.78 (7.52–10.27)	9.23 (7.90–10.79)
65		8.08 (6.92–9.42)	8.86 (7.58–10.35)
75			20.37 (17.20–24.11)

Table 3

Estimates and simultaneous 95% credible intervals of quadrant-level association differences under the UGGM with a within-quadrant structured CAR covariance model and an unstructured covariance

Hypothesis parameter	Estimates (credible intervals) structured CAR Σ_T with $\sigma_j \sim U(0,1000)$	Estimates (credible intervals) unstructured Σ_T
Overall quadrant-level association differences		
Left/right across		
$\tilde{\rho}_{LR} - \tilde{\rho}_{A}$	0.921 (0.360, 1.494)	0.966 (0.348, 1.539)
Left/right vs up/down		
$\tilde{\rho}_{LR} - \tilde{\rho}_{UD}$	0.849 (0.313, 1.454)	0.823 (0.285, 1.463)
Across vs up/down		
$\tilde{\rho}_{A} - \tilde{\rho}_{UD}$	−0.072 (−1.158, 1.070)	−0.143 (−1.223, 1.097)
Pairwise quadrant-level association differences		
Left/right vs across		
$\tilde{\rho}_{12} - \tilde{\rho}_{24}$	0.728 (0.089, 1.437)	0.801 (0.134, 1.480)
$\tilde{\rho}_{12} - \tilde{\rho}_{13}$	1.077 (0.453, 1.612)	1.099 (0.435, 1.644)
$\tilde{\rho}_{34} - \tilde{\rho}_{24}$	0.765 (0.178, 1.435)	0.832 (0.207, 1.493)
$\tilde{\rho}_{34} - \tilde{\rho}_{13}$	1.113 (0.476, 1.642)	1.130 (0.438, 1.671)
Left/right vs up/down		
$\tilde{\rho}_{12} - \tilde{\rho}_{23}$	0.887 (0.329, 1.497)	0.856 (0.303, 1.512)
$\tilde{\rho}_{12} - \tilde{\rho}_{14}$	0.774 (0.135, 1.441)	0.759 (0.082, 1.465)
$\tilde{\rho}_{34} - \tilde{\rho}_{23}$	0.923 (0.306, 1.570)	0.887 (0.296, 1.554)
$\tilde{\rho}_{34} - \tilde{\rho}_{14}$	0.811 (0.239, 1.433)	0.790 (0.224, 1.457)
Across vs up/down		
$\tilde{\rho}_{24} - \tilde{\rho}_{23}$	0.158 (−1.039, 1.315)	0.055 (−1.133, 1.284)
$\tilde{\rho}_{24} - \tilde{\rho}_{14}$	0.046 (−1.138, 1.188)	−0.042 (−1.190, 1.182)
$\tilde{\rho}_{13} - \tilde{\rho}_{23}$	−0.190 (−1.257, 1.017)	−0.243 (−1.307, 1.052)
$\tilde{\rho}_{13} - \tilde{\rho}_{14}$	−0.303 (−1.367, 0.928)	−0.430 (−1.415, 0.979)
DIC	893.394	844.952
# burn-in samples	1000	1000
# MCMC iterations	11 000	11 000

Table 4

Parameter estimates and corresponding 95% credible intervals under the UGGM with a within-quadrant structured CAR covariance model and an unstructured covariance

Fixed effects	**Estimates (credible intervals) under structured CAR Σ_T with $\sigma_j \sim U(0, 1000)$**	**Estimates (credible intervals) under unstructured Σ_T**
Overall intercept		
α	−4.032 (−5.456, −2.669)	−4.541 (−5.893, − 3.278)
β_j: quadrant j		
β_1	−0.253 (−1.813, 1.151)	−0.502 (−2.239, 1.604)
β_2	−0.091 (−1.522, 1.296)	−0.175 (−1.671, 1.199)
β_3	−0.749 (−2.442, 0.764)	−0.504 (−2.277, 1.099)
$\gamma_{k(j)}$: Tooth k nested within quadrant j		
$\gamma_{1(1)}$	−0.209 (−1.288, 0.898)	0.087 (−1.353, 1.594)
$\gamma_{2(1)}$	−1.618 (−2.858, −0.467)	−1.485 (−3.066, 0.084)
$\gamma_{3(1)}$	−4.108 (−6.179, −2.364)	−4.170 (−6.350, −2.231)
$\gamma_{4(1)}$	−1.111 (−2.235, −0.019)	−0.895 (−2.174, 0.487)
$\gamma_{1(2)}$	−0.373 (−1.457, 0.688)	−0.149 (−1.585, 1.282)
$\gamma_{2(2)}$	−2.304 (−3.756, −1.018)	−2.306 (−3.947, −0.732)
$\gamma_{3(2)}$	−2.489 (−3.922, −1.197)	−2.399 (−4.006, −0.929)
$\gamma_{4(2)}$	−0.967 (−2.145, 0.134)	−0.657 (−1.843, 0.562)
$\gamma_{1(3)}$	−7.018 (−10.540, −4.279)	−6.602 (−10.120, −3.696)
$\gamma_{2(3)}$	−6.861 (−10.360, −4.147)	−6.430 (−9.936, −3.516)
$\gamma_{3(3)}$	−5.532 (−8.083, −3.365)	−5.097 (−7.692, −2.872)
$\gamma_{4(3)}$	0.721 (−0.365, 1.935)	1.094 (−0.114, 2.473)
$\gamma_{1(4)}$	−6.928 (−10.380, −4.383)	−6.903 (−10.390, −4.106)
$\gamma_{2(4)}$	−5.432 (−7.934, −3.437)	−5.288 (−7.790, −3.154)
$\gamma_{3(4)}$	−4.140 (−5.965, −2.572)	−4.016 (−5.983, −2.219)
$\gamma_{4(4)}$	1.078 (0.076, 2.200)	1.468 (0.369, 2.722)
Pairwise quadrant-level association		
$\tilde{\rho}_{12}$	0.889 (0.647, 0.977)	0.901 (0.626, 0.983)
$\tilde{\rho}_{13}$	−0.188 (−0.718, 0.477)	−0.199 (−0.745, 0.518)
$\tilde{\rho}_{14}$	0.115 (−0.523, 0.708)	0.141 (−0.546, 0.729)
$\tilde{\rho}_{23}$	0.003 (−0.631, 0.603)	0.045 (−0.621, 0.639)
$\tilde{\rho}_{24}$	0.161 (−0.509, 0.736)	0.099 (−0.552, 0.716)
$\tilde{\rho}_{34}$	0.925 (0.766, 0.986)	0.932 (0.769, 0.986)
Overall quadrant-level association		
$\tilde{\rho}_{LR} = 0.5(\tilde{\rho}_{12} + \tilde{\rho}_{34})$	0.907 (0.775, 0.970)	0.916 (0.753, 0.977)
$\tilde{\rho}_{A} = 0.5(\tilde{\rho}_{13} + \tilde{\rho}_{24})$	−0.014 (−0.573, 0.569)	−0.049 (−0.608, 0.590)

Fixed effects	**Estimates (credible intervals) under structured CAR Σ_T with $\sigma_j \sim U(0, 1000)$**	**Estimates (credible intervals) under unstructured Σ_T**
$\tilde{\rho}_{UD} = 0.5(\tilde{\rho}_{14} + \tilde{\rho}_{23})$	0.059 (−0.538, 0.615)	0.093 (−0.552, 0.644)
DIC	893.394	844.952
# burn-in samples	1000	1000
# MCMC iterations	11 000	11 000

† We set $\beta_4 = 0$ for the reference quadrant (Quadrant 4) and $\gamma_{5(j)} = 0$ for $j = 1, \ldots, 4$ for the reference tooth (the last molar of each quadrant).