This is to certify that the
dissertation entitled

A Hierarchical Bayesian Approach to Model Spatially Correlated
Binary Data with Applications to Dental Research

presented by

Yanwei Zhang

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Statistics

David Todem
Major Professor's Signature

05/08/08
Date

MSU is an affirmative-action, equal-opportunity employer

A HIERARCHICAL BAYESIAN APPROACH TO MODEL SPATIALLY CORRELATED BINARY DATA WITH APPLICATIONS TO DENTAL RESEARCH

By

Yanwei Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics

2008

ABSTRACT

A HIERARCHICAL BAYESIAN APPROACH TO MODEL SPATIALLY CORRELATED BINARY DATA WITH APPLICATIONS TO DENTAL RESEARCH

By

Yanwei Zhang

Statistical analysis of multivariate binary data measured repeatedly in time or cross-sectionally clustered in space, besides the difficulties of non-continuous nature of data, raises a number of challenges. For instance, dental data from oral health research community are always discrete, clustered spatially and repeated in time. The researchers are interested in the risk factors and spatial symmetry property of caries prevalence incidence. It is well believed, for example, that the caries outcomes adjacent to each other are highly likely to be correlated, which necessitates the use of methodologies for correlated discrete data. Generalized estimating equation(GEE) based approach might help answer marginal mean and pairwise association types of research questions about correlated units of interest. When association among units is of primary research concern, GEE suffers seriously from less efficiency. Methodologies for analyzing multivariate categorical data clustered in space, with both marginal mean and association being of research interest, need to continue. In this thesis, we will introduce complete likelihood based approaches for analyzing spatially correlated binary data. Specifically, we are going to discuss a class of methods that attempt to explicitly take some very unique spatial structure features into consideration for valid and efficient inferences at tooth level. Furthermore, we proposed different models by using latent variables with hierarchical levels to account for the spatial dependence

of the data features from different points of view. The hierarchical structure of the model and local identifiability of latent variable models make the statistical inference appropriate within Bayesian framework through the MCMC based posterior sampling algorithm. Comparison among the performances of different models was made under Bayesian model selection criterion (DIC) for missing data problem. Finally, we gave Bayesian hypothesis testing for the spatial symmetries of caries incidence by providing semitendinous credible regions for the differences of quantities that were used to measure spatial association strength. The methodology is illustrated by using dental data from Signal Tandmobiel (STM) project.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. David Todem for his valuable advice and continuing encouragements during my years as a research assistant at the department of epidemiology. His help and expectation made my academic life challenging but a rewarding one. I am also grateful to the financial support from his funded project. Dr. Todem's hardworking, dedication, and creativeness have been and will always be inspiring me to achieve my career goal.

Appreciation is extended to Dr. Ramamoorthi and Dr. Gardiner for making it possible for me to become an applied statistician of being Bayesian in principle. I would like to thank Dr. Ramamoorthi, in the first year of my studying in east lansing, who helped me went through the toughest time in my life so far. Dr. Gardiner gave me some advice that helped me make transitions from a mathematician to an applied statistician. I also would like to thank Dr. Melfi and Dr. Cui for being my thesis committee members.

Special thanks to my parents, they are the greatest parents I could possibly ask for. Especially, my father has been always supportive of my efforts for carrying out my career goals. It was my father who tried as much as possible to make my education financially possible. I dedicate this work to all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background and objectives

In biomedical studies, it is common in practice for a binary disease outcome to be measured either repeatedly across time or cross-sectionally across spatial spots. The motivating example for this research comes from dental research. The caries status of teeth are evaluated as binary outcomes with 1 indicating the presence of caries and 0 otherwise. The caries prevalence incidences are suspected to have a certain spatial symmetry property in terms of the quadrants configuration within the mouth, which is well believed by dentists in practice. It is well known, for example, that the dental caries outcomes adjacent to one another are likely to be correlated. Specifically, there are four quadrants within the mouth and all the quadrants are believed to be correlated to one another. Within each quadrant, the adjacent teeth are also likely to be correlated and the correlation might be affected by the quadrant. Hence, it necessitates the use of methodologies for correlated data to analyze dental data

When a patient first visits a dentist, either for a check-up or a more serious dental issue, the dentist will normally conduct a full examination to gain an understanding of the patient's overall dental health as well as the patient's particular dental

problem(s), if any. Because of the complexity and diversity of dental issues: and the numerous teeth involved, it is difficult for the dental health researchers to analyze the dental data, except in a most general and superficial way with respect to quadrant,tooth position, age, sex, geographical region, etc. In dental practice, it is of interest to find out some patterns in terms of caries of the teeth, which will help the dentist efficiently examine oral health of the patients and provide people informative guidance for intervention of caries. Researchers have been working on different methodologies to analyze the dental data to address caries incidence pattern related questions. The traditional method for analyzing dental data is based on the number of Decayed/Missing/Filled Surfaces (DMFS) or Decayed/Missing/Filled Teeth (DMFT), introduced by Klein *et al.* (1938). DMFT and DMFS can roughly express the caries prevalence numerically and are obtained by calculating the number of Decayed (D), Missing (M) and Filled (F) teeth (T) or surfaces (S) within the mouth. The DMFT evaluation method is a well-known technique and has been used for many years to analyze the effects of variables, such as fluoride, on the dental health of given populations. This approach operates the analysis at the mouth level, which is not informative, in terms of caries pattern, to dentists and patients for oral health examination and caries interventions. Dentists and patients are really interested in spatial symmetry patterns of caries. For example, if one caries was found on one specific tooth within a specific quadrant, which tooth will be the next that is highly likely to have caries. If the dentists has some information about the spatial symmetry of the caries, they may efficiently locate or predict which is the next tooth with high risk of developing caries. If so, dentists and patients may be able to pay more attentions to the teeth with high risk. Due to the spatial configuration of the quadrants and teeth within each quadrant, the nature of the data requires the methodology for correlated binary data.

Lesaffre *et al.* (2006) proposed a several methods to analyze the dental data from

the Signal Tandmobiel (STM) project. Their approach was based on the Generalized Estimating Equation (GEE)(Zeger and Liang, 1986) to deal with correlated data. Lesaffre's approach used logistic regression model framework to model marginal caries incidence using exchangeable working correlation matrix to account the dependence of the data. Their GEE based approach is not able to capture the special correlation structure among quadrants and among teeth within the same quadrant. Roy (2006) proposed a model-based approach for imputing these missing values. His method exploited the spatial correlation among teeth without considering the different strength of spatial dependence among quadrants. Vanobbergen *et al.*(2007) proposed ALR(Alternating Logistic regression)(Carey *et al.* (1993)) approach to investigate spatial correlation respect to caries patterns in primary dentition in 7-year-old children. At the population level, symmetry in the prevalence of caries experience across the midline was tested at the tooth and tooth surface levels under ALR model. ALR simultaneously modeled marginal expectation of each binary variable as well as the association between paries of outcomes using GEE. Liang *et al.* (1992) showed that GEE estimates only can reasonably efficient when covariance structure of the response variables is correctly specified. Meanwhile, ALR models have issues of convergence when the cluster size is large.

GEE based logistic regression models and ALR models are both marginal model, which means they did not take care of the heterogeneity and dependence among quadrants and teeth nested within corresponding quadrants. The estimate of parameters of interest for fixed effect is consistent, but it might be inefficient and seriously biased. The GEE based approach, as a distribution free methodology, does not lend itself to classical tools for model checking. GEE is based on the first order moment and ALR is trying to model the higher order moment of the data while still only focusing on pairwise association without trying to model the joint relationship among the observations. More importantly, it is infeasible to address to the spatial symmetry of

association strength among quadrants and the teeth within corresponding quadrants since all these higher order moments characteristics are unobserved. Hence, searching for alternative solutions continues.

The valid and efficient joint model for the spatially correlated binary dental data is to incorporate latent variables to induce the dependence structure among quadrants and the nested dependence structure among teeth within corresponding quadrants. Meanwhile the latent variables also can generate a flexible multivariate distributions for the binary dental data. Without obvious multivariate distributions for the multivariate spatially correlated binary data, the joint model for accounting the nature of the data is not straightforward. Another way to model the dental data is using mixture models. Specifically, we can view the distribution of the caries status of the tooth of interest as being a mixture of bernoulli distributions with different probabilities of success. The probability of the incidence of caries is modeled by a logistic regression model that takes the design structure, quadrant and tooth position within the corresponding quadrant, into account. Generalized latent variables and mixture models allow factorization of the joint distributions of the multivariate correlated binary data into the product of a conditional distributions, given the latent variables and allocation random variables that induce the unobserved heterogeneities and dependence structures among the observations. The objective of this thesis is to develop a new methodology for complex and likelihood based analysis of multivariate spatially correlated binary caries experience from the dental data, which can help us examine spatial symmetry of the quadrants, association strength among teeth within each specific quadrants. In this thesis, we proposed Bayesian generalized latent variable model (BGLVM) and Bayesian mixture of generalized latent variable model (BMGLVM) to give flexible multivariate distributions of the spatially correlated binary dental data with dependence structure induced by the latent variables. BGLVM and BMGLVM are specified from Frequentist's point of view but implemented under

Bayesian framework. The BGLVM uses logistics regression model giving a flexible multivariate distribution for the dental data with two level of latent variables inducing dependence structure for corresponding level of spatial configuration. For the BGLVM, the dependence structures among quadrants and teeth nested within quadrants are induced by the latent variable models whose covariance structure, modeled by undirected graphical Gaussian model or conditional autoregressive model. For the BMGLVM, the dependence among quadrants is induced by the weights of the mixture components of the mixture model and the dependence among teeth within the same quadrant is induced by generalized latent variable model in the same way as in BGLVM.

1.2 Principles for the analysis

The principle of our approach for modeling the multivariate spatially correlated dental data is based on the concept of latent variables that are incorporated into the likelihood based model for generating flexible multivariate distributions for the observations and inducing multilevel dependence structures due to unobserved heterogeneity from the complex structure of the multivariate correlated binary data. Specifically, two level of random vectors are introduced into to the model via latent variables, which are used to induce spatial dependence structures among subunits at their corresponding level and generate flexible distributions for the subunits. The joint distribution for each of the two levels of latent variables is given by Undirected Graphical Gaussian Model (UGGM)(Dempster,1972, Giudici and Green 1999) with respect to different spatial configurations of the subunits at corresponding level. Each level of latent variables is used to induce spatial dependence among subunits at the corresponding level. The first level of spatial dependence structure is the spatial association among four quadrants. The four quadrants are adjacent in spatial frame and also coexist in the

same oral biological environment, which make them correlated in some unobserved structure. the second level of spatial dependence structure is the spatial association among the teeth within the same quadrant. It is reasonable to believe that teeth adjacent to one another are likely to be correlated. Meanwhile, we know the oral biological environment is very complicate in the way that the associations exist not only between teeth adjacent to each other, but also with other teeth in the same quadrant. The UGGMs for the latent variables will be based on different precision matrixes: one is unstructured type and the other is Markovian type based on CAR model (Cressie (1991)).

In this thesis, we are trying to combine the merits of frequetist's and Bayesian's in model formulations and implement. Specifically, the design structure based models are formulated within the framework of frequentist for considering the marginal identifiability of the model. The latent variables are incorporated hierarchically in the graphical structure of Bayesian model and models are implemented in Baeysian principle. Since our models are based on latent variable approach, local identifiability and model complexity will raise lots of technical problems within frequentist's framework. For example, computational feasibility in optimization, singularity of the information matrix and accuracy and computational feasibility of high dimensional integration approximation by using either adaptive Gaussain quadrature or MCMC based approaches. Bayesian provides a way to avoid all the above technical concerns by using Gibbs sampling to obtain the posterior distributions of the quantities of interest. We use noninformative priors for the parameters of interest, since posterior inference will not rely on the subjective prior information and it will also give the comparable result with frequentist's as sample size increase. Meanwhile, we use independent proper conjugate priors to the parameters of interest, which will ensure the validity of the posterior samples obtained by Gibbs sampling and improve the convergence of the MCMC based posterior sampling algorithm. More importantly,

Bayesian approach can be helpful in complex modeling situations where a frequentist analysis is difficult or does not exist. Lee and Song demonstrated better performance of a Baysian approach in small samples compared with ML estimation. Frequentist's results rely on the asymptotic arguments, but Bayesian inference is feasible as long as the posterior sampling algorithm converge which can be increased easily in large number of MCMC iterations. All the inferences will be based on credible intervals within Bayesian framework and implemented in *WinBUGS*. The appropriate model will be chosen by a formal Bayesian model selection criteria based on the DIC for missing data problems (Celeux *et al.* 2006).

1.3 Outline of the thesis

In chapter 2, we will systematically describe the principles of generalized latent variable approaches for joint modeling correlated discrete data. We will also describe the generalized latent variable model context within the Bayesian framework for analyzing the dental from STM. We use multivariate spatial latent variables at both quadrant level and tooth nested within quadrant level to model a very flexible multivariate distribution for the binary vectors and induce spatial dependence among tooth through the dependence structure of the spatial latent vectors in the generalized linear model settings. The joint relationship among spatial latent will be modeled under the context of undirected graphical model and conditional autoregressive model correspondingly. Model fitting and statistical inferences about the parameters of interest are going to be under Bayesian framework.

In chapter 3, we will describe the finite mixture model within the Bayesian framework for analyzing the dental from STM. We use Dirichlet process to model the mixing proportions and multivariate spatial latent variables to model a very flexible multivariate distribution for the mixture component and induce spatial dependence

among teeth through the dependence structure of the spatial latent vectors in the latent variable model settings. The joint relationship among spatial latent will be modeled under the context of undirected graphical model and conditional autoregressive model correspondingly. Model fitting and statistical inferences about the parameters of interest are going to be under Bayesian framework.

In Chapter 4, we will summarize our work and give some routes for the future work.

CHAPTER 2

Bayesian Generalized Latent Variable Models

2.1 Introduction

Dental caries is a common oral disease that results in demineralization of the tooth. In oral health research, the number of Decayed/Missing/Filled Surfaces (DMFS) or Decayed/Missing/Filled Teeth (DMFT), introduced by Klein *et al.* (1938), are often analyzed. The two scores are the sums of binary indicators of caries on the teeth and tooth surfaces for the primary dentition. This approach operates the analysis at the mouth level. Leroux *et al.* (2006) mentioned dental data presents an unique set of challenges for statistical analysis, including large cluster sizes, multilevel data structures (e.g., teeth within patients, sites or surfaces within teeth), complex correlation structures. Lesaffre *et al.* (2006) proposed several methods to analyze the dental data from the Signal Tandmobiel (STM) project. They used GEE based logistic model and log-linear model to model marginal mean with exchangeable working correlation matrix to account for the dependence of the data. Vanobbergen *et al.*(2007) proposed ALR(Alternating Logistic regression) approach to investigate spatial correlation

with respect to caries activities patterns in primary dentition in 7-year-old children. ALR simultaneously models marginal expectation of each binary variable as well as the association between paries of outcomes. Zhu *et al.* (2005) proposed a generalized latent variable model framework to analyze multivariate spatially correlated data, which gave an appropriate approach to complex spatially correlated data with large cluster sizes and multilevel data structures. Their approach is sensitive to Euclidian space, and can not take care of multi-level dependence structure of the dental data. More importantly, their method is EM based and implemented via MCMC, which is computationally intensive for high dimensional correlated latent variables posterior sampling and without fisher information matrix as byproduct. The purpose of this article is to introduce a Bayesian Generalized Latent Variable Model (BGLVM) framework for general spatial topology structures to explain multi-level correlations introduced by "between-cluster" and "within-cluster" random effects. Specifically, the "between-cluster" random effects are used to induce dependence among quadrants and "within-cluster" random effects are used to induce dependence among teeth within the same quadrant. The BGLVM, implemented using Gibbs sampling with non-informative priors, allows us to model the "between-cluster" and "within-cluster" correlation structures explicitly. It is possible for us to examine the spatial symmetry of quadrants in terms of caries incidence, and capture the special spatial association structure between quadrants for the same subject of interest and among teeth within quadrants, which can help us efficiently characterize the caries incidence at tooth level.

2.2 The Spatial Dependence Structures

2.2.1 Notation

To model the observations, let y_{ijk} denote the kth response variable within jth cluster of ith subject of interest, where $k = 1, ..., K, j = 1, ..., J, i = 1, ..., n$. Let $y_{ij} = (y_{ij1}, ..., y_{ijk}, ..., y_{ijK})'$ denote the response vector within jth cluster of ith subject. Let $y_i = (y'_{i1}, ..., y_{ij}, ..., y'_{iJ})'$ denote the collection of response variables of ith subject. let $y = (y'_1, ..., y'_i, ..., y'_n)'$ denote the collection of response variables of all subjects in this study.

For modeling the latent variables, we use undirected graphical Gaussian model. let $Q_i = (Q_{i1}, ..., Q_{ij}, ..., Q_{iJ})'$ denote the latent variables at cluster level for i subject, where $i = 1, ..., n$. Let $T_{ij} = (T_{ij1}, ..., T_{ijk}, ..., T_{ijK})'$ denote the intermediate level latent variables that are nested within the jth cluster associated with the ith subject. Let $T_i = (T'_{i1}, ..., T_{ij}, ..., T'_{iJ})'$ denote the collection of all latent variables at intermediate level associated with the ith subject in the study. Let $L_i = (Q'_i, T'_i)'$ denote the collection of latent variables at both levels associated with the ith subject.

2.2.2 Principles of our modeling approach

The dental data shows a two-level spatial association structures, i.e., the first level spatial association structures are among quadrant(V)-(VIII). For the convenience of indexing the data, we will use quadrant(I) instead of quadrant (V) and corresponding index for the other quadrant. The second level spatial association structure is, nested within corresponding quadrant, the spatial correlation among teeth.

In general, the valid approaches for analyzing correlated data without explicit multivariate distribution consist are based on either GEE or random effect models. The former is suitable for marginal mean or pairwise associations between response outcomes orientated statistical problems and the latter is for subject specific statis-

tical issues. The dental data is spatially correlated and has information about teeth spatial configurations that need to be incarnated in the model to provide explicit structure for inducing dependence among quadrants and teeth at their corresponding levels. The main contribution of this paper is to develop a methodology to model this unique spatial dependence of the deciduous dentition. There is no explicit multivariate distribution available for the spatially correlated binary dental caries experience outcomes. Generalized latent variable models (Skrondal & Rabe-Hesketh(2004)) are commonly used to generate flexible multivariate distributions and induce unobserved heterogeneity for correlated data with implicit multivariate distribution.

To take the unique spatial structure of dental data into account, we use two levels of latent variables to take care of the spatial dependence of the teeth within the mouth for each subject. For the ith subject, at the higher level, we introduce the quadrant level latent vector Q_i that is used to tight the four quadrants by inducing dependence structure among quadrants. The latent vector at higher level is also used to generate flexible multivariate distributions for the quadrant specific response vectors. The joint distribution of this spatial latent vector is given by Undirected graphical Gaussian model with spatial configurations of the quadrants taken into account. The quadrant-wise observation vectors $\{y_{ij} : j = 1, ..., J\}$ will be conditionally independent given Q_i for $i = 1, ..., n$. At the intermediate level level, quadrant-specific spatial latent vector T_{ij} is introduced, which is used to tight the five teeth within the same quadrant by inducing dependence structure among teeth within the same quadrant. Similarly, the intermediate level spatial latent vector is also used to generate flexible univariate distributions for the tooth specific response outcomes. The joint distribution of this spatial latent vector is given by Undirected graphical Gaussian model with spatial configurations of the teeth and the quadrant taken into account. The observations $\{y_{ijk} : k = 1, ..., K\}$ will be conditionally independent given T_{ij} for $j = 1, ..., J$ and $i = 1, ..., n$. Meanwhile, the intermediate level spatial latent vectors $\{T_{ij} : j =$

$1, ..., J\}$ are conditional independent given the higher level spatial latent vector Q_i for $i = 1, ..., n$. In order to assess the spatial symmetry of the caries experience of deciduous dentition, we will examine the association among latent variables at higher level. Due to the complexity of oral biological system, we will give flexible covariance structure for the undirected graphical Gaussian models and formal model selection procedure will be used to choose appropriate one for the data.

2.3 Models

2.3.1 Generalized Latent variable model

Sammel(1997) proposed an joint model for different outcomes in Generalized linear model framework with normal latent variables introduced to different models. Moustaki(2000) extended this framework to a class of generalized latent trait models. Both of the approaches are based on EM algorithm for model fitting and the computational hurdles arise seriously as the number of latent variables increases. One of the primary difficulties is in integrating out the latent variables, although standard approximation can be used, the accuracy will decrease with the dimension of the latent variables. Dunson(2000) proposed a model allows observed and latent variables to have distribution in exponential family. Wang's (2003) multivariate spatial latent variable model was extended by Zhu *et al.* (2005) into generalized linear latent variable models for repeated measurements of spatially correlated multivariate data. A MCEMG(Monte Carlo EM Gradient) algorithm was used for model implement, which was based on numerical approximations to marginalize the score functions and Hessian matrix over latent variables. It is well known that MCEMG is seriously computationally intensive and less accurate as the dimension of latent variables increases.

In this paper, we propose a Bayesian generalized linear latent variable models with two levels of spatial latent vectors. The joint distributions of the latent vectors are given by Undirected Graphical Gaussian model(UGGM) (Dempster,1972,

Giudici and Green 1999). In order to test the spatial symmetry property of tooth caries experience within the subject, we proposed statistical hypothesis testing for all possible situations under Bayesian framework. Under the latent variable models, it is assumed that, given the two levels of the spatial latent vectors, the teeth are mutually conditionally independent then we can specify the complete likelihood. We will use MCMC approach to perform posterior inference for the quantities of interest using non-informative priors, which will give the data more flexibility to decide what is going on and also can give comparable inference results to Frequentist's.

Response Models

We model the kth response variable within the jth quadrant of the ith subject, y_{ijk}, which is a binary indicator of caries experience of $tooth_{ik(j)}$. Conditional on the corresponding two levels of latent variables for the kth tooth position within the jth quadrant of the ith subject, the response model is given by an exponential family distribution with the probability density function in a general form

$$p(y_{ijk}|L_i,\alpha,\gamma,\beta,\varphi) = p(y_{ijk}|\eta_{ijk},\varphi) = \exp\{\frac{\eta_{ijk}y_{ijk} - b_i(\eta_{ijk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}, \quad (2.1)$$

where $\eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{ik(j)}$ (McCullagh and Nelder *et al.* 1989).

We assume that the link function $g(\cdot)$ is a canonical link that relates the mean of y_{ijk} to a linear predictor as follows

$$g(E[y_{ijk}|\eta_{ijk}]) = \eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{ik(j)},$$

where $\alpha, \beta = (\beta_1, ..., \beta_j, ..., \beta_J)', \gamma = (\gamma_{1(1)}, ..., \gamma_{K(1)}, \gamma_{1(2)}, ..., \gamma_{K(J)})'$ are the regression coefficients for the fixed effects with constraints $\sum_j \beta_j = 0$ and $\sum_k \gamma_{k(j)} = 0; j = 1, ..., J$ for identifiability of the marginal mean. Q_{ij} and $T_{ik(j)}$ are the random effects that are used to generate flexible multivariate distributions and induce dependence unobserved heterogeneity of the spatially correlated binary dental caries outcomes. It

is assumed that the quadrant level spatial latent vector $\{Q_i\}$ are identically independent Gaussian with zero mean and covariance matrix Σ_Q. Furthermore, we assume that, given the quadrant level spatial latent vectors $\{Q_i : i = 1, ..., n\}$, the tooth level spatial latent vectors $\{T_{ij} : j = 1, ..., J, i = 1, ..., n\}$ are mutually independently multivariate Gaussian with mean zeros, covariance matrix $\{\Sigma_T^j : j = 1, ..., J\}$ correspondingly. The generalized linear model relates the response variables to quadrant-specific and tooth-specific covariates and the latent variables.

Under the latent variable model approach, we can assume that the response variables are conditionally mutually independent, given the vectors of latent variables $L = \{L_i = (Q_i', T_{i1}', ..., T_{ij}', ..., T_{iJ}')' : i = 1, ..., n\}$. The joint probability density of y conditional on the set of latent variables L and $\{\alpha, \beta', \gamma', \varphi\}$ is as follows

$$\begin{aligned} & p(y|L, \alpha, \beta', \gamma', \varphi) = \prod_{i=1}^{n} \prod_{j=1}^{J} \prod_{k=1}^{K} p(y_{ijk}|\eta_{ijk}, \varphi) \\ & = \exp[\sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K} \{\frac{\eta_{ijk} y_{ijk} - b_i(\eta_{ijk})}{a_i(\varphi)} + c_i(y_{ijk}, \varphi)\}] \end{aligned} \tag{2.2}$$

Structure Models for Latent Variables

In the response model, given the two levels of spatial latent variables, the conditional independence assumption allows the specification of complete likelihood for the response model. In our modeling approach, the two levels of spatial latent vectors are used to induce the dependence structure of the teeth of interest. In order to incorporate appropriate spatial latent vectors into the model, we need to choose the ones that can really represent the design structure and characterize the random mechanism of data generating process. The objective of these latent processes is to generate flexible distributions for observations and induce the dependency among observations. UGGMs need to work on specific nodes spatial configurations and we list the possible graphs for both quadrant and tooth nested within quadrant levels as below.

As shown above graphs, the four quadrants can be viewed as four nodes in a graph. If two nodes are not directly connected, they are said to be conditionally indepen-

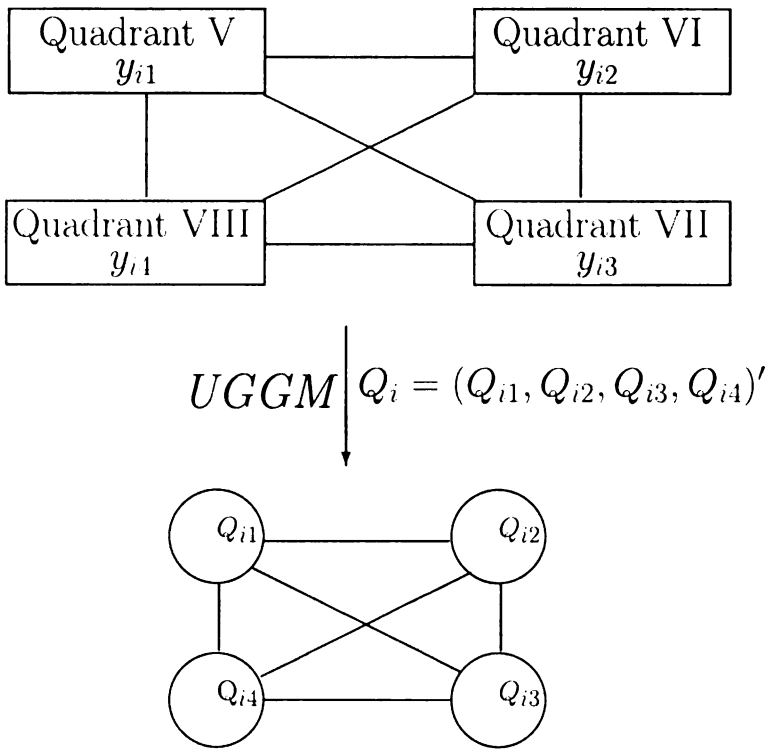

Figure 2.1. The response vectors y_{i1}, y_{i2}, y_{i3} and y_{i4} are tighten by spatial latent vector $Q_i = (Q_{i1}, Q_{i2}, Q_{i3}, Q_{i1})'$ whose joint distribution is given by UGGM with unstructured precision matrix.

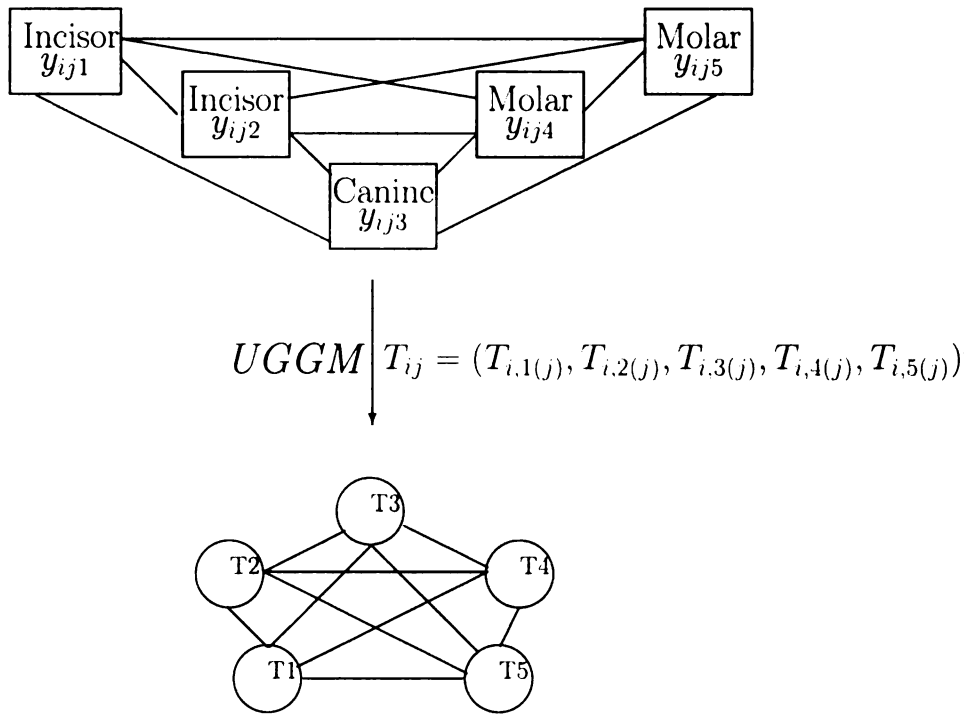

Figure 2.2. The response variables y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4} and y_{ij5} are tighten by spatial latent vector $T_{ij} = (T_{i,1(j)}, T_{i,2(j)}, T_{i,3(j)}, T_{i,4(j)}, T_{i,5(j)})'$ whose joint distribution is given by UGGM with unstructured precision matrix.

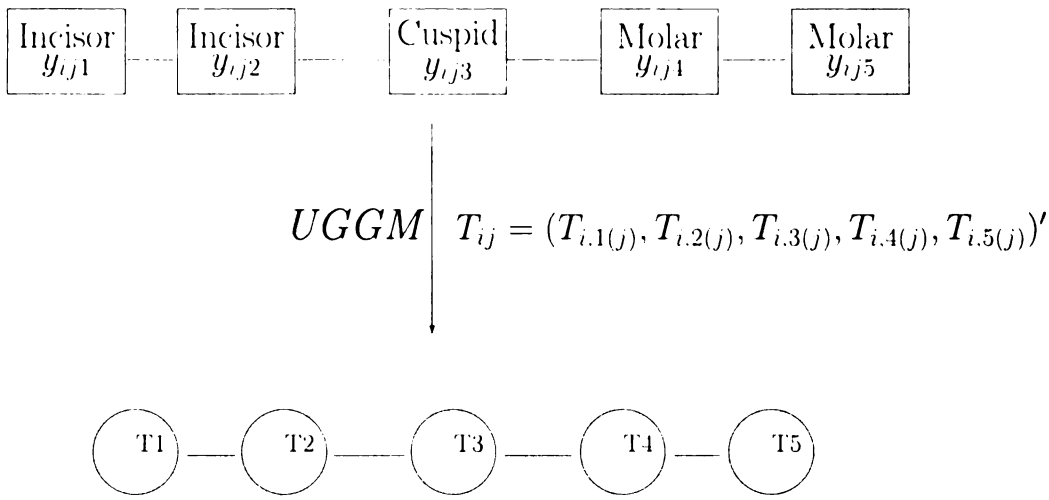

Figure 2.3. Note: The response variables y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4} and y_{ij5} are tighten by spatial latent vector $T_{ij} = (T_{i.1(j)}, T_{i.2(j)}, T_{i.3(j)}, T_{i.4(j)}, T_{i.5(j)})'$ whose joint distribution is given by UGGM with precision matrix under CAR model assumption.

dent given the other nodes in the graph. The graphical model, for the ith subject, is used to describe the spatial configuration of the nodes and characterize the association strength between nodes of interest by partial correlation of the corresponding between random variables $Q_i = (Q_{i1}, Q_{i2}, Q_i 3, Q_i 4)'$ that are assigned to the nodes. As matter of fact, in statistics, partial correlation measures the degree of association between two random variables, with the effect of a set of controlling random variables removed. We can assign an multivariate Gaussian distributed random vector to Q_i, i.,e., $Q_i \sim \mathcal{N}(\mathbf{0}, \Sigma_Q)$, which will lead to undirected graphical Gaussian model. After introducing latent variable vector Q, modeled by UGGM, the quadrants, i.e., the quadrant-wise response vectors $\{y_{ij} : j = 1, ..., J\}$, are conditionally mutually independent. Considering the nested spatial structure between quadrants and teeth within quadrants, it is necessary to notice that the nested dependence structure is essential to make the model valid for the problem of interest. The second level of spatial latent vectors, nested within the corresponding quadrant, need to be incorporated into the model. Similarly, within one specific quadrant, say the jth quadrant, a quadrant specific UGGM with random nodes $T_{ij} = (T_{i1(j)}, T_{i2(j)}, T_{i3(j)}, T_{i4(j)}, T_{i5(j)})'$ are introduced. The spatial associations among teeth within jth quadrant are induced by T_{ij}, which is mutually independent conditional on Q_i. Furthermore, we assume the $T_{ij} \sim \mathcal{N}(\mathbf{0}, \Sigma_T^j)$: $j = 1, ..., J$. After introducing latent variable vector T_{ij}, modeled by UGGM, the teeth within jth quadrant are conditionally mutually independent.

We know Gaussian random variables are determined by the first two moments. For the identifiability, we already assume the mean structures of the two levels of spatial latent variables are vectors of zeros, then the problem will become issues about the covariance structures. The general covariance matrix will be unstructured with symmetric and positive definite constraints. The unstructured covariance matrix can be simplified if we assume Markovian properties for the nodes, somehow as shown in the third graph. The Markovian type covariance matrix can be incorporated within

spatial statistics by CAR model (Cressie (1991)). The choice of the two types of covariance structure for the spatial latent vectors at tooth nested within quadrant level is made through model selection in Bayesian framework via Deviance Information Criterion(DIC) for missing data problem proposed by Celeux *et al.*(2001), which is an extension of the DIC introduced in Spiegelhalter *et al.*(2002) for Bayesian model selections.

Undirected Graphical Gaussian Model

In this section, we review the graphical Gaussian model (Dempster,1972) required for this paper. Let $G = (V, E)$ be an undirected graph with vertex set $V = \{1, ..., k, ..., K\}$ and edge set $E = \{e_{kk'} : k \neq k' = 1, ..., K\}$, where $e_{kk'} = 1$ or 0 according to whether vertices k and k', $1 \leq k \neq k' \leq K$ are directly connected in G or not. In the undirected graphical Gaussian model, the edges set describes the associate structures of the vertex set. Random vector is assigned to edges set to represent the association strength between corresponding vertexes. The undirected graphical Gaussian model consists of all k dimensional normal distribution, say $X = \{X_1, ..., X_k, ..., X_K\}$, with $X \sim N(\mathbf{0}, \Sigma)$ and precision matrix $\Omega = \Sigma^{-1} = \{\omega_{kk'} : k \neq k' = 1, ..., K\}$, where Σ is unknown but satisfies the following restrictions in terms of the pairwise conditional independencies determined by the Markov properties (Drton and Perlman (2004)):

$$e_{kk'} = 0 \Leftrightarrow X_k \perp X_{k'} | X_{V \setminus \{kk'\}} \Leftrightarrow \rho_{kk'} = 0; \quad \forall k \neq k' = 1, ..., K,$$

where $\{\rho_{kk'}\}$ is the so called the partial correlation between the kth and $k'th$ vertex in the graph, defined as $\rho_{kk'} = -\omega_{kk'}/\sqrt{\omega_{kk} * \omega_{k'k'}}$. This partial correlation is a measurement of association between two quadrants of interest with the effect of the rest quadrant being removed. We will use partial correlation to examine the spatial symmetry property of caries prevalence.

Conditional Autoregressive Models

For the vector of univariate variables $\nu = (\nu_1, \nu_2, ..., \nu_K)'$, the zero-centered CAR specification, where s is the number of spatial nodes of interest, following Cressie(1991), sets

$$(\nu_k|\nu_{-k}, \sigma_k^2) \sim N(\rho \sum_{\nu_{k'} \in \nu_{-k}} b_{kk'}\nu_k', \sigma_k^2); \;\; k = 1, ..., K, \tag{2.3}$$

where $\nu_{-k} = \nu \setminus \{\nu_k\}$. Following Brook's (1964) lemma the resulting joint density for ν takes the form

$$f(\nu|\sigma^2) \propto \exp\{-\frac{1}{2}\nu^T D_{\sigma^2}^{-1}(I - \rho B)\nu\} \tag{2.4}$$

where B is $K \times K$ matrix with $B = (b_{kk'})$ and $b_{kk} = 0$ and $D_{\sigma 2}$ us an $K \times K$ diagonal matrix with non-zero entries $\{\sigma_k^2 : k = 1, ..., K\}$. The precision matrix $D_{\sigma^2}^{-1}(I - \rho B)$ need to be symmetric, which yields the conditions

$$b_{kk'}\sigma_{k'}^2 = b_{k'k}\sigma_k^2: \;\; \forall k, k' = 1, ..., K. \tag{2.5}$$

If the precision matrix is positive definite, then (4) is a proper distribution. Under above parameterizations, the precision matrix $D_{\sigma^2}^{-1}(I - \rho B)$ is nonsingular if $\rho \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$ where $\lambda_{min}, \lambda_{max}$ are the smallest and largest eigenvalues of B respectively. It usually assumes that the $D_{\sigma 2} = \sigma^2 M$, where M is diagonal matrix with diagonal elements M_{kk} proportional to the conditional variance of σ_k^2. Meanwhile, σ^2 controls the overall variability and ρ represent the overall spatial association. Weights matrix B with $B_{kk'}$ need to reflect the spatial association between nodes k and k'. *GoeBUGS*(2004) sets $B_{kk'} = b_{kk'} = 1/n_k$, for $k \neq k'$ and $M_{kk} = 1/n_k$ where n_k is the number of nodes which is adjacent to node k. Under the above settings, the spatial latent vector ν will follow a proper distribution, i.e.

$$\nu \sim N(\mathbf{0}, \sigma^2(I - \rho B)^{-1}M) \tag{2.6}$$

2.4 Bayesian Estimations and Statistical Inference

2.4.1 Identifiability of the models

Frequently, models with latent variables are not globally identifiable. One can integrate out the latent spatial variable vectors to obtain a marginal likelihood to assess whether parameters are redundant. The likelihood of the latent variable model is parameterized by Σ_Q and $\{\Sigma_T^j : j = 1, ..., J\}$. The identifiability problem become to examine if the parameters involved in the covariance are redundant, which might be problematic within frequentist's framework. Dawid (1979) and Gelfand & Sahu (1999) discussed model identifiability issues within Bayesian framework. In particular, Suppose that the Bayesian model is denoted by the likelihood $L(\theta; y)$ and the prior $f(\theta)$ and we partition the parameters of interest as $\theta = (\theta_1, \theta_2)$. If

$$f(\theta_2|\theta_1, y) = f(\theta_2|\theta_1) \tag{2.7}$$

then we say that θ_2 is not identifiable, where $f(\theta_2|\theta_1, y) \propto L(\theta_1, \theta_2; y) f(\theta_2|\theta_1) f(\theta_1)$. That is, if observing data y does not increase our prior knowledge about θ_2 given θ_1, then θ_2 is not identifiable by the data. Dawid's formal definition of Bayesian model nonidentifiability states that θ_2 is not identifiable if and only if $L(\theta_1, \theta_2; y)$ is free of θ_2. In order to make our model identifiable, we need to not only take care of marginal identifiability of the model through integrating out the latent variables, but also put some constraints to the covariance matrix of the Gaussian spatial latent vectors at both levels.

2.4.2 Prior distributions

In this section, prior distributions are chosen for the regression parameters θ and association parameters ξ. Gibbs sampling algorithm is applied to simulate the samples from the posterior distributions of the quantities of interest. Zhao *et al.*(2006), Zeger *et al.*(1991) and Dunson *et al.*(2000) all suggested noninformative conjugate prior

distributions for the parameters of interest, which can wash out the effect of priors as sample size increases. Bedrick *et al.*(1996) noted that normal prior distributions were suggested for the logistic regression coefficient θ.

$$\theta \sim N(\mu, \mathbf{F}), \tag{2.8}$$

where μ is the a vector of location parameters and $\mathbf{F}$ is the covariance matrix. It is common to take μ as vector of zeros and $\mathbf{F}$ as diagonal matrix with very larger entries.

We are interested in the joint posterior distribution of $(\theta, \xi|y)$. Under mild condition in (Geman and Geman *et al.*(1984)), Gibbs sampler can obtain the joint posterior distribution by sampling from the conditional posterior distributions $(\theta|y, \xi)$ and $(\xi|y, \theta)$ correspondingly. To simplify the sampling from the conditional posterior distributions, we choose hierarchical independent priors for θ and ξ in this hierarchical Bayesian model, i.e. $(\xi|y, \theta) = (\xi|y)$, which is true as long as the priors satisfy $p(\theta, \xi) = p(\theta)p(\xi)$. We proposed two covariance structures for the Guassian spatial latent variable models. In the generalized linear model setting with Gaussian random effects, the proper noninformative conjugate priors will be Inverse Gamma(IG) for signal variance component and Inverse Wishart distribution for a variance-covariance matrix.

Let $\Omega_Q = \Sigma_Q^{-1}$ and $\{\Omega_{T_j} = \Sigma_{T_j}^{-1} : j = 1, ..., J\}$ denote the precision matrixes of the two levels of spatial latent vectors correspondingly. At higher level, the precision matrix for the spatial latent vector $\{Q_i : i = 1, ..., n\}$ is unstructured. Wishart priors (Dunson *et al.*(2000), O'Malley and Alan M. Zaslavsky *et al.*(2006)) is applied as conjugate non-informative priors for the precision matrix Ω_Q under unstructured situation.

$$\Omega_Q \sim Wishart(\upsilon_Q, \Lambda_Q), \tag{2.9}$$

with the degrees of υ_Q and the precision matrix Λ_Q. In practice, the common

noninformative Wishart prior is chosen by specifying $\Lambda_Q = I_{v_Q \times v_Q}$ and $v_Q = rank(\Sigma_Q)+1$. It will yield a prior under which the marginal distribution of each correlation parameter is $U(1,1)$(O'Malley *et al.*(2006)). At intermediate level, we have two precision matrix structure for the spatial latent vectors $\{T_{ij} : j = 1, ..., J, i = 1, ..., n\}$ and we will give noninformative priors correspondingly.

(1) Unstructured precision matrix in the UGGM: Conditional on the higher level spatial latent vector Q_i, the intermediate level spatial latent vectors $\{T_{ij} : j = 1, ..., J\}$ are conditionally independent. So, we give independent priors to the precision matrix $\{\Omega_{T_j} : j = 1, ..., J\}$. Similarly, independent Wishart processes are assigned as priors for these precision matrixes.

$$\Omega_{T_j} \sim Wishart(v_{T_j}, \Lambda_{T_j}): \quad j = 1, ..., J, \tag{2.10}$$

with the degrees of $v_{T_j} = rank(\Sigma_{T_j}) + 1$ and the precision matrix $\Lambda_{T_j} = I_{v_{T_j} \times v_{T_j}}$.

(2) CAR model based precision matrix in the UGGM: Conditional on the higher level spatial latent vector Q_i, the intermediate level spatial latent vectors $\{T_{ij} : j = 1, ..., J\}$ are conditionally independent. So, we give independent priors precision matrix $\{\Omega_{T_j}, : j = 1, ..., J\}$ that are parameterized by $\{\sigma_j^2, \rho_j : j = 1, ..., J\}$. Similarly, independent Inverse Gamma (Dunson *et al.*(2000)) distributions, proper conjugate priors, are assigned as priors to the overall variation parameters $\{\sigma_j^2 : j = 1, ..., J\}$ and independent uniform distribution with supports constraints in *section 3.1.4* to the overall spatial association parameters $\{\rho_j : j = 1, ..., J\}$, improper priors *GeoBUGS*(2004) for the over quadrant specific spatial association parameters, respectively.

$$\sigma_j^2 \sim IG(\varepsilon, \varepsilon): \quad j = 1, ..., J, \tag{2.11}$$

and

$$\rho_j \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1}): \quad j = 1, ..., J. \tag{2.12}$$

where ε is very small positive number and λ_{min}^{-1}, λ_{max}^{-1} are as defined in *section 3.1.4*.

2.4.3 Posterior computations

MCMC techniques are used for the posterior computations in the models proposed in section 3. The posterior distributions of parameters of interest can be obtained in standard way (Dunson *et al.*(2000), Zeger *et al.*(1991)).

Given the precision matrixes Ω_Q and $\{\Omega_{T_i} : i = 1, ..., I\}$, the joint posterior distribution for the regression parameters and latent variables at both higher and intermediate level is

$$\begin{aligned} p(\theta, Q, T|y) &\propto p(y|\theta, Q, T)\pi(\theta, Q, T) \\ &\propto \exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk} - b_i(\eta_{ijk})}{a_i(\varphi)} + c_i(y_{ijk}, \varphi)\} - \frac{1}{2}\theta' F^{-1}\theta\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} Q_i'\Omega_Q Q_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J} T_{ij}'\Omega_{T_i}T_{ij}\right\}, \end{aligned} \tag{2.13}$$

where $\sum_{i,j,k}$ denote $\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K}$, $\pi(\cdot)$ denote the joint prior density, $Q = (Q_1', ..., Q_i', ..., Q_n')'$ with $Q_i = (Q_{i1}, ..., Q_{ij}, ..., Q_{iJ})'$, $T = (T_1', ..., T_i', ..., T_n')'$ with $T_i = (T_{i1}', ..., T_{ij}', ..., T_{iJ}')'$ and $T_{ij} = (T_{i,1(j)}, ..., T_{i,k(j)}, ..., T_{i,K(j)})'$. Furthermore, $\theta = (\alpha, \beta', \gamma', \varphi)'$ and $\eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{i,k(j)}$.

If the MCMC algorithm is a Gibbs sampler, the full conditional distribution of each of the unknowns in (13) needs to be specified, which can be obtained in a standard way Dunson *et al.*(2000), Zeger*et al.*(1991)). For the fixed effect θ, the full conditional distribution is

$$p(\theta|Q, T, y) \propto \exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk} - b_i(\eta_{ijk})}{a_i(\varphi)} + c_i(y_{ijk}, \varphi)\} - \frac{1}{2}\theta' F^{-1}\theta\right\}. \tag{2.14}$$

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

$$p(Q_i|T, y; \theta) \propto \exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk} - b_i(\eta_{ijk})}{a_i(\varphi)} + c_i(y_{ijk}, \varphi)\} - \frac{1}{2}\sum_{i=1}^{n} Q_i'\Omega_Q Q_i\right\}. \tag{2.15}$$

The full conditional distribution for the Gaussian spatial latent vectors T_{ij} is

$$p(T_{ij}|Q,y;\theta) \propto \exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk}-b_i(\eta_{ijk})}{a_i(\varphi)}+c_i(y_{ijk},\varphi)\}-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}T'_{ij}\Omega_{T_i}T_{ij}\right\}. \tag{2.16}$$

The full conditional distributions of precision matrix Ω_Q is

$$p(\Omega_Q|Q,T,y;\theta) = Wishart(\upsilon_Q+N,\Lambda_Q+\sum_{i=1}^{n}Q_iQ'_i). \tag{2.17}$$

The posterior distributions for $\{\Omega_{T_j}, j=1,...,J\}$ can be obtained in terms of different precision matrix structures correspondingly.

(1) Unstructured precision matrix:

$$p(\Omega_{T_j}|Q,T,y;\theta) = Wishart(\upsilon_{T_i}+N,\Lambda_{T_j}+\sum_{n=1}^{N}T_{ij}T'_{ij}). \tag{2.18}$$

(2) CAR model based precision matrix:

(2.1) Overall precision parameters:

$$p(\tau_j|Q,T,y;\theta) \propto \tau_j^{\varepsilon-1}\exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk}-b_i(\eta_{ijk})}{a_i(\varphi)}+c_i(y_{i,j,k},\varphi)\}-\tau_j\varepsilon\right\}. \tag{2.19}$$

(2.2) Overall spatial parameters:

$$p(\rho_j|Q,T,y;\theta) \propto \exp\left\{\sum_{i,j,k}\{\frac{\eta_{ijk}y_{ijk}-b_i(\eta_{i,j,k})}{a_i(\varphi)}+c_i(y_{ijk},\varphi)\}\right\}I_{(\lambda_{min}^{-1},\lambda_{max}^{-1})}(\rho_j). \tag{2.20}$$

All the posterior distributions, except for $\{\rho_j : j=1,,...,J\}$, are proper based on their proper conjugate priors. The uniform priors for the overall spatial parameters are not conjugate, which might lead to improper posterior distributions. The simplest technique for verifying if the posterior distributions of the parameters is proper is to verify if the posterior distribution is proper for reduced data by discarding all but a single outcome per subject leading to a reduced data set consisting of independent outcomes, are proper (O'Brien and Dunson, 2004). Since the covariance structures

do not appear in the reduced data likelihood and also the support for the spatial association parameters is finite, i.e., $\{\rho_j \in (\lambda_{min}^{-1}, \lambda_{max}^{-1}) : j = 1, ..., J\}$, so the posterior distributions of the spatial association parameters $\{\rho_j : j = 1, ..., J\}$ are proper. The algorithm for the posterior computation is through sampling θ, Q, T, and ξ respectively from the above conditional distributions.

2.4.4 Missing data issue

In medical research, missing data is a very common problem. Little and Rubin *et al.*(2002) gave a comprehensive framework for dealing with missing data. We will follow their framework to incorporate missing data into our model. Let Y denote the data that would occur in the absence of missing values. we write $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} denotes the observed values and Y_{mis} denotes the missing values. Let $f(Y|\psi) \equiv f(Y_{obs}, Y_{mis}|\psi)$ denote the probability or density of the joint distribution of Y_{obs} and Y_{mis}. From frequentist's point of view, the inference is based on the marginal probability density of Y_{obs} is obtained by integrating out the missing data Y_{mis}:

$$f(Y_{obs}|\psi) = \int f(Y_{obs}, y_{mis}|\psi) dy_{mis}.$$

More generally, we define a *missing-data indicator* as follows:

$$M_{ijk} = \begin{cases} 1, & y_{ijk} \ missing, \\ 0, & y_{ijk} \ observed. \end{cases} \tag{2.21}$$

The full model treats M as random variables and specifies the joint distribution of M and Y. That is,

$$f(Y, M|\psi, \varrho) = f(Y|M; \psi) f(M|Y, \varrho), \qquad (\psi, \varrho) \in \Omega_{\psi,\varrho},$$

where $\Omega_{\psi,\varrho}$ is the parametric space of (ψ, ϱ).

The actual observed data is (Y_{obs}, M). The distribution of the observed data is obtained by integrating Y_{mis} out of the joint density of $Y = (Y_{obs}, Y_{mis})$ and M. That is,

$$f(Y_{obs}, M|\psi, \varrho) = \int f(Y_{obs}, y_{mis}|M; \psi) f(M|Y_{obs}, y_{mis}, \varrho) dy_{mis}. \tag{2.22}$$

The full likelihood of (ψ, ϱ) is proportional to the above, i.e.

$$L_{full}(\psi, \varrho|Y_{obs}, M) \propto f(Y_{obs}, M|\psi, \varrho). \tag{2.23}$$

If the the distribution of missing-data mechanism does not depend on the missing values Y_{mis}, then

$$\begin{aligned} f(Y_{obs}, M|\psi, \varrho) &= f(M|Y_{obs}, \varrho) \int f(Y_{obs}, y_{mis}|M; \psi) dy_{mis} \\ &= f(M|Y_{obs}, \varrho) f(Y_{obs}|M; \psi). \end{aligned} \tag{2.24}$$

Under MAR (missing at random) assumption and ψ and ϱ are distinct, the likelihood-based inferences for ψ will be the same as likelihood-based inferences for ψ from $f(Y_{obs}|\psi)$.

From Bayesian point of view, missing data is treated as random as well as the parameters of interest. One of the advantages of the Bayesian hierarchical approach implemented in *WinBUGS* is that missing data from the response variables can be routinely handled. In most statistical packages, incomplete cases (in either the response or the covariates) are removed from any analysis. *WinBUGS* generates a sample to replace missing responses from the posterior distribution of the response variable under MAR assumption.

2.4.5 Bayesian Model Selection

The formal procedure for choosing an appropriate Bayesian hierarchical model for the observed data necessities methods to compare alternative models within Bayesian

framework. The DIC (deviance information criterion, Spiegelhalter *et al.* (2002)) is a hierarchical modeling selection criterion that can be viewed as a generalization of the AIC (Akaike information criterion,Akaike, 1973) and BIC (Bayesian information criterion, Kass and Raftery, 1995). It is particularly useful in Bayesian model selection problems where the posterior distributions of parameters have been obtained by Markov chain Monte Carlo (MCMC) simulation. The DIC-statistic is a measure of model complexity and goodness of fit with the definition as

$$DIC = \overline{D(\psi)} + pD,$$

where $\overline{D(\psi)}$ is the posterior mean of the deviance $D(\psi) = -2\log(f(y|\psi))$, which is a measurement of goodness of fit of the proposed model for the observed data. Let $D(\overline{\psi})$ be the deviance evaluated at the posterior mean of ψ. Let $pD = \overline{D(\psi)} - D(\overline{\psi})$ denote the effective number of parameters in the model, which is a penalty for the complexity of the model. The quantities $\overline{D(\psi)}$ and $D(\overline{\psi})$ can be obtained routinely from an MCMC simulation chain. Our hierarchial models contains two levels of latent variables, which necessitates the model selections to be based on the DIC for missing data problems (Celeux *et al.*,2006). In terms of our problem, we have to deal with both missing data and latent variables to get a complete DICs. In order to deal with missing data, we consider the complete likelihood (21) and the deviance function has the form

$$\begin{aligned} D(o) &= -2\log\left\{f(Y_{obs}, M|\psi, \varrho)\right\} \\ &= -2\log\left\{\int f(Y_{obs}, y_{mis}|M;\psi) f(M|Y_{obs}, y_{mis}, \varrho) dy_{mis}\right\}, \end{aligned} \tag{2.25}$$

where $o = (\psi', \varrho')'$. Pettitt *et al.* (2006) gave an approximation for (21) in the form

$$D(o) = -2\log\left\{f(y_{obs}, \widehat{y_{mis}}|M;\psi) f(M|y_{obs}, \widehat{y_{mis}}, \varrho)\right\}, \tag{2.26}$$

where $\widehat{y_{mis}}$ is the posterior predictor for the missing data Y_{mis}.

In order to deal with the latent vectors, we need to compute the complete DICs in Celeux *et al.* (2006). Let $E_{\psi}[\theta|y, q, t]$ 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

$$DIC(y,q,t) = -4E_{\psi}[\log(f(y,q,t|\psi))|y,q,t] + 2\log(f(y,q,t|E_{\psi}[\psi|y,q,t])). \quad (2.27)$$

As in the EM algorithm, we can then integrate Q and T out from (26) to get

$$\begin{aligned} DIC &= E_{Q,T}[DIC(y,Q,T)|y] \\ &= -4E_{\psi,Q,T}[\log(f(y,Q,T|\psi))|y] + 2E_{Q,T}[\log(f(y,Q,T|E_{\psi}[\psi|y,Q,T]))|y]. \end{aligned} \quad (2.28)$$

All the integrations can be obtained trivially through Monte Carlo integration approximation using the MCMC posterior samples in the coda file of *WinBUGS*.

Combining (2.25)-(2.28), we have the DIC for Bayesian generalized latent variable models with missing data in the form.

$$\begin{aligned} DIC &= E_{Q,T}[DIC(y_{obs}, \widehat{y_{mis}}, M, Q, T)|y_{obs}, M] \\ &= -4E_{\psi,\varrho,Q,T}[\log(f(y_{obs}, \widehat{y_{mis}}, Q, T|\psi, M, \varrho)f(M|y_{obs}, \widehat{y_{mis}}, \varrho))|y_{obs}, M] \\ &\quad +2E_{Q,T}[\log(f(y_{obs}, \widehat{y_{mis}}, Q, T|\widetilde{\psi}, \widetilde{\varrho})f(M|y_{obs}, \widehat{y_{mis}}, \widetilde{\varrho}))|y_{obs}, M], \end{aligned} \quad (2.29)$$

where $(\widetilde{\psi}) = E_{\psi}[\psi|y_{obs}, M, Q, T]$ and $\widetilde{\varrho} = E_{\varrho}[\varrho|y_{obs}, M]$.

2.4.6 Spatial symmetry hypothesis testing

The spatial symmetry property in our problem means the joint caries experience presentations for response variables at quadrant level are highly associated with one another. Dentists do believe that spatial symmetry exist in mouth. Lesaffre *et al.*(2006) showed empirically that the caries experience for left and right quadrants are more strongly associated than the other cases. Unfortunately, few literatures have discussed this issue comprehensively. In our UGGM at quadrant level, we know the partial correlation parameters $\{\rho_{jj'} : j \neq j' = 1, ..., J\}$ measures the strength of the spatial association among two different nodes(quadrants). One of the major concerns of the spatial symmetry in mouth can be formulated as the following hypothesis situation:

Hypothesis testing for pairwise comparisons among spatial association strength parameters

In order to assess the spatial symmetry of the four quadrants, we need to introduce different "Neighborhoods" relationships that can explain the relative spatial structures of the quadrants of interest. Spatial symmetry is assessed at the quadrant level, instead of tooth level. At quadrant level, We define the vector of teeth to be "Horizontal Neighbors" to each other, if the two quadrants are both in either "Upper Jaw" or "Lower Jaw", and to be "Vertical Neighbors" to one another, if the two quadrants are both in either "Left Jaw" or "Right Jaw" and to be "Across Neighbors" to one another, if the two quadrants are either in "Left Jaw" or "Right Jaw". The assessment of quadrant spatial symmetry in terms of cries prevalence will be based on "Left-right", i.e., " Horizontal Neighbors", "Up-down", i.e.,"Vertical Neighbors" and "Across", i.e., "Across Neighbors".

There are two ways to assess the spatial symmetry among quadrants in terms of caries prevalence incidence through statistical hypothesis statement. The first one is based on the so called "overall" spatial symmetry assessments via a weighted statistic and the second is the so called "specific" spatial symmetry assessment that is the direct comparisons of the spatial symmetry measurements.

First of all, the weighted statistics for assessing the overall spatial associations in terms of "Left-right", "Up-down" and "Across" can be formulated as below:

$$\rho_{LR} = \frac{1}{2}(\rho_{56} + \rho_{78});$$

$$\rho_{UD} = \frac{1}{2}(\rho_{67} + \rho_{58});$$

$$\rho_{A} = \frac{1}{2}(\rho_{68} + \rho_{57}).$$

The statistical hypothesis testing about the overall spatial association in terms of "Left-right" V.S. "Up-down", "Left-right" V.S. "Across" and "Across" V.S. "Up-down" can be formulated as follows:

(1) Left-right Versus Up-down

$$H_0 : \rho_{LR} = \rho_{UD} \qquad V.S. \qquad H_a : \rho_{LR} \neq \rho_{UD}; \tag{2.30}$$

(2) Left-right Versus Across

$$H_0 : \rho_{LR} = \rho_{A} \qquad V.S. \qquad H_a : \rho_{LR} \neq \rho_{A}; \tag{2.31}$$

(3) Across Versus Up-down

$$H_0 : \rho_{A} = \rho_{UD} \qquad V.S. \qquad H_a : \rho_{A} \neq \rho_{UD}. \tag{2.32}$$

Secondly, if the assessment is based on the direct comparisons of spatial symmetry measurement, there are twelve possible hypothesis testing situations for the spatial symmetries in terms of partial correlation between quadrants.

(1.1) Left-right Versus Up-down The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 6 and quadrant 7, with quadrant 6 as reference.

$$H_0 : \rho_{56} = \rho_{67} \qquad V.S. \qquad H_a : \rho_{56} \neq \rho_{67}; \tag{2.33}$$

(1.2) Left-right Versus Up-down The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 5 and quadrant 8, with quadrant 5 as reference.

$$H_0 : \rho_{56} = \rho_{58} \qquad V.S. \qquad H_a : \rho_{56} \neq \rho_{58}; \tag{2.34}$$

(1.3) Left-right Versus Up-down The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 6 and quadrant 7, with quadrant 7 as reference.

$$H_0 : \rho_{78} = \rho_{67} \qquad V.S. \qquad H_a : \rho_{78} \neq \rho_{67}; \tag{2.35}$$

(1.4) Left-right Versus Up-down The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 5 and quadrant 8, with quadrant 8 as reference.

$$H_0 : \rho_{78} = \rho_{58} \qquad V.S. \qquad H_a : \rho_{78} \neq \rho_{58}; \tag{2.36}$$

(2.1) Left-right Versus Across The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 6 and quadrant 8, with quadrant 6 as reference.

$$H_0 : \rho_{56} = \rho_{68} \qquad V.S. \qquad H_a : \rho_{56} \neq \rho_{68}; \tag{2.37}$$

(2.2) Left-right Versus Across The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 5 and quadrant 7, with quadrant 5 as reference.

$$H_0 : \rho_{56} = \rho_{57} \qquad V.S. \qquad H_a : \rho_{56} \neq \rho_{57}; \tag{2.38}$$

(2.3) Left-right Versus Across The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 6 and quadrant 8, with quadrant 8 as reference.

$$H_0 : \rho_{78} = \rho_{68} \qquad V.S. \qquad H_a : \rho_{78} \neq \rho_{68}; \tag{2.39}$$

(2.4) Left-right Versus Across The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 5 and quadrant 7, with quadrant 7 as reference.

$$H_0 : \rho_{78} = \rho_{57} \qquad V.S. \qquad H_a : \rho_{78} \neq \rho_{57}; \tag{2.40}$$

(3.1) Across Versus Up-down The association between quadrant 5 and quadrant 7 V.S. the association between quadrant 5 and quadrant 8, with quadrant 5 as reference.

$$H_0 : \rho_{57} = \rho_{58} \qquad V.S. \qquad H_a : \rho_{57} \neq \rho_{58}; \tag{2.41}$$

(3.2) Across Versus Up-down The association between quadrant 5 and quadrant 7 V.S. the association between quadrant 6 and quadrant 7, with quadrant 7 as reference.

$$H_0 : \rho_{57} = \rho_{67} \qquad V.S. \qquad H_a : \rho_{57} \neq \rho_{67}; \tag{2.42}$$

(3.3) Across Versus Up-down The association between quadrant 6 and quadrant 8 V.S. the association between quadrant 5 and quadrant 8, with quadrant 8 as reference.

$$H_0 : \rho_{68} = \rho_{58} \qquad V.S. \qquad H_a : \rho_{68} \neq \rho_{58}; \tag{2.43}$$

(3.4) Across Versus Up-down The association between quadrant 6 and quadrant 8 V.S. the association between quadrant 6 and quadrant 7, with quadrant 6 as reference.

$$H_0 : \rho_{68} = \rho_{67} \qquad V.S. \qquad H_a : \rho_{68} \neq \rho_{67}. \tag{2.44}$$

Simultaneous credible intervals

Pairwise spatial symmetry hypothesis testing is based on credible intervals for the differences between two partial correlations corresponding to two different nodes (quadrants) in the UGGM. In Bayesian statistics, a credible interval is a posterior probability interval, used for purposes similar to those of confidence intervals in frequentist statistics. Suppose that parameter ς is of interest, a $(1-\alpha)100\%$ credible interval for the parameter ς of interest is any set C such that $P_{\pi(\varsigma|y)}(\varsigma \in C) = 1 - \alpha$, where $\pi(\varsigma|y)$ is the posterior distribution of parameter ς given the observed data y. There are two ways to assess the spatial symmetry among quadrants in terms of caries prevalence

incidence through statistical hypothesis statement. The first one is based on the so called "overall" spatial symmetry assessments via a weighted statistic and the second is the so called "specific" spatial symmetry assessment that is the direct comparisons of the spatial symmetry measurements.

Since we are performing a multiple spatial symmetry comparisons among quadrants in terms of all possible hypothesis testing situations, it is necessary to give a simultaneous credible regions (Besag *et al.* (1995)) to control type S error rate (Gelman *et al.*), i.e., the similar concept as type I error rate in frequentist's framework. The $100K/M\%$ simultaneous credible regions for overall spatial associations differences are based on order statistics (Besag *et al.* (1995))

$$\left\{[(\rho_I - \rho_{II})^{[M+1-t^*]}, (\rho_I - \rho_{II})^{[t^*]}] : (I, II) \in Neighborhood\right\},$$

where

$$t^* = \min\{t : \#\left\{(\rho_I - \rho_{II})^{[M+1-t^*]} \leq (\rho_I - \rho_{II})^{(t)} \leq (\rho_I - \rho_{II})^{[t^*]}\right\} \geq K\},$$

and $\{(\rho_I - \rho_{II})^{(t)} : t = 1, ..., M, (I, II) \in Neighborhood\}$ are the posterior samples of $\{(\rho_I - \rho_{II}) : (I, II) \in Neighborhood\}$. Here, $Neighborhood = \{("LR", "UD"), ("LR", "A"), ("A", "UD")\}$.

Similarly, the $100K/M\%$ simultaneous credible regions for specific spatial associations difference are given by

$$\{[(\rho_{ii'} - \rho_{jj'})^{[M+1-t^*]}, (\rho_{ii'} - \rho_{jj'})^{[t^*]}] : i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\},$$

where

$$t^* = \min\{t : \#\left\{(\rho_{i,i'} - \rho_{jj'})^{[M+1-t^*]} \leq (\rho_{ii'} - \rho_{jj'})^{(t)} \leq (\rho_{ii'} - \rho_{jj'})^{[t^*]}\right\} \geq K\},$$

and $\{(\rho_{ii'} - \rho_{jj'})^{(t)} : t = 1, ..., M, i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\}$ are the posterior samples of $\{\rho_{ii'} - \rho_{jj'} : i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\}$.

2.4.7 Example

Now we show how the above methodology works for dental data and need to specify all the functions and general notations. In our study, all of the responses are binary, so we have the following: $a_i(\varphi) = 1$, $b_i(\eta_{ijk}) = \log(1 + \exp(\eta_{ijk}))$, $c_i(y_{ijk}, \varphi) = 0$, $E[y_{ijk}|\eta_{ijk}] = \frac{\exp(\eta_{ijk})}{1+\exp(\eta_{ijk})}$, $g(x) = \log(\frac{x}{1-x})$, for $k = 1, ..., 5, j = 1, ..., 4, i = 1, ..., n$. Hence, the parameters of interest in the observational model is $\theta = (\alpha, \beta', \gamma')'$ and $\xi = \xi^{-1}(\Sigma_Q, \Sigma_T^1, ..., \Sigma_T^j, ..., \Sigma_T^J)'$, then $\log p(y_{ijk}|\eta_{ijk}) = \log p(y_{ijk}|Q_i, T_{i,k(j)}, \theta) = \eta_{ijk}y_{ijk} - \log(1 + \exp(\eta_{ijk}))$. The canonical parameter $\{\eta_{ijk} : k = 1, .., 5, j = 1, ..., 4, i = 1, ..., n\}$ is defined as follows:

$$\eta_{ijk} = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{i,k(j)}. \tag{2.45}$$

Priors for parameters of interest are given by noninformative proper conjugate priors, which will give comparable results as frequentist's as the sample size increases. More specifically, the priors are given as follows:

$$\alpha \sim N(0, 1000); \tag{2.46}$$

$$\beta_j \sim N(0, 1000); \ \forall j = 1, ..., J - 1, \tag{2.47}$$

$$\gamma_{k(j)} \sim N(0, 1000): \ \forall k = 1, ..., K - 1, j = 1, ..., J; \tag{2.48}$$

with constraints $\sum_j \beta_j = 0$ and $\sum_k \gamma_{k(j)} = 0$, $j = 1, ..., J$ for identifiability of the observation model. For the priors of precision matrix, O'Malley and Zaslavsky (2006) proposed scaled Wishart distribution as conjugate proper priors

$$\Omega_Q \sim Wishart(4 + 1, I), \tag{2.49}$$

where I is 4×4 identity matrix.

For the priors of the precision matrix $\{\Omega_{T_j} : j = 1, ..., J\}$, there are two different models for the the structures of the precision matrix.

(1) Unstructured precision matrix:

$$\Omega_{Tj} \sim Wishart(5 + 1, II), \quad \forall j = 1, ..., 4, \tag{2.50}$$

where II is 5×5 identity matrix.

(2) CAR model based precision matrix:

$$\tau_j \sim Gamma(0.001, 0.001); \quad \forall j = 1, ..., 4, \tag{2.51}$$

$$\rho_j \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1}). \quad \forall j = 1, ..., 4, \tag{2.52}$$

where $\{\sigma_j^{-2} = \tau_j : j = 1, ..., 4\}$ are the quadrant specific parameters for overall variability and $\{\rho_i : j = 1, ..., 4\}$ are the quadrant specific parameters for overall spatial effects. λ_{min} and λ_{max} are as defined in CAR models in *section 3.1.4*.

To construct 95% simultaneous credible regions, we use 11,000 MCMC iterations with 1000 burn in, i.e., $M = 10,000$ and $K = 9,500$. The 95% simultaneous credible regions are more convertive simultaneous confidence regions than frenquentist's for the multiple hypothesis statements since they have a type S error rate between 0% and 2.5% (Gelman *et al.*).

2.5 The Signal Tandmobiel Project Example

In the Signal-Tandmobiel project, there are 4,468 7-year-old schoolchildren (born in 1989) from 179 schools in Flanders (Belgium) who were selected by a stratified clustered random sample. The mean age of the children on the day of examination was 7.1 years (SD = 0.4). The 15 strata were obtained by combining the 3 types of educational system (public, municipal and private schools) with geographical areas (the

Table 2.1. Prevalence of caries experience(% affected) in the deciduous dentition of 7-year-old children n=4.351.

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

5 Flemish provinces). The schools represented the clusters. This sample represents about 7% of the corresponding Flemish population. The sampling procedure aimed at selecting each child in Flanders with equal probability. A more detailed description of the design of the Signal-Tandmobiel project is reported in Vanobbergen *et al.* (2000).

2.5.1 Primary results

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 suggested a spatial symmetrical pattern in terms of caries experience.

In Vanobbergen *et al.* (2007), pairwise associations were assessed in terms of odds ratio of caries experience via ALR model. The results are shown in table 2.

The above result shows that it is left-right spatial symmetry is the most notable. Decayed teeth of discordant contralateral pairs tend to aggregate on the right or the left side of the subject's mouth than would be expected by chance alone (Vanobbergen *et al.*(2007)).

Table 2.2. Odds ratios and 95% confidence intervals for the 2×2 association models for caries on deciduous molars on tooth in 7-year-old children.

First Molar (ALR model)			
54	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)			
55	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)

2.5.2 The results from our approach

Our generalized latent variable models are implemented in *WinBUGS*, using noninformative priors for the parameters of interest. After 1,000 burn-in, the posterior distributions of the quantities are based on 10,000 MCMC iterations. There are two possible models indexed by the precision matrix structure for spatial latent vector at intermediate level. The choice for appropriate model is based on the DIC for missing data problem (Celexu *et al.*(2006)). In this part, we will give the results for both overall and specific spatial symmetries assessment through simultaneous credible regions for the differences of interest. The results start from the overall spatial symmetry assessment under different model assumptions, in table 3-4, based on 95% simultaneous credible regions of the differences that are corresponding to their hypothesis testing situations. It was then followed by the results for specific spatial symmetry assessments in table 5-6.

Based on the results from two different models, the posterior inferences about the spatial symmetries are similar, which tells us both models work fairly well. Bayesian

Table 2.3. Credible intervals of overall spatial association strength comparisons Based on UGGM with unstructured covariance structure

Simultaneous Spatial Effects	Credible intervals
left/righ .v.s. across	
$\rho_{LR} - \rho_A$	(0.807, 1.238)
left/righ .v.s. upper/down	
$\rho_{LR} - \rho_{UD}$	(0.312, 1.471)
across .v.s. upper/down	
$\rho_A - \rho_{UD}$	(-0.775, 0.491)
DIC	593.300
N.burnin	1000
N.interation	11000

Table 2.4. Credible intervals of overall spatial association strength comparisons Based on UGGM with CAR model based covariance structure

Simultaneous Spatial Effects	Credible intervals
left/righ .v.s. across	
$\rho_{LR} - \rho_A$	(0.807, 1.236)
left/righ .v.s. upper/down	
$\rho_{LR} - \rho_{UD}$	(0.310, 1.427)
across .v.s. upper/down	
$\rho_A - \rho_{UD}$	(-0.779, 0.410)
DIC	780.500
N.burnin	1000
N.interation	11000

Table 2.5. Credible intervals of specific spatial association strength comparisons Based on UGGM with unstructured covariance structure

Simultaneous Spatial Effects	Credible intervals
left/righ .v.s. across	
$\rho_{56} - \rho_{68}$	(0.134, 1.581)
$\rho_{56} - \rho_{57}$	(0.394, 1.711)
$\rho_{78} - \rho_{68}$	(0.237, 1.589)
$\rho_{78} - \rho_{57}$	(0.133, 1.728)
left/righ .v.s. upper/down	
$\rho_{56} - \rho_{67}$	(0.235, 1.551)
$\rho_{56} - \rho_{58}$	(0.117, 1.485)
$\rho_{78} - \rho_{67}$	(0.230, 1.601)
$\rho_{78} - \rho_{58}$	(0.215, 1.504)
across .v.s. upper/down	
$\rho_{68} - \rho_{67}$	(-1.303, 1.313)
$\rho_{68} - \rho_{58}$	(-1.327, 1.204)
$\rho_{57} - \rho_{67}$	(-1.442, 1.109)
$\rho_{57} - \rho_{58}$	(-1.488, 1.042)
DIC	593.300
N.burnin	1000
N.interation	11000

Table 2.6. Credible intervals of specific spatial association strength comparisons Based on UGGM with CAR model based covariance structure

Spatial Effects	Credible intervals
left/righ .v.s. across	
$\rho_{56} - \rho_{68}$	(0.068, 1.404)
$\rho_{56} - \rho_{57}$	(0.445, 1.601)
$\rho_{78} - \rho_{68}$	(0.236, 1.416)
$\rho_{78} - \rho_{57}$	(0.477, 1.662)
left/righ .v.s. upper/down	
$\rho_{56} - \rho_{67}$	(0.297, 1.458)
$\rho_{56} - \rho_{58}$	(0.067, 1.455)
$\rho_{78} - \rho_{67}$	(0.291, 1.524)
$\rho_{78} - \rho_{58}$	(0.262, 1.450)
across .v.s. upper/down	
$\rho_{68} - \rho_{67}$	(-1.020, 1.209)
$\rho_{68} - \rho_{58}$	(-1.078, 1.146)
$\rho_{57} - \rho_{67}$	(-1.258, 0.970)
$\rho_{57} - \rho_{58}$	(-1.381, 0.950)
DIC	780.500
N.burnin	1000
N.interation	11000

model selection is based on DICs, the smaller the DIC, the better the model. It is common in practice that if the difference between the DICs of two different models are more than 10 then the model with smaller DIC is the better one. Hence, from the results from table 3 to table 6, conditional on the data, the model with unstructured precision matrix is the better one. Specifically, the appropriate hierarchical generalized latent variable model consist of two levels of latent vectors. The first level of Gaussian spatial latent vector has unstructured precision matrix. The second level of Gaussian spatial latent vectors also have unstructured precision matrix. Furthermore, the choice for the unstructured covariance structure can be explained by the following two facts. (1) The oral biological environment is so complected that the higher level Gaussian spatial latent vector might not be able to account for the heterogeneity from four quadrant-wise response vectors sufficiently and leave some residuals to the intermediate level spatial latent vectors. (2) At intermediate level, Gaussian spatial latent vector with CAR model based precision matrix are not sophisticated to account for both the residual heterogeneity and the one from the teeth within corresponding quadrants. Hence, the second level of Gaussian spatial latent vectors need more complicated precision matrix than Markovian type (CAR model based covariance matrix). Based on the chosen model, the conclusion of the hypothesis testing about both overall and specific spatial symmetry among quadrants are as follows: (1) 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 with lower bounds are all positive. (2) The difference of spatial associations between across and up-down is not significant at type S error rate between 0% and 2.5% (Gelman (2006)), since 95% simultaneous credible intervals of the difference between across spatial association and up-down spatial association includes zero.

2.6 Discussion

In this chapter, we propose a flexible class of Bayesian Generalized latent variable models for multivariate spatially correlated binary data with multi-level dependence structure. Our approach is to model the response variables by distributions in the exponential family and impose a multivariate spatial correlation structure on the latent variables, which accounts for the multi-level spatial dependence structures. Statistical inference is based on posterior sampling from the posterior distributions of the parameters of interest. We have used undirected graphical Gaussian model(UGGM) for constructing the precision matrix structures of multivariate spatial latent vectors at both higher and intermediate levels. One consideration is the parameterizations of both the observational and latent variable models, for the identifiability of the model, we constrain sum to zero for the fixed effects and the spatial process has mean zeros. Noninformative conjugate priors are applied for the parameters of interest, which will give a comparable inference results to the frequentist's as the sample size increases. We proposed two possible models to account for the dependence structure in the dental data. Bayesian model selection is based on DIC for missing data problem. Spatial symmetry hypothesis is assessed by simultaneous credible intervals for multiple comparisons of pairwise spatial association strength. The results from both models show the generalized latent variables model work well and consistent to one another and also comparable to the results in existing literatures. It concluded that the left-right spatial association is the strongest and the spatial associations for across and up-down are not different significantly at type S error rate between 0% and 2.5%. For the data example, we have assumed that the Gaussian spatial latent process $\{Q_i : i = 1, ..., n\}$ at higher level and $\{T_{ij} : j = 1, ..., J, i = 1, ..., n\}$ at intermediate level are sufficient to induce the unobserved heterogeneities from the data at corresponding levels. It would be interesting to introduce non-Gaussian latent process to model the underlying spatial dependence among quadrants and teeth nested within the corresponding

quadrant, which can lead to a richer class of the latent processes $\{Q_i : i = 1, ..., n\}$ and $\{T_{ij} : j = 1, ..., J, i = 1, ..., n\}$. Finally, our model selection is based on DIC and it will be optimal when the model selection is simultaneous through Reversible Jump Monte Carlo Markov Chain(RJMCMC) (Green (1995)) or Birth and Death Monte Carlo Markov Chain(BDMCMC)(Stephens (2000)) . It will be more interesting to consider the symmetry pattern of quadrants for a longitudinal study, which will lead to the spatial-temporal analysis.

CHAPTER 3

Bayesian Finite Mixture of Generalized Latent Variable Models

3.1 Introduction

As we have noticed in the above chapter that the dental showed a unique nested dependence structure among the caries experience response variables for the teeth of interest, which lead to a wide heterogeneity of distribution for the multivariate spatially correlated binary response variables. Finite mixture of distributions have provided a mathematical-based approach to model various random phenomena with the flexible distribution. It is obvious that mixture distributions are extremely useful in the modeling of heterogeneity in a cluster analysis context. It is of great interest that we can view the quadrant-wise multivariate binary response vectors as from a certain number of *underlying* subpopulation or clusters. Each of the underlying cluster is characterized by the corresponding underlying cluster-specific parameters and some common parameter to describe the marginal distribution of the binary response

variable with respect to the spatial configurations for each quadrant-wise response vector. The spatial symmetry among quadrants, in terms of the caries prevalence, can be measured by the probabilities that two different quadrant-wise response vectors will fall into the same underlying cluster that is indexed by a corresponding cluster-specific multivariate distribution.

Zhang *et al.* (2007) proposed a Bayesian Generalized Latent Variable Model (BGLVM) to analyze the dental data from the STM project. Their approach used a hierarchical generalized latent variable model to take care of the multiple level nested dependence structure of the dental data. The multiple level spatial latent variables are used to generate a flexible multivariate distribution for the multivariate binary outcomes and induce the unique nested dependence structure. The joint behavior of the multiple level spatial latent variables are described by Gaussian undirected graphical model with different ways to account for the covariance matrix structures. Spatial symmetry checking was based on the partial correlation parameters of the graphical models. Model implement and hypothesis testing are within Bayesian framework. Since we know mixture model is very flexible method of modeling, it is interesting to view the same problem from the mixture model point of view in stead of generalized latent variable model. It is also very helpful to give a general framework to use mixture model for analyzing spatially correlated multivariate binary data.

Fernández and Green *et al.* (2002) proposed a Bayesian mixture model to analyze spatial correlated data, which gives an appropriate approach in the case of finite, typically irregular, patterns of points or regions with prescribed spatial relationships. The spatial association strength was assessed through parameters that are used to adjust the variability of mixing weights in the mixture from one location to another. Their approach is sensitive to Euclidian space, and can not take care of multi-level correlations induced by both "between-cluster" and "within-cluster" spatial configuration of the data. Fernández and Green focused specifically on Poisson distributed

data with applications in disease mapping, which are quite different from the situation what the dental data are facing. For the estimation of the true risk pattern, their approach is based on a continuously distributed Markov random fields to model the mixture weight for the corresponding component via logit-normal model. They did not consider other mixture components that can yield flexible distributions for the outcomes and induce complex heterogeneity structure. However, their approach introduced spatial mixture models as an interesting new tool for those modeling heterogeneity in spatial data. Zhou and Wakefield *et al.* (2006) proposed a Bayesian mixture model for partitioning gene expression data, which is essentially an approach of clustering the observed data by a mixture model with unknown number of cluster inferred by the data. The aim of their research in which time ordered gene expression data are collected is to determine genes that co-express, that is, have similar patterns of expression, which provided a probabilistic framework for partitioning or clustering, which naturally provides a measure of similarity among genes in terms of expression. Under their approach, partitioning and estimation are conducted simultaneously, and the number of partitions can be treated as a random parameter, which will give the method a certain flexibility in applications. It is noticeable that as always for parametric hierarchial modeling, the measures of uncertainty are only as reliable as the model, so extensive model checking should be carried out in applications. It is necessary to give flexibility to the mixture components rather than as what they did via modeling a marginal parametric mean structure. Extension needs to incorporate covariates at various stages and other external information need to be taken into account. It is also meaningful to give the framework for analyzing non-normal data under mixture models for clustering.

The purpose of this article is to introduce a Bayesian Mixture of Generalized Latent variable Model (BMGLVM) framework for general spatial topology structures to explain multi-level correlations. The BMGLVM, implemented via Gibbs sampling

with non-informative priors, allows us to model the "between-cluster" and "within-cluster" correlation structure explicitly. It is possible for us to examine the spatial symmetry of quadrants in terms of caries incidence, and capture the special spatial association structure among quadrants for the same subject of interest and among teeth within quadrants, which can help us efficiently characterize the pattern of caries incidence at tooth level.

3.2 The Spatial Dependence Structures

3.2.1 Notation

To model the observations, let y_{ijk} denote the kth response variable within jth cluster of ith subject of interest, where $k = 1, ..., K, j = 1, ..., J, i = 1, ..., n$. Let $y_{ij} = (y_{ij1}, ...y_{ijk}, ..., y_{ijK})'$ denote the response vector within jth cluster of ith subject. Let $y_i = (y'_{i1}, ..., y_{ij}, ..., y'_{iJ})'$ denote the collection of response variables of ith subject. let $y = (y'_1, ..., y'_i, ..., y'_n)'$ denote the collection of response variables of all subjects in this study.

A multinomial model is applied for the allocation process associated with mixture models. let $Q_i = (Q'_{i1}, ..., Q_{ij}, ..., Q'_{iJ})'$ denote the mixture component allocation random variables for the ith subject, where $Q_{ij} = (Q_{ij1}, ..., Q_{ijm}, ..., Q_{ijM})'$ and M is the number of mixture components in the mixture model, for $i = 1, ..., n$ and $j = 1, ..., J$. It is assumed that Q_{ij}'s are identically independently multinomial distributed. For modeling the latent variables, we use conditional autoregressive model. Let $T_{im} = (T_{i,1(m)}, ..., T_{i,k(m)}, ..., T_{i,K(m)})'$ denote the latent variables associated with the mth mixture components for the ith subject of interest. Let $T_i = (T'_{i1}, ..., T'_{im}, ..., T'_{iM})'$ denote the collection of latent variables at intermediate level for the ith subject in the study. Let $L = \{Q'_i, T'_i : i = 1, ..., n\}$ denote the collection of all allocation random variables and latent variables for all subjects.

3.2.2 Principles of our modeling approach

The dental data shows a two-level spatial association structures, i.e., the first level spatial association structures are among quadrant(V)-(VIII). For the convenience of indexing the data, we will use quadrant(I) instead of quadrant (V) and corresponding index for the others. The second level spatial association structure is, nested within corresponding quadrant, the spatial correlation among teeth.

In general, the valid approaches for analyzing correlated data without explicit multivariate distribution consist are based on either GEE and random effect models. The former is suitable for marginal mean or pairwise associations between response outcomes orientated statistical problems and the latter is for subject specific statistical issues. The dental data is spatially correlated and has information about teeth spatial configurations that need to be incarnated in the model to provide explicit structure for inducing dependence among quadrants and teeth at their corresponding levels. The main contribution of this paper is to develop a methodology to model this unique spatial dependence of the deciduous dentition. There is no explicit multivariate distribution available for the spatially correlated binary dental caries experience outcomes. Mixture models(McLachlan and Peel (2000)) are commonly used to are generate flexible multivariate distributions and induce unobserved heterogeneity for correlated data with implicit multivariate distribution.

To take the unique spatial structure of dental data into account, we use two levels of latent variables to take care of the spatial dependence of the teeth within the mouth for each subject. At higher level, the mixture component allocation random vectors $\{Q_{ij} : j = 1, ..., J\}$ for the ith subject are used to allocate the quadrant-wise response vector y_{ij} to its corresponding subgroup that is characterized by the mixture components of the mixture model. The mixture component allocation process has the function to mix the multiple mixture components into a flexible multivariate distributions and induce the dependence among quadrants. Given the mixture

component allocation process, the quadrant-wise response vectors $\{y_{ij} : j = 1, ..., J\}$ are conditionally mutually independent. At intermediate level, conditional on the allocation status of the mixture component process, we introduce spatial latent vectors, $\{T_{im} : m = 1, ..., M, i = 1, ..., n\}$, that are used to tight the generate the mth mixture component flexibly and induce dependence structure among teeth. The joint distribution of this spatial latent vector is given by Undirected graphical Gaussian model with spatial configurations of the teeth taken into account. The observations $\{y_{ijk} : k = 1, ..., K, j = 1, ..., J\}$ will be conditionally independent given Q_i and T_i for $i = 1, ..., n$. Meanwhile, the intermediate level spatial latent vectors $\{T_{im} : m = 1, ..., M\}$ are conditional independent given the higher level spatial latent vector Q_i for $i = 1, ..., n$. In order to assess the spatial symmetry of the caries experience of deciduous dentition, we will examine the pairwise comparisons for the similarity scores that will be defined later on. Due to the complexity of oral biological system, we will give flexible covariance structure for the undirected graphical Gaussian models and leave the number of mixture components to be unknown. A formal model selection procedure will be used to choose appropriate mixture model for the data.

3.3 Models

3.3.1 Bayesian Mixture Models

Finite mixture models with regression structure have a long and extensive literature and have been commonly used. McLachlan and Peel *et al.* (2000) gave a very general framework for mixture model with non-normal components to deal with overdispersed data. Mixture models are used to facilitate the modeling of the heterogeneity from the overdispersed and correlated data by generating flexible distributions of the responses variable of interest and inducing dependence structures among response variables. Conditional on mixture component allocation process Q_{ij}, for mixture model with

M components, $y_i = (y'_{i1}, ..., y_{ij}, ..., y'_{iJ})'$ has contribution to the likelihood as

$$p(y_i|Q_i;\theta) = \prod_{j=1}^{J} \prod_{m=1}^{M} \{\pi_{jm} p_m(y_{ij}|Q_{ijm} = 1;\theta)\}^{Q_{ijm}}, \tag{3.1}$$

where $\{\pi_{jm} : m = 1, ..., M, j = 1, ..., J\}$ are the mixture proportions and $p_m(y_{ij}|Q_{ijm} = 1;\theta)$ is the mth components of the mixture model.

It is known that the estimation for mixture models is straightforward using EM algorithm but with difficulties and challenges. Bayesian estimation for mixture models is feasible and well defined as long as the posterior simulation algorithm converges. Key initial papers on the Bayesian analysis of mixture models using MCMC methods include Diebolt and Robert (1994) and Escobar and West (1995). Provided that suitable (proper conjugate) priors are used, the posterior density will be proper. *WinBUGS* can be used to provide valid posterior samples of the quantities of interest. However, there are some difficulties that have to be addressed with the Bayesian approach in the context of mixture models. First of all, improper priors might yield improper posterior distributions. Secondly, when the number of components M is unknown, the parameter space is ill-defined, which prevents the use of classical testing procedures and priors. Finally, label switching occurs when some of the labels of the mixture components permute. The effect of label switching is important when the solution is calculated iteratively because there is the possibility that the labels of the components may be switched on different iterations. In this paper, we will discuss the methods that have been proposed for overcoming the problems mentioned above.

3.3.2 Response Models

We model the kth response variable within the jth quadrant of the ith subject, y_{ijk}, which is a binary indicator of caries experience of $tooth_{i,k(j)}$. The response model is specified hierarchically. At higher level, the mixture model (3.1) will give a flexible multivariate distributions for the quadrant-wise binary data $\{y_{ij} : j = 1, ..., J\}$ and

induce the dependence structure among the four quadrants. Simultaneously, there exists mixture component allocation random indicator $Q_i = (Q'_{i1}, ..., Q'_{ij}, ..., Q'_{iJ})'$, where $Q_{ij} = (Q_{ij1}, ..., Q_{ijm}, ..., Q_{ijM})'$ is a random binary vector with only element being 1, for $j = 1, ..., J, i = 1, ..., n$. At intermediate level, condition on the Q_i, for instance, $Q_{ijm} = 1$, i.e., y_{ij} follows the mth mixture component in the mixture model. Meanwhile, there exists a spatial latent vector $T_{im} = (T_{i,1(m)}, ..., T_{i,k(m)}, ..., T_{i,K(m)})'$ that is used to tight the J binary response variables $(y_{ij1}, ..., y_{ijk}, ..., y_{ijK})'$. The joint distribution of T_{im} is given by undirected graphical Gaussian model(UGGM) with spatial configurations of the K teeth taking into account. Essentially, T_{im} is used to generate flexible multivariate distribution for the binary response vector and induce the dependence for y_{ijk}. Conditional on Q_i and $\{T_{im} : m = 1, ..., M\}$, the binary response variable y_{ijk} can be modeled by an exponential family distribution with the probability density function as the general form

$$p_m(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)}; \alpha, \gamma, \beta, \phi) = \exp\{\frac{\eta_{imk} y_{ijk} - b_i(\eta_{imk})}{a_i(\phi)} + c_i(y_{ijk}, \phi)\}, \tag{3.2}$$

where $\eta_{imk} = \alpha_m + \beta_k + T_{i,k(m)}$ (McCullagh and Nelder *et al.* 1989).

We assume that the link function $g(\cdot)$ is a canonical link that relates the mean of y_{ijk} to a linear predictor as follows

$$g(E[y_{ijk}|Q_{ijm} = 1; \eta_{imk}]) = \eta_{imk} = \alpha_m + \beta_k + T_{i,k(m)},$$

where $\alpha = (\alpha_1, ..., \alpha_m, ..., \alpha_M)'$ overall component mean with increasing order constraints and $\beta = (\beta_1, ..., \beta_k, ..., \beta_K)'$ are the regression coefficients of generalized latent variable models with constraints $\sum_k \beta_k = 0$. Furthermore, $T_{im} = (T_{i,1(m)}, ..., T_{i,k(m)}, ..., T_{i,K(m)})'$ are the Gaussian with mean zero and covariance matrix $\{\Sigma_T^m : m = 1, ..., M\}$. We assume that $\{Q_i : i = 1, ..., n\}$ and $\{T_{im} : m = 1, ..., M, i = 1, ..., n\}$ are mutually independent, which relate the response variables

to quadrant-specific and tooth-specific covariates and the latent variables.

Under the mixture model and latent variable model approach, we can assume that the response variables are conditionally mutually independent,given the vectors of latent variables $L = \{L_i = (Q_i', T_{i1}', ..., T_{im}', ..., T_{iM}')' : i = 1, ..., n\}$. The joint probability density of y conditional on the set of latent variables L and $\{\pi', \alpha', \beta', \varphi\}$, where $\pi = (\pi_1', ..., \pi_j', ..., \pi_J')'$ with $\pi_j = (\pi_{j1}, ..., \pi_{jm}, ..., \pi_{jM})'$, is as follows

$$p(y|L; \pi', \alpha', \beta', \phi) = \prod_{i,j,m} \left\{ \pi_{jm} \prod_{k=1}^{K} p_m(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)}, \phi) \right\}^{Q_{ijm}}$$
$$= \exp\left\{ \sum_{i,j,m} \left\{ Q_{ijm} \left\{ \log(\pi_{jm}) + \sum_{k=1}^{K} [\frac{\eta_{imk} y_{ijk} - b_i(\eta_{imk})}{a_i(\phi)} + c_i(y_{ijk}, \phi)] \right\} \right\} \right\}, \quad (3.3)$$

where $\prod_{ijm}$ and $\sum_{i,j,m}$ denote $\prod_{i=1}^{n} \prod_{j=1}^{J} \prod_{m=1}^{M}$ and $\sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{m=1}^{M}$ correspondingly.

3.3.3 The Structure Model for Latent Variables

In the response model, given the two levels of latent variables, the conditional independence assumption allows the specification of complete likelihood for the response model. In our modeling approach, the two levels of spatial latent vectors are used to induce the dependence structure of the teeth of interest. In order to incorporate appropriate spatial latent vectors into the model, we need to choose the ones that can really represent the design structure and characterize the random mechanism of data generating process. The objective of these latent processes is to generate flexible distributions for observations and induce the dependency among observations.

At higher level, it is assumed there exist independent mixture component allocation processes, say, $Q_{i1}', ..., Q_{ij}', ..., Q_{iJ}'$, with

$$Q_{ij} = (Q_{ij1}, ..., Q_{ijm}, ..., Q_{ijM})' \sim Mult_M(1, \pi_j), \quad j = 1, ..., J, i = 1, .., n. \quad (3.4)$$

At intermediate level, we will follow the approach in Zhang *et al.* (2007) by incorporating appropriate spatial latent vectors to formulate flexible mixture components.

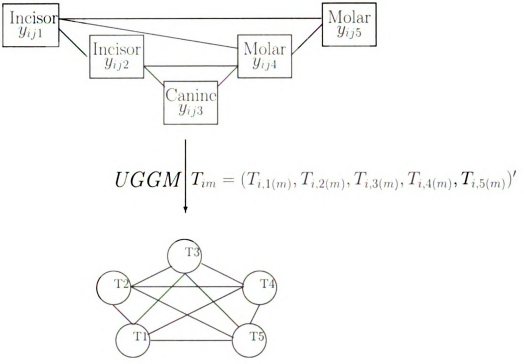

Figure 3.1. The response variables y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4} and y_{ij5} are allocated to the mth cluster and tighten by spatial latent vector $\boldsymbol{T}_{im} = (T_{i,1(m)}, T_{i,2(m)}, T_{i,3(m)}, T_{i,4(m)}, T_{i,5(m)})'$ whose joint distribution is given by UGGM with unstructured precision matrix.

Undirected Graphical Gaussian Models (UGGMs)t are used to give the joint distribution for the spatial latent vectors. The UGGMs will take the spatial configurations of the teeth within quadrants into account. The spatial configurations of the teeth within quadrants are as below.

As shown above graphs, the five teeth within each quadrant can be viewed as five nodes in a graph. If two nodes are not directly connected, they are said to be conditionally independent given the other nodes in the graph. For the mth mixture component, a UGGM is used to describe the spatial configuration of the nodes

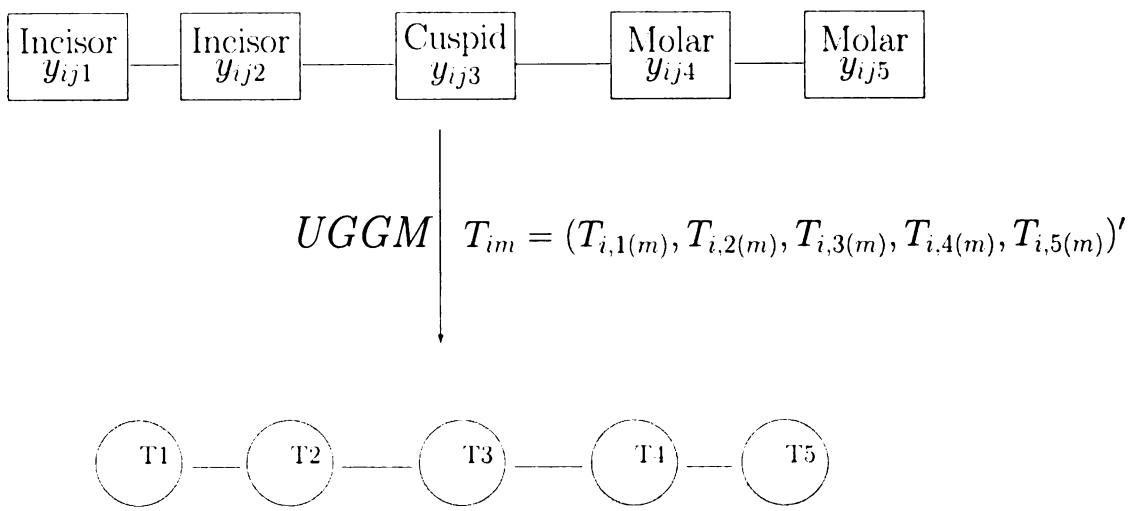

Figure 3.2. Note: The response variables y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4} and y_{ij5} are are allocated to the mth cluster and tighten by spatial latent vector $T_{im} = (T_{i,1(m)}, T_{i,2(m)}, T_{i,3(m)}, T_{i,4(m)}, T_{i,5(m)})'$ whose joint distribution is given by UGGM with precision matrix under CAR model assumption.

and manifest the associations among nodes of interest by assigning random variables $T_{im} = (T_{i,1(m)}, T_{i,2(m)}, T_{i,3(m)}, T_{i,4(m)}, T_{i,5(m)})'$ to the nodes in the graph. Meanwhile, $\{T_{im} : m = 1, ..., M\}$ are mutually independent conditional on Q_i. A UGGM assumes that

$$T_{im} = (T_{i,1(m)}, T_{i,2(m)}, T_{i,3(m)}, T_{i,4(m)}, T_{i,5(m)})' \sim N(\mathbf{0}, \Sigma_T^m), \quad m = 1, ..., M, \quad (3.5)$$

where Σ_T^m is a symmetrical and positive definite matrix for $m = 1, ..., M$.

We know Gaussian random variables are determined by the first two moments. For the identifiability, we already assume the mean structures of the two levels of spatial latent variables are vectors of zeros, then the problem will become issues about the covariance structures $\{\Sigma_T^m : m = 1, ..., M\}$. A general covariance matrix will be unstructured with symmetric and positive definite constraints. The unstructured covariance matrix can be simplified if we assume Markovian properties for the nodes, somehow as shown in the second graph. The Markovian type covariance matrix can be incorporated within spatial statistics by CAR model (Cressie (1991)). The choice of the two types of covariance structure for the spatial latent vectors at tooth nested within quadrant level is made through model selection in Bayesian framework via Deviance Information Criterion(DIC) for missing data problem proposed by Celeux *et al.*(2004), which is an extension of the DIC introduced in Spiegelhalter *et al.*(2002) for Bayesian model selections.

Undirected Graphical Gaussian Model

In this section, we review the graphical Gaussian model (Dempster,1972) required for this paper. Let $G = (V, E)$ be an undirected graph with vertex set $V = \{1, ..., k, ..., K\}$ and edge set $E = \{e_{kk'} : k \neq k' = 1, ..., K\}$, where $e_{kk'} = 1$ or 0 according to whether vertices k and k', $1 \leq k \neq k' \leq K$ are directly connected in G or not. In the undirected graphical Gaussian model, the edges set describes the associate structures of the vertex set. Random vector is assigned

to edges set to represent the association strength between corresponding vertexes. The undirected graphical Gaussian model consists of all k dimensional normal distribution, say $X = \{X_1, ..., X_k, ..., X_K\}$, with $X \sim N(\mathbf{0}, \Sigma)$ and precision matrix $\Omega = \Sigma^{-1} = \{\omega_{kk'} : k \neq k' = 1, ..., K\}$, where Σ is unknown but satisfies the following restrictions in terms of the pairwise conditional independences determined by the Markov properties (Drton and Perlman (2004)):

$$e_{kk'} = 0 \Leftrightarrow X_k \perp X_{k'} | X_{V \setminus \{kk'\}}; \ \forall k \neq k' = 1, ..., K.$$

Conditional Autoregressive Models

For the vector of univariate variables $\nu = (\nu_1, \nu_2, ..., \nu_K)'$, the zero-centered CAR specification, where K is the number of spatial nodes of interest, following Cressie(1991), sets

$$(\nu_k | \nu_{-k}, \sigma_k^2) \sim N(\rho \sum_{\nu_{k'} \in \nu_{-k}} b_{kk'} \nu'_k, \sigma_k^2); \ k = 1, ..., K, \tag{3.6}$$

where $\nu_{-k} = \nu \setminus \{\nu_k\}$. Following Brook's (1964) lemma the resulting joint density for ν takes the form

$$f(\nu | \sigma^2) \propto exp\{-\frac{1}{2} \nu^T D_{\sigma^2}^{-1} (I - \rho B) \nu\}, \tag{3.7}$$

where B is $K \times K$ matrix with $B = \{b_{kk'} : k \neq k' = 1, ..., K\}$ and $b_{kk} = 0, \forall k = 1, ..., K$ and D_{σ^2} us an $K \times K$ diagonal matrix with non-zero entries $\{\sigma_k^2 : k = 1, ..., K\}$. The precision matrix $D_{\sigma^2}^{-1}(I - \rho B)$ need to be symmetric, which yields the conditions

$$b_{kk'} \sigma_{k'}^2 = b_{k'k} \sigma_k^2; \quad \forall k, k' = 1, ..., K. \tag{3.8}$$

If the precision matrix is positive definite, then (3.7) is a proper distribution. Under above parameterizations, the precision matrix $D_{\sigma^2}^{-1}(I - \rho B)$ is nonsingular if $\rho \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$ where $\lambda_{min}, \lambda_{max}$ are the smallest and largest eigenvalues of B

respectively. It usually assumes that the $D_{\sigma 2} = \sigma^2 M$, where M is diagonal matrix with diagonal elements M_{kk} proportional to the conditional variance of σ_k^2. σ^2 controls the overall variability and ρ represent the overall spatial association. Weights matrix B with $B_{kk'}$ need to reflect the spatial association between nodes k and k'. *GoeBUGS*(2004) sets $B_{kk'} = b_{kk'} = 1/n_k$, for $k \neq k'$ and $M_{kk} = 1/n_k$ where n_k is the number of nodes which is adjacent to node k. Under the above settings, the spatial latent vector ν will follow a proper distribution, i.e.,

$$\nu \sim N(\mathbf{0}, \sigma^2 (I - \rho B)^{-1} M). \tag{3.9}$$

3.4 Bayesian Estimations and Statistical Inference

3.4.1 Identifiability of the models

Based on the framework of the mixture of generalized latent variable models, we have to deal with the model identifiability issues at both mixture model level and generalized latent variable level. At mixture model level, we need to deal with *label switching* issue. The interchanging of component labels is generally handled by a constraints on the mixing proportions of the form

$$\pi_{j1} \leq \pi_{j2} \leq \cdots \leq \pi_{jM}, \quad j = 1, ..., J,$$

or on the component means of the form

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_M.$$

Frequently, models with latent variables are not globally identifiable. One can integrate out the latent spatial variable vectors to obtain a marginal likelihood to assess whether parameters are redundant. The contribution to the likelihood from the latent variable model is parameterized by $\{\Sigma_T^m : m = 1, ..., M\}$. The identifiability problem become to examine if the parameters involved in the covariance are redundant, which

might be problematic within frequentist's framework. Dawid (1979) and Gelfand & Sahu (1999) discussed model identifiability issues within Bayesian framework. In particular, Suppose that the Bayesian model is denoted by the likelihood $L(\theta; y)$ and the prior $f(\theta)$ and we partition the parameters of interest as $\theta = (\theta_1, \theta_2)$. If

$$f(\theta_2|\theta_1, y) = f(\theta_2|\theta_1), \tag{3.10}$$

then we say that θ_2 is not identifiable, where $f(\theta_2|\theta_1, y) \propto L(\theta_1, \theta_2; y) f(\theta_2|\theta_1) f(\theta_1)$. That is, if observing data y does not increase our prior knowledge about θ_2 given θ_1, then θ_2 is not identifiable by the data. Dawid's formal definition of Bayesian model nonidentifiability states that θ_2 is not identifiable if and only if $L(\theta_1, \theta_2; y)$ is free of θ_2. In order to make our model identifiable, we need to not only take care of marginal identifiability of the model through integrating out the latent variables, but also put some constraints to the covariance matrix of the Gaussian spatial latent vectors at both levels.

3.4.2 Prior distributions

In this section, prior distributions are chosen for the parameters $\theta = (\pi', \alpha', \beta')'$ and and association parameters ξ. The priors are assigned hierarchically to the corresponding parameters of interest. Gibbs sampling algorithm is applied to simulate the samples from the posterior distributions of the quantities of interest. At higher level, McLachlan and Peel (2000) used a non-informative conjugate proper prior to mixture proportions in the form:

$$\pi_j = (\pi_{j1}, ..., \pi_{jm}, ..., \pi_{jM})' \sim Dirichlet((\varphi_1, ..., \varphi_M)'), \quad j = 1, ..., J, \tag{3.11}$$

where $(\varphi_1, ..., \varphi_M)$ is the weights vector for the mixture proportions. At intermediate level, Zhao *et al.*(2006), Zeger *et al.*(1991) and Dunson *et al.*(2000) all suggested noninformative conjugate prior distributions for the parameters of interest, which

can wash out the effect of priors as sample size increases. Bedrick *et al.*(1996) noted that normal prior distributions were suggested for the logistic regression coefficient θ.

$$(\alpha', \beta')' \sim N(\mu, \mathbf{F}), \tag{3.12}$$

where μ is the a vector of location parameters π and $\mathbf{F}$ is the covariance matrix. It is common to take μ as vector of zeros and $\mathbf{F}$ as diagonal matrix with very larger entries. We are interested in the joint posterior distribution of $(\theta, \xi|y)$. Under mild condition in (Geman and Geman *et al.*(1984)), Gibbs sampler can obtain the joint posterior distribution by sampling from the conditional posterior distributions $(\theta|y, \xi)$ and $(\xi|y, \theta)$ correspondingly. To simplify the sampling from the conditional posterior distributions, we choose hierarchical independent priors for θ and ξ in this hierarchical Bayesian model, i.e. $(\xi|y, \theta) = (\xi|y)$, which is true as long as the priors satisfy $p(\theta, \xi) = p(\theta)p(\xi)$. We proposed two covariance structures for the Guassian spatial latent variable models. In the generalized linear model setting with Gaussian random effects, the proper noninformative conjugate priors will be Inverse Gamma(IG) for signal variance component and Inverse Wishart distribution for a variance-covariance matrix.

Based on the relative relationship among nodes in the graph of the UGGM, we give noninformative priors to precision matrix parameter correspondingly.

(1) Unstructured precision matrix in the UGGM:

For the ith subject, conditional on the higher level spatial latent vector Q_i, the intermediate level spatial latent vectors $\{T_{im} : m = 1, ..., M\}$ are conditionally independent. So, we give independent priors to the precision matrix $\{\Omega_{T_m} : m = 1, ..., M\}$. Similarly, independent Wishart processes are assigned as priors for these precision matrixes.

$$\Omega_{T_m} \sim Wishart(v_{T_m}, \Lambda_{T_m}):\ m = 1, ..., M, \tag{3.13}$$

with the degrees of $v_{T_m} = rank(\Sigma_{T_m}) + 1$ and the precision matrix $\Lambda_{T_m} = I_{v_{T_m} \times v_{T_m}}$.

(2) CAR model based precision matrix in the UGGM:

For the *ith* subject, conditional on the higher level spatial latent vector Q_i, the intermediate level spatial latent vectors $\{T_{im} : m = 1, ..., M\}$ are conditionally independent. So, we give independent priors precision matrix $\{\Omega_{T_m} : m = 1, ..., M\}$ that are parameterized by $\{\sigma_m^2, \rho_m : m = 1, ..., M\}$. Similarly, independent Inverse Gamma (Dunson *et al.*(2000)) distributions, proper conjugate priors, are assigned as priors to the overall variation parameters $\{\sigma_m^2 : m = 1, ..., M\}$ and independent uniform distribution with supports constraints in *section 3.3.2* to the overall spatial association parameters $\{\rho_m : m = 1, ..., M\}$, improper priors *GeoBUGS*(2004) for the over quadrant specific spatial association parameters, respectively.

$$\sigma_m^2 \sim IG(\varepsilon, \varepsilon); \quad m = 1, ..., M, \tag{3.14}$$

and

$$\rho_m \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1}); \quad m = 1, ..., M, \tag{3.15}$$

where ε is very small positive number and $\lambda_{min}^{-1}, \lambda_{max}^{-1}$ are as defined in *section 3.3.2.*

3.4.3 Posterior computations

Let $(\pi', \alpha', \beta', Q', T', \Sigma_T)'$ denote the current state of the Markov chain. We will follow the steps (1)-(3) to obtain the posterior samples of quantities of interest from their posterior distributions. McLachlan & Peel (2000) and Fernández & Green (2002) gave a general posterior sampling algorithm for the mixture model.

Step 1: Posterior sampling for mixture proportions;

$$\pi_j = (\pi_{j1}, ..., \pi_{jm}, ..., \pi_{jM})' \sim Dirichlet((\varphi_1 + N_{j1}, ..., \varphi_m + N_{jm}, ..., \varphi_M + N_{jM})'), \tag{3.16}$$

where $N_{jm} = \sum_{i=1}^{n} Q_{ijm}$ for $j = 1,..,J, m = 1,...,M$ and $\varphi = (\varphi_1,...,\varphi_M)'$ is a known weight vector.

Step 2: Posterior sampling for mixture component allocation random variables;

$$Q_{ij} = (Q_{ij1},...,Q_{ijM})' \sim Multi_M(1,(\tau_{j1},...,\tau_{jM})'),\ \ j = 1,...,J, i = 1,...,n, \tag{3.17}$$

where $\tau_{jm} = \frac{\pi_{jm} p_m(y_{ij}|Q_{ijm}=1,T_{im};\theta)}{\sum_{k=1}^{M} \pi_{jk} p_k(y_{ij}|Q_{ijk}=1,T_{ik};\theta)}$, with

$$p_m(y_{ij}|Q_{ijm} = 1, T_{im};\theta) = \prod_{k=1}^{K} p_m(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)};\theta),$$

and $p_m(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)};\theta)$ is defined in (3) with

$$logit\left\{E(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)};\theta)\right\} = \alpha_m + \beta_k + T_{i,k(m)}.$$

Step 3: Posterior sampling for generalized latent variable models.

Conditional on the mixture component allocation process at higher level, the posterior distributions of parameters and latent variables in the generalized latent variable model can be obtained in standard way (Dunson *et al.*(2000), Zeger*et al.*(1991)). Given the precision matrix $\{\Omega_{T_m} : m = 1,...,M\}$, the joint posterior distribution for the regression parameters and latent variables at intermediate level is

$$\begin{aligned} &p(\theta,T|Q,y;\pi) \propto p(y|Q,T,y;\theta) f(\theta,T) \\ &\propto \exp\left\{\sum_{i,j,m} Q_{ijm}\left\{\log(\pi_{jm}) + \sum_k Q_{ijm}\{\frac{\eta_{imk} y_{ijk} - b_i(\eta_{imk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}\right\}\right\} \\ &\times \exp\left\{-\tfrac{1}{2}\theta'\mathbf{F}^{-1}\theta - \tfrac{1}{2}\sum_{i=1}^{n}\sum_{m=1}^{M} T'_{im}\Omega_{T_m}T_{im}\right\}, \end{aligned} \tag{3.18}$$

where $\sum_{i,j,m}$ denotes $\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{m=1}^{M}$, $f(\cdot)$ denote the joint prior density, $Q = (Q'_1,....,Q'_i,....,Q'_n)'$ with $Q_i = (Q'_{i1},....,Q'_{ij},....,Q'_{iJ})'$ and $Q_{ij} = (Q_{ij1},....,Q_{ijm},....,Q_{ijM})'$, $T = (T'_1,....,T'_i,....,T'_n)'$ with $T_i = (T'_{i1},....,T'_{im},....,T'_{iM})'$ and $T_{im} = (T_{i,1(m)},....,T_{i,k(m)},....,T_{i,K(m)})'$. Furthermore, $\theta = (\alpha',\beta',\varphi)'$ and $\eta_{imk} = \alpha_m + \beta_k + T_{i,k(m)}$. In practice, we set μ to be vector of zeros and $\mathbf{F}$ to

be diagonal matrix with large diagonal elements. If the MCMC algorithm is a Gibbs sampler, the full conditional distribution of each of the unknowns in (19) needs to be specified, which can be obtained in a standard way Dunson *et al.*(2000), Zeger *et al.*(1991)). For the fixed effect θ, the full conditional distribution is

$$\begin{aligned} & p(\theta|Q,T,y;\pi) \\ & \propto \exp\left\{\sum_{i,j,m} Q_{ijm}\left\{\log(\pi_{jm}) + \sum_k Q_{ijm}\{\frac{\eta_{imk}y_{ijk}-b_i(\eta_{imk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}\right\}\right\} \\ & \times \exp\left\{-\frac{1}{2}\theta' F^{-1}\theta\right\}. \end{aligned} \tag{3.19}$$

The full conditional distribution for the Gaussian spatial latent vectors T_{im} is

$$\begin{aligned} & p(T_{im}|Q,y;\pi,\theta) \\ & \propto \exp\left\{\sum_{i,j,m} Q_{ijm}\left\{\log(\pi_{jm}) + \sum_k Q_{ijm}\{\frac{\eta_{imk}y_{ijk}-b_i(\eta_{imk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}\right\}\right\} \\ & \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{m=1}^{M} T'_{im}\Omega_{T_m}T_{im}\right\}. \end{aligned} \tag{3.20}$$

The full conditional distributions of precision matrix $\{\Omega_{T_m} : m = 1,...,M\}$ can be obtained in terms of different precision matrix structures correspondingly.

(1) Unstructured precision matrix:

$$p(\Omega_{T_m}|Q,T,y;\pi,\theta) = Wishart(\upsilon_{T_m} + N_m, \Lambda_{T_m} + \sum_{i=1}^{n} T_{im}T'_{im}), \tag{3.21}$$

where $N_m = \sum_{j=1}^{J} N_{jm} = \sum_{i=1}^{n}\sum_{j=1}^{J} Q_{ijm}$.

(2) CAR model based precision matrix:

(2.1) Overall precision parameters:

$$\begin{aligned} & p(\tau_m|Q,T,y;\pi,\theta) \\ & \propto \exp\left\{\sum_{i,j,m} Q_{ijm}\left\{\log(\pi_{jm}) + \sum_k Q_{ijm}\{\frac{\eta_{imk}y_{ijk}-b_i(\eta_{imk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}\right\}\right\} \\ & \times \tau_m^{\varepsilon-1}\exp\{-\tau_m\varepsilon\}. \end{aligned} \tag{3.22}$$

(2.2) Overall spatial parameters:

$$p(\rho_m|Q,T,y;\pi,\theta)$$
$$\propto \exp\left\{\sum_{i,j,m} Q_{ijm}\left\{\log(\pi_{jm}) + \sum_k Q_{ijm}\{\frac{\eta_{imk}y_{ijk}-b_i(\eta_{imk})}{a_i(\varphi)} + c_i(y_{ijk},\varphi)\}\right\}\right\}$$
$$\times I_{(\lambda_{min}^{-1},\lambda_{max}^{-1})}(\rho_m). \tag{3.23}$$

All the posterior distributions, except for $\{\rho_m : m = 1,,...,M\}$, are proper based on their proper conjugate priors. The uniform priors for the overall spatial parameters are not conjugate, which might lead to improper posterior distributions. The simplest technique for verifying if the posterior distributions of the parameters is proper is to verify if the posterior distribution is proper for reduced data by discarding all but a single outcome per subject leading to a reduced data set consisting of independent outcomes, are proper (O'Brien and Dunson, 2004). Since the covariance structures do not appear in the reduced data likelihood and also the support for the spatial association parameters is finite, i.e., $\{\rho_m \in (\lambda_{min}^{-1},\lambda_{max}^{-1}) : m = 1,...,M\}$, so the posterior distributions of the spatial association parameters $\{\rho_m : m = 1,...,M\}$ are proper. The algorithm for the posterior computation is through sampling π, Q, θ, T, and ξ respectively from the above conditional distributions.

3.4.4 Bayesian Model Selection

The formal procedure for choosing an appropriate Bayesian hierarchical model for the observed data necessities methods to compare alternative models within Bayesian framework. The DIC (deviance information criterion, Spiegelhalter *et al.* (2002)) is a hierarchical modeling selection criterion that can be viewed as a generalization of the AIC (Akaike information criterion,Akaike, 1973) and BIC (Bayesian information criterion, Kass and Raftery, 1995). It is particularly useful in Bayesian model selection problems where the posterior distributions of parameters have been obtained by Markov chain Monte Carlo (MCMC) simulation. The DIC-statistic is a measure of

model complexity and goodness of fit with the definition as

$$DIC = \overline{D(\vartheta)} + pD,$$

where $D(\vartheta)$ is the deviance given the model parameters $\vartheta = (\pi', \theta', \xi')'$, defined as

$$D(\vartheta) = -2\log(p(y|\vartheta)) + 2\log(h(y)),$$

where $h(y)$ is some fully specified standardizing term which is function of the data alone. $\overline{D(\vartheta)}$ is the posterior mean of the deviance, a measurement of goodness of fit of the proposed model for the observed data. $D(\bar{\vartheta})$ is the deviance evaluated at the posterior mean of ϑ and $pD = \overline{D(\vartheta)} - D(\bar{\vartheta})$ is the effective number of parameters in the model, a penalty for the complexity of the model. The quantities $\overline{D(\vartheta)}$ and $D(\bar{\vartheta})$ can be trivially computed from an MCMC simulation chain. Rather than the conventional DIC introduced in Spiegelhalter *et al.* (2002), our hierarchial models containing two levels of latent variables, which necessitates the model selections to be based on the DIC for missing data problems (Celeux *et al.*,2006). MCMC methods, such as the Gibbs sampler, can be employed conveniently to produce posteriors for parameters that are marginalized over latent spatial vectors. We computed the complete DICs (Celeux *et al.*,2006) by using the MCMC simulation results to get both the measurement of goodness of fit and the number of effective parameters associated with each models and used these statistics to select the most appropriate model. In terms of our problem, we have to deal with latent variables to get a complete DICs.

In order to compute the complete DICs, Celeux *et al.* (2006) gave a definition of complete data DIC, by defining the complete data estimator $E_\vartheta[\vartheta|y,q,t]$, which does not suffer from identifiability problems since the components are identified by $(q', t')'$, the realization of the spatial latent vectors $(Q', T')'$, and then obtain DIC for the complete model as

$$DIC(y,q,t) = -4E_\vartheta[\log(p(y,q,t|\vartheta))|y,q,t] + 2\log(p(y,q,t|E_\vartheta[\vartheta|y,q,t])). \quad (3.24)$$

As in the EM algorithm, we can then integrate this quantity to define

$$\begin{aligned} DIC &= E_{Q,T}[DIC(y,Q,T)|y] \\ &= -4E_{\vartheta,Q,T}[\log(p(y,Q,T|\vartheta))|y] + 2E_{Q,T}[\log(p(y,Q,T|E_{\vartheta}[\vartheta|y,Q,T]))|y]. \end{aligned} \tag{3.25}$$

More specifically, notice that

$$\begin{aligned} E_{\vartheta,Q,T}[\log(p(y,Q,T|\vartheta))|y] &= E_{\vartheta}\left\{E_Q(E_T[\log(p(y,Q,T|\vartheta))|y,Q;\vartheta]|y,\vartheta)|y\right\} \\ &= E_{\vartheta}\left\{E_T(E_Q[\log(p(y,Q,T|\vartheta))|y,T;\vartheta]|y,\vartheta)|y\right\}, \end{aligned} \tag{3.26}$$

also

$$\begin{aligned} \log(p(y,Q,T|\vartheta)) &= \textstyle\sum_{i,j,m}\left\{Q_{ijm}\log(\pi_{jm}p(T_{im}|\Sigma_T^m))\right\} \\ &+ \textstyle\sum_{i,j,m,k}\left\{Q_{ijm}\log(p_m(y_{ijk}|Q_{ijm}=1,T_{i,k(m)};\theta))\right\}, \end{aligned} \tag{3.27}$$

where $p_m(y_{ijk}|Q_{ijm}=1,T_{i,k(m)};\theta)$ is given in (3) and $p(T_{im}|\Sigma_T^m)$ is given in (6). Interchanging the order of Q and T in the integrations by Fubini's theorem , we can have

$$\begin{aligned} &E_{\vartheta,Q,T}[\log(f(y,Q,T|\vartheta))|y] \\ &= E_{\vartheta}\left\{E_T(E_Q[\log(f(y,Q,T|\vartheta))|y,T;\vartheta]|y,\vartheta)|y\right\} \\ &= E_{\vartheta}\left\{E_T(\textstyle\sum_{i,j,m}\left\{E_Q(Q_{ijm}|y,T;\vartheta)\log(\pi_{jm}p(T_{im}|\Sigma_T^m))\right\}|y,\vartheta)|y\right\} \\ &+ E_{\vartheta}\left\{E_T(\textstyle\sum_{i,j,m,k}\left\{E_Q(Q_{ijm}|y,T;\vartheta)\log(p_m(y_{ijk}|Q_{ijm}=1,T_{im};\theta))\right\}|y,\vartheta)|y\right\}, \end{aligned} \tag{3.28}$$

where $E_Q(Q_{ijm}|y,T;\vartheta)$ is given as below:

$$E_Q(Q_{ijm}|y,T;\vartheta) = P(Q_{ijm}=1|y,T;\vartheta) = \frac{\pi_{jm}p_m(y_{ij}|Q_{ijm}=1,T_{im};\theta)}{\sum_{k=1}^{M}\pi_{jk}p_k(y_{ij}|Q_{ijm}=1,T_{ik};\theta)},$$

with

$$p_m(y_{ij}|T_{im};\theta) = \prod_{j=1}^{J} p_m(y_{ijk}|Q_{ijm}=1,T_{i,k(m)};\theta),$$

and

$$
\begin{aligned}
&E_{Q,T}[\log(p(y,Q,T|E_{\vartheta}[\vartheta|y,Q,T]))|y] \\
&\quad = E_T\left\{E_Q[\log(p(y,Q,T|E_{\vartheta}[\vartheta|y,Q,T]))|y]\right\} \\
&\quad = E_T\left\{\sum_{i,j,m}\left\{E_Q(Q_{ijm}|y,T;\hat{\vartheta})\log(\widehat{\pi_{jm}}p(T_{im}|\widehat{\Sigma_T^m}))\right\}|y\right\} \\
&\quad + E_T\left\{\sum_{i,j,m,k}\left\{E_Q(Q_{ijm}|y,T;\hat{\vartheta})\log(p_m(y_{ijk}|Q_{ijm}=1,T_{im};\hat{\theta}))\right\}|y\right\}
\end{aligned}
\tag{3.29}
$$

where $\hat{\vartheta}, \{\widehat{\pi_{jm}} : j = 1,...,J, m = 1,...,M\}, \{\widehat{\Sigma_T^m} : m = 1,...,M\}, \hat{\theta}$ are posterior means of $\vartheta, \{\pi_{jm} : j = 1,...,J, m = 1,...,M\}, \{\Sigma_T^m : m = 1,...,M\}, \theta$ correspondingly.

All the integration can be obtained routinely by Monte Carlo integration approximation using the MCMC posterior samples in the coda file of *WinBUGS*.

Spatial symmetry hypothetical testing

The spatial symmetry property in our problem means the joint caries experience presentations for response variables at quadrant level are highly associated with one another. Dentists do believe that spatial symmetry exist in mouth. Lesaffre *et al.*(2006) showed empirically that the caries experience for left and right quadrants are more strongly associated than the other cases. Unfortunately, few literatures have discussed this issue comprehensively. The mixture of generalized latent variable models provides a way to examine the spatial symmetry of the four quadrants in terms of joint caries experiences. Under the mixture model, if there are two quadrant-wise binary response vectors $y_{ij} = (y_{ij1},...,y_{ijk},...,y_{ijK})'$ and $y_{ij'} = (y_{ij'1},...,y_{ij'k},...,y_{ij'K})'$ who have the exact joint probabilistic behaviors, then the mixture component allocation processes will always assign the two binary response vectors to the same mixture component. Specifically, the mixture component allocation processes $Q_{ij} = (Q_{ij1},...,Q_{ijm},...,Q_{ijM})'$ and $Q_{ij'} = (Q_{ij'1},...,Q_{ij'm},...,Q_{ij'M})'$ will satisfy $\sum_{m=1}^{M} P(Q_{ijm} = Q_{ij'm} = 1|y) = 1$. Hence, the strength of the similarity of two quadrant-wise responses vectors that is defined by $S_{jj'}$, say, the jth quadrant and

the $j'th$ quadrant, can be measured by the below quantity

$$S_{jj'} = \frac{1}{n}\sum_{i=1}^{n}\sum_{m=1}^{M} P(Q_{ijm} \equiv Q_{ij'm} = 1|y) \tag{3.30}$$

Hypothetical testing for pairwise comparisons among spatial association strength parameters.

In order to assess the spatial symmetry of the four quadrants, we need to introduce different "Neighborhoods" relationships that can explain the relative spatial structures of the quadrants of interest. Spatial symmetry is assessed at the quadrant level, instead of tooth level. At quadrant level, We define the vector of teeth to be "Horizontal Neighbors" to each other, if the two quadrants are both in either "Upper Jaw" or "Lower Jaw", and to be "Vertical Neighbors" to one another, if the two quadrants are both in either "Left Jaw" or "Right Jaw" and to be "Across Neighbors" to one another, if the two quadrants are either in "Left Jaw" or "Right Jaw". The assessment of quadrant spatial symmetry in terms of cries prevalence will be based on "Left-right", i.e., " Horizontal Neighbors", "Up-down", i.e.,"Vertical Neighbors" and "Across", i.e., "Across Neighbors".

There are two ways to assess the spatial symmetry among quadrants in terms of caries prevalence incidence through statistical hypothesis statement. The first one is based on the so called "overall" spatial symmetry assessments via a weighted statistic and the second is the so called "specific" spatial symmetry assessment that is the direct comparisons of the spatial symmetry measurements.

First of all, the weighted statistics for assessing the overall spatial associations in terms of "Left-right", "Up-down" and "Across" can be formulated as below:

$$S_{LR} = \frac{1}{2}(S_{56} + S_{78});$$

$$S_{UD} = \frac{1}{2}(S_{67} + S_{58});$$

$$S_A = \frac{1}{2}(S_{68} + S_{57}).$$

The statistical hypothesis testing about the overall spatial association in terms of "Left-right" V.S. "Up-down", "Left-right" V.S. "Across" and "Across" V.S. "Up-down" can be formulated as follows:

(1) Left-right Versus Up-down

$$H_0 : S_{LR} = S_{UD} \qquad V.S. \qquad H_a : S_{LR} \neq S_{UD}; \tag{3.31}$$

(2) Left-right Versus Across

$$H_0 : S_{LR} = S_A \qquad V.S. \qquad H_a : S_{LR} \neq S_A; \tag{3.32}$$

(3) Across Versus Up-down

$$H_0 : S_A = S_{UD} \qquad V.S. \qquad H_a : S_A \neq S_{UD}. \tag{3.33}$$

Secondly, if the assessment is based on the direct comparisons of spatial symmetry measurement, there are twelve possible hypothesis testing situations for the spatial symmetries in terms of partial correlation between quadrants.

(1.1) Left-right Versus Up-down The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 6 and quadrant 7, with quadrant 6 as reference.

$$H_0 : S_{56} = S_{67} \qquad V.S. \qquad H_a : S_{56} \neq S_{67}; \tag{3.34}$$

(1.2) Left-right Versus Up-down The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 5 and quadrant 8, with quadrant 5 as reference.

$$H_0 : S_{56} = S_{58} \qquad V.S. \qquad H_a : S_{56} \neq S_{58}; \tag{3.35}$$

(1.3) Left-right Versus Up-down The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 6 and quadrant 7, with quadrant 7 as reference.

$$H_0 : S_{78} = S_{67} \qquad V.S. \qquad H_a : S_{78} \neq S_{67}; \tag{3.36}$$

(1.4) Left-right Versus Up-down The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 5 and quadrant 8, with quadrant 8 as reference.

$$H_0 : S_{78} = S_{58} \qquad V.S. \qquad H_a : S_{78} \neq S_{58}; \tag{3.37}$$

(2.1) Left-right Versus Across The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 6 and quadrant 8, with quadrant 6 as reference.

$$H_0 : S_{56} = S_{68} \qquad V.S. \qquad H_a : S_{56} \neq S_{68}; \tag{3.38}$$

(2.2) Left-right Versus Across The association between quadrant 5 and quadrant 6 V.S. the association between quadrant 5 and quadrant 7, with quadrant 5 as reference.

$$H_0 : S_{56} = S_{57} \qquad V.S. \qquad H_a : S_{56} \neq S_{57}; \tag{3.39}$$

(2.3) Left-right Versus Across The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 6 and quadrant 8, with quadrant 8 as reference.

$$H_0 : S_{78} = S_{68} \qquad V.S. \qquad H_a : S_{78} \neq S_{68}; \tag{3.40}$$

(2.4) Left-right Versus Across The association between quadrant 7 and quadrant 8 V.S. the association between quadrant 5 and quadrant 7, with quadrant 7 as reference.

$$H_0 : S_{78} = S_{57} \qquad V.S. \qquad H_a : S_{78} \neq S_{57}; \tag{3.41}$$

(3.1) Across Versus Up-down The association between quadrant 5 and quadrant 7 V.S. the association between quadrant 5 and quadrant 8, with quadrant 5 as reference.

$$H_0 : S_{57} = S_{58} \qquad V.S. \qquad H_a : S_{57} \neq S_{58}; \tag{3.42}$$

(3.2) Across Versus Up-down The association between quadrant 5 and quadrant 7 V.S. the association between quadrant 6 and quadrant 7, with quadrant 7 as reference.

$$H_0 : S_{57} = S_{67} \qquad V.S. \qquad H_a : S_{57} \neq S_{67}; \tag{3.43}$$

(3.3) Across Versus Up-down The association between quadrant 6 and quadrant 8 V.S. the association between quadrant 5 and quadrant 8, with quadrant 8 as reference.

$$H_0 : S_{68} = S_{58} \qquad V.S. \qquad H_a : S_{68} \neq S_{58}; \tag{3.44}$$

(3.4) Across Versus Up-down The association between quadrant 6 and quadrant 8 V.S. the association between quadrant 6 and quadrant 7, with quadrant 6 as reference.

$$H_0 : S_{68} = S_{67} \qquad V.S. \qquad H_a : S_{68} \neq S_{67}. \tag{3.45}$$

Simultaneous credible intervals

Pairwise spatial symmetry hypothesis testing is based on credible intervals for the differences between two partial correlations corresponding to two different nodes (quadrants) in the UGGM. In Bayesian statistics, a credible interval is a posterior probability interval, used for purposes similar to those of confidence intervals in frequentist statistics. Suppose that parameter ς is of interest, a $(1-\alpha)100\%$ credible interval

for the parameter ς of interest is any set C such that $P_{\pi(\varsigma|y)}(\varsigma \in C) = 1 - \alpha$, where $\pi(\varsigma|y)$ is the posterior distribution of parameter ς given the observed data y.

Since we are performing a multiple spatial symmetry comparisons among quadrants in terms of all possible hypothesis testing situations, it is necessary to give a simultaneous credible regions (Besag *et al.* (1995)) to control type S error rate (Gelman *et al.*), i.e. the similar concept as type I error rate in frequentist's framework. The $100K/M\%$ simultaneous credible regions is based on order statistics (Besag *et al.* (1995))

$$\left\{[(S_I - S_{II})^{[M+1-t^*]}, (S_I - S_{II})^{[t^*]}] : (I, II) \in Neighborhood\right\},$$

where

$$t^* = \min\{t : \#\left\{(S_I - S_{II})^{[M+1-t^*]} \leq (S_I - S_{II})^{(t)} \leq (S_I - S_{II})^{[t^*]}\right\} \geq K\},$$

and $\{(S_I - S_{II})^{(t)} : t = 1, ..., M, (I, II) \in Neighborhood\}$ are the posterior samples of $\{(S_I - S_{II}) : (I, II) \in Neighborhood\}$. Here, $Neighborhood = \{("LR", "UD"), ("LR", "A"), ("A", "UD")\}$.

Similarly, the $100K/M\%$ simultaneous credible regions for specific spatial associations difference are given by

$$\{[(S_{ii'} - S_{jj'})^{[M+1-t^*]}, (S_{ii'} - S_{jj'})^{[t^*]}] : i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\},$$

where

$$t^* = \min\{t : \#\left\{(S_{ii'} - S_{jj'})^{[M+1-t^*]} \leq (S_{ii'} - S_{jj'})^{(t)} \leq (S_{ii'} - S_{jj'})^{[t^*]}\right\} \geq K\},$$

and $\{(S_{ii'} - S_{jj'})^{(t)} : t = 1, ..., M, i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\}$ are the posterior samples of $\{(S_{ii'} - S_{jj'}) : i \neq i', j \neq j', (i, i') \neq (j, j'), i, j = 1, ..., J\}$.

Table 3.1. Prevalence of caries experience(% affected) in the deciduous dentition of 6,7,8-year-old children n=4,351.

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

3.5 The Signal Tandmobiel Project Example

In the Signal-Tandmobiel project, there are 4,468 schoolchildren who were among 6,7,8-year-old, (born in 1989) from 179 schools in Flanders (Belgium) and were selected by a stratified clustered random sample. The mean age of the children on the day of examination was 7.1 years (SD = 0.4). The 15 strata were obtained by combining the 3 types of educational system (public, municipal and private schools) with geographical areas (the 5 Flemish provinces). The schools represented the clusters. This sample represents about 7% of the corresponding Flemish population. The sampling procedure aimed at selecting each child in Flanders with equal probability. A more detailed description of the design of the Signal-Tandmobiel project is reported in Vanobbergen *et al.* (2000).

3.5.1 Primary results

The population prevalence data of caries experience in the deciduous dentition at the tooth is shown in table 1 for the 6,7,8-year-old children. The descriptive observations suggested a symmetrical distribution of caries experience at the population level.

In Vanobbergen *et al.* , the Null hypothesis of population symmetry at tooth level was tested for all deciduous molars. The results are shown in table 2.

The above result shows that it is left-right spatial symmetry is the most notable.

Table 3.2. Odds ratios and 95% confidence intervals for the 2×2 association models for caries on deciduous molars on tooth in 7-year-old children.

First Molar (ALR model)			
54	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)			
55	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)

Decayed teeth of discordant contralateral pairs tend to aggregate on the right or the left side of the subject's mouth than would be expected by chance alone (Vanobbergen *et al.*(2006)).

Zhang *et al.*(2007) proposed a Bayesian generalized latent variable models(BGLVMs) that is a complete likelihood approach for analyzing the dental data and gave a 95% simultaneous credible intervals, in table 3, for the differences of the partial correlations, which are used to measure the association strength among different nodes (quadrants). The simultaneous credible intervals for the spatial symmetry testing situations are given as follow.

The above result also shows that the spatial symmetry, in terms of the caries prevalence, between left and right quadrant is stronger than the ones either between upper quadrant and down quadrant or across quadrants.

3.5.2 The results from our approach

Now we show how the above methodology works for dental data and need to specify all the functions and general notations. In our study, all of the responses

Table 3.3. Credible intervals of spatial association strength comparisons based on BGLVMs and UGGM with unstructured covariance structure

Simultaneous Spatial Effects	Credible intervals
left/righ .v.s. across	
$\rho_{56} - \rho_{68}$	(0.134, 1.581)
$\rho_{56} - \rho_{57}$	(0.394, 1.711)
$\rho_{78} - \rho_{68}$	(0.237, 1.589)
$\rho_{78} - \rho_{57}$	(0.433, 1.728)
left/righ .v.s. upper/down	
$\rho_{56} - \rho_{67}$	(0.235, 1.551)
$\rho_{56} - \rho_{58}$	(0.117, 1.485)
$\rho_{78} - \rho_{67}$	(0.230, 1.601)
$\rho_{78} - \rho_{58}$	(0.215, 1.504)
across .v.s. upper/down	
$\rho_{68} - \rho_{67}$	(-1.303, 1.313)
$\rho_{68} - \rho_{58}$	(-1.327, 1.204)
$\rho_{57} - \rho_{67}$	(-1.442, 1.109)
$\rho_{57} - \rho_{58}$	(-1.488, 1.042)
DIC	593.300
N.burnin	1000
N.interation	11000

are binary, so we have the following: $a_i(\varphi) = 1$, $b_i(\eta_{imk}) = \log(1 + \exp(\eta_{imk}))$, $c_i(y_{ijk}, \varphi) = 0$, , $g(x) = \log(\frac{x}{1-x})$, , $E[y_{ijk}|Q_{ijm} = 1, \eta_{imk}] = \frac{\exp(\eta_{imk})}{1+\exp(\eta_{imk})}$, for $k = 1, ..., 5, m = 1, ...M, j = 1, ..., 4, i = 1, ..., n$. Hence, the parameters of interest in the observational model is $\theta = (\pi', \alpha', \beta')'$ and $\xi = \xi^{-1}((\Sigma_T^1, ..., \Sigma_T^m, ..., \Sigma_T^M)')$, then $p(y_{ijk}|Q_{ijm} = 1, \eta_{imk}, \varphi) = p_m(y_{ijk}|Q_{ijm} = 1, \eta_{imk})$, and $\log p(y_{ijk}|Q_{ijm} = 1, \eta_{imk}) = \log p(y_{ijk}|Q_{ijm} = 1, T_{im}, \theta) = \eta_{imk} y_{ijk} - \log(1 + \exp(\eta_{imk}))$. The canonical parameter $\{\eta_{imk} : k = 1, .., K, m = 1, ..., M, i = 1, ..., n\}$ is defined as follows:

$$\eta_{imk} = \alpha_m + \beta_k + T_{i,k(m)}. \tag{3.46}$$

Priors for parameters of interest are given by noninformative proper conjugate priors, which will give comparable results as frequentist's when sample size large enough, in which case the sample can provide enough information for parameter estimates and prior information will be washed away, also conjugate priors will make the posterior proper if the prior is proper, in which case the Gibbs sampler can efficiently provide the appropriate posterior samples from the target posterior distributions. More specifically, the priors are given as follows:

$$\pi_j = (\pi_{j,1}, ..., \pi_{j,m}, ..., \pi_{j,M})' \sim Dirichlet(\varphi); \quad j = 1, ..., J, \tag{3.47}$$

where φ is a M-dimensional vector of ones with M being prespecified, and

$$\alpha_m \sim N(0, 1000); \quad m = 1, ..., M, \tag{3.48}$$

and

$$\beta_k \sim N(0, 1000); \quad k = 1, ..., 4. \tag{3.49}$$

We assume the order restriction to the mixture component effect α, i.e., $\alpha_1 \leq \alpha_2 \leq , ..., \leq \alpha_M$ for the label switching problems with the mixture model. For identifiability of the generalized latent variable model, we assume $\sum_{k=1}^{5} \beta_k = 0$. For the priors of precision matrix, O'Malley and Zaslavsky (2006) proposed scaled Wishart distribution as conjugate proper priors

For the priors of the precision matrix $\{\Omega_{T_m} = \Sigma_{T_m}^{-1} : m = 1, ..., M\}$, there are two different models for the the structures of the precision matrix. (1) Unstructured precision matrix, the common noninformative conjugate proper prior is Wishart distribution, i.e.

$$\Sigma_{T_m}^{-1} = \Omega_{T_m} \sim Wishart((5+1), I): \quad m = 1, ..., M, \tag{3.50}$$

where I is 5×5 identity matrix, which will give a noninformative conjugate proper priors for the precision matric $\Omega_{T_m} = \Sigma_{T_m}^{-1}, m = 1, ..., M$.

(2) Covariance matrix with structure under CAR model:

$$\sigma_m^{-2} = \tau_m \sim Gamma(0.001, 0.001); \quad m = 1, ..., M, \tag{3.51}$$

and

$$\rho_m \sim U(\lambda_{min}^{-1}, \lambda_{max}^{-1}). \quad m = 1, ..., M, \tag{3.52}$$

where $\{\sigma_m^2 : m = 1, ..., M\}$ are the quadrant specific parameters for overall variability and $\{\rho_m : m = 1, ..., M\}$ are the quadrant specific parameters for overall spatial effects. λ_{min} and λ_{max} are as defined in CAR models in *section 3.3.2.*

Our mixture of generalized latent variable models are implemented in *WinBUGS*, using noninformative priors for the parameters of interest. After 1000 burn in, the posterior inference is based on 11000 iterations. The model selection in terms of number of mixture components at higher level and covariance matrix structure for spatial latent vectors at intermediate level is based on DIC for missing data problem (Celeux *et al.* (2006)).

Based on the above results from table (4)-(7) for four different models, the posterior inferences about the spatial similarity in terms of caries prevalence are roughly similar, which is because all the models work fairly well. Bayesian model selection is based on DICs of both of the models, the smaller the DIC, the better the model. It is common that if the difference between two different models are more than 10

Table 3.4. Credible intervals of spatial similarity comparisons based on mixture model with 2 components and UGGM with unstructured covariance structure

Spatial Effects	Credible intervals (95 %)
left/righ .v.s. across	
$S_{56} - S_{68}$	(-0.021, 0.194)
$S_{56} - S_{57}$	(0.014, 0.229)
$S_{78} - S_{68}$	(-0.014, 0.208)
$S_{78} - S_{57}$	(0.028, 0.236)
left/righ .v.s. upper/down	
$S_{56} - S_{67}$	(0.000, 0.222)
$S_{56} - S_{58}$	(-0.007, 0.208)
$S_{78} - S_{67}$	(0.014, 0.229)
$S_{78} - S_{58}$	(0.000, 0.215)
across .v.s. upper/down	
$S_{68} - S_{67}$	(-0.111, 0.097)
$S_{68} - S_{58}$	(-0.111, 0.097)
$S_{57} - S_{67}$	(-0.111, 0.097)
$S_{57} - S_{58}$	(-0.111, 0.097)
DIC	635.400
N.burnin	1000
N.interation	11000

Table 3.5. Credible intervals of spatial similarity comparisons based on based on mixture model with 2 components and UGGM with CAR model based covariance structure

Spatial Effects	Credible intervals (95 %)
left/righ .v.s. across	
$S_{56} - S_{68}$	(0.007, 0.208)
$S_{56} - S_{57}$	(0.042, 0.243)
$S_{78} - S_{68}$	(0.021, 0.222)
$S_{78} - S_{57}$	(0.056, 0.257)
left/righ .v.s. upper/down	
$S_{56} - S_{67}$	(0.035, 0.236)
$S_{56} - S_{58}$	(0.021, 0.222)
$S_{78} - S_{67}$	(0.049, 0.213)
$S_{78} - S_{58}$	(0.035, 0.229)
across .v.s. upper/down	
$S_{68} - S_{67}$	(-0.104, 0.083)
$S_{68} - S_{58}$	(-0.101, 0.083)
$S_{57} - S_{67}$	(-0.104, 0.083)
$S_{57} - S_{58}$	(-0.104, 0.083)
DIC	452.200
N.burnin	1000
N.interation	11000

Table 3.6. Credible intervals of spatial similarity comparisons based on mixture model with 3 components and UGGM with unstructured covariance structure

Spatial Effects	Credible intervals (95 %)
left/righ .v.s. across	
$S_{56} - S_{68}$	(0.007, 0.188)
$S_{56} - S_{57}$	(0.035, 0.215)
$S_{78} - S_{68}$	(0.007, 0.188)
$S_{78} - S_{57}$	(0.035, 0.215)
left/righ .v.s. upper/down	
$S_{56} - S_{67}$	(0.028, 0.208)
$S_{56} - S_{58}$	(0.014, 0.195)
$S_{78} - S_{67}$	(0.028, 0.201)
$S_{78} - S_{58}$	(0.014, 0.195)
across .v.s. upper/down	
$S_{68} - S_{67}$	(-0.090, 0.076)
$S_{68} - S_{58}$	(-0.090, 0.076)
$S_{57} - S_{67}$	(-0.090, 0.076)
$S_{57} - S_{58}$	(-0.090, 0.076)
DIC	537.500
N.burnin	1000
N.interation	11000

Table 3.7. Credible intervals of spatial similarity comparisons based on based on mixture model with 3 components and UGGM with CAR model based covariance structure

Spatial Effects	Credible intervals (95 %)
left/righ .v.s. across	
$S_{56} - S_{68}$	(0.014, 0.215)
$S_{56} - S_{57}$	(0.042, 0.243)
$S_{78} - S_{68}$	(0.035, 0.229)
$S_{78} - S_{57}$	(0.056, 0.250)
left/righ .v.s. upper/down	
$S_{56} - S_{67}$	(0.042, 0.243)
$S_{56} - S_{58}$	(0.021, 0.215)
$S_{78} - S_{67}$	(0.056, 0.257)
$S_{78} - S_{58}$	(0.035, 0.229)
across .v.s. upper/down	
$S_{68} - S_{67}$	(-0.090, 0.097)
$S_{68} - S_{58}$	(-0.090, 0.097)
$S_{57} - S_{67}$	(-0.090, 0.097)
$S_{57} - S_{58}$	(-0.090, 0.097)
DIC	348.600
N.burnin	1000
N.interation	11000

then the model with smaller DIC is the better one. Hence the model (shown in table 7) with 3 components and CAR model based covariance matrix for the corresponding spatial latent vectors is more appropriate than the other models for the observed data. Specifically, at higher level, the quadrant-wise response vectors follow a mixture model with 3 components, and were assigned mixture label for each response vectors by its mixture component allocation process. Conditional on the mixture label, at the intermediate level, the Gaussian spatial latent vectors, modeled by UGGM with CAR model based covariance matrix, were introduced to specify the corresponding mixture component. It's noticeable that our model tried to account for the heterogeneity from the dental data hierarchically in two parts. The first part is through the mixture of flexible multivariate distributions, which gives much more flexibility for the distributions of the quadrant-wise response vectors than what was done in BGLVM (Zhang *et al.* (2007)) at the quadrant level. The second part is through the generalized latent variable models that is similar to what was done in Zhang's *et al.*(2007) at intermediate level. The choice of the model is reasonable, since the mixture model can take more than enough heterogeneity from the quadrant-wise response vectors, which makes the intermediate level Gaussian spatial latent vectors with CAR model based precision matrix structure sophisticated enough to explain the left heterogeneity of the dental data. Based on the chosen model, the conclusion of the hypothesis testing about spatial symmetry among quadrants is as follows: (1) Left-right spatial association relationship is the strongest, which is shown in terms of 95% credible intervals of the differences between left-right and across and the differences between left-right and up-down with lower bounds are all positive. (2) The difference of spatial association between across and up-down is not significant at type S error rate between 0% and 2.5% (Gelman (2006)), since the 95% credible interval of the difference between across and up-down includes zero.

3.6 Discussion

In this paper, we propose a flexible class of Bayesian mixture of generalized latent variable models for multivariate spatially correlated binary data with multi-level nested covariance structure. Our approach is to model the response variables in a hierarchical structure. At higher level, we model the quadrant-wise response vectors by a mixture of generalized latent variable models. At intermediate level, the response variables within quadrants are assumed to be from the canonical exponential family with the canonical parameters modeled by the generalized latent variable models. Meanwhile we imposed a multivariate spatial correlation structure on the latent variables, which induces the spatial correlation structures among the teeth within the same quadrant. Statistical inference is based on the posterior distributions of the parameters of interest. The spatial symmetry among quadrants is assessed by the similarity score defined in (31). There are two considerations in the model specifications. The first one is that we used the order constraints for the component marginal means to deal with the label switching issues for the Bayesian mixture model. The second consideration is the parameterizations for generalized latent variable models. For the identifiability of the model, we use sum to zero constraint fixed effects for the tooth position and assume spatial process has mean zero. Noninformative conjugate priors are applied for the parameters of interest, which will give a comparable inference results to the frequentist's as the sample size increases. We proposed four models to account for both number of mixture component at higher level and the covariance structure of Gaussian spatial latent vectors at intermediate level. The choices of the number of mixture component and covariance structure are based on DIC for missing data problem. Spatial hypothesis about the spatial symmetry of quadrants is based on simultaneous credible intervals for the differences of pairwise similarity scores of interest. The results from our model show the mixture of generalized latent variables models work fairly well and also comparable to the results in

existing literatures. It concluded that the left-right spatial association is the strongest and the spatial associations for across and up-down are not different significantly at type S error rate between 0% and 2.5% (Gelman (2006)). For the data example, we have assumed that the mixture component allocation process $\{Q_i : i = 1, ..., n\}$ at higher level and $\{T_{im} : m = 1, ..., M, i = 1, ..., n\}$ at intermediate level are sufficient to generate flexible multivariate distribution and induce dependence among teeth to account for the wide heterogeneities in the dental data. It would be interesting to introduce different probability models to latent variables at both higher and intermediate level. For instance, non-Gaussian latent process to model the underlying spatial dependence among teeth, which can lead to a richer class of the latent processes $\{T_{im} : m = 1, ..., M, i = 1, ..., n\}$. Finally, Other approaches for dealing with label switching problems associated with Bayesian mixture model may be interesting. It will be optimal when the model selection is simultaneous through either Reversible Jump Monte Carlo Markov Chain (MJMCMC)(Green *et al.* (1995)) or Birth and Death Monte Carlo Markov Chain (BDMCMC) (Stephens (2002)). It will be more interesting to consider the symmetry pattern of quadrants for a longitudinal study, which will lead to the spatial-temporal analysis.

CHAPTER 4

Discussion and Future Research

4.1 Bayesian generalized latent variable models

We have described generalized latent variable models for analyzing multilevel spatially correlated binary outcomes, i.e., the multivariate binary caries experience outcomes from STM project, which is similar to the mixed model with random effects being two levels of Gaussian spatial latent vectors at both a quadrant level and tooth nested within quadrant level. It is noticeable that our model is formulated in a hierarchically dynamic structure which is not only feasible but also relatively easier within Bayesian framework, when compared to Frequentist's approach where multilevel dynamic model is either very difficult or infeasible to formulate. The hierarchial structure of our model's specification makes our approach valid for the dental data with multilevel dependence among the subunits of interest, because it approximates the way in which the multilevel correlated binary outcomes were generated. Our approach can be viewed as a graph with three levels of tree structure. At the higher level, there exists a quadrant level Gaussian spatial latent vector that tights the four quadrant-wise binary response vectors together to induce the dependence among the quadrants and generate flexible multivariate distributions for each response vector.

At this level, our model provides both fixed effects corresponding to quadrant location and random effects presented by the higher level Gaussian spatial latent vector. Conditional on the Gaussian spatial latent vector at quadrant level, the quadrant-wise response vectors are mutually independent. The joint probabilistic behavior of the quadrant level Gaussian spatial latent vector is given by the UGGMs with mean vector of zeros and unstructured covariance matrix. At the intermediate level, there exist four Gaussian spatial latent vectors that are nested within corresponding quadrants in which the toot is located. In other words, the four intermediate level Gaussian spatial latent vectors are characterized by quadrant index. Each of the four latent vectors is used to tight the corresponding five binary caries response variables together to induce the dependence among the teeth within the same quadrant and generate flexible distributions for each response variable. At this level, our model provides both fixed effects for tooth location and random effects, i.e., the intermediate level Gaussian spatial latent vector nested within the corresponding quadrant, that generates flexible distributions for binary caries experience outcomes and induce the dependence among the teeth within the same quadrant. For the model identifiability, it is assumed that the Gaussian spatial latent vectors at intermediate level are mutually independent given the Gaussian spatial latent vector at quadrant level. Conditional on the Gaussian spatial latent vectors at both higher and intermediate level, all the binary response of caries experience in the mouth are mutually independent. This hierarchical model specification makes complete likelihood approach feasible, which will improve the efficiency of the estimation of the model parameters. At the lower level, a liner mixed model is specified to describe the log odds of the caries experience for each tooth of interest. An important feature of our model is that it allows irregularly spaced multilevel measurements under different spatial configurations, where the measurements are characterized by a hierarchical spatial dependence structure. The common way to implement the generalized latent variable models is through EM

algorithm in frequentist's framework, where the marginal likelihood is approximated by using an adaptive Gauss-Hermite quadrature approach to numerically integrate out the low dimensional latent variables in the model. For a high dimensional latent variable models, a Monte Carlo EM approach is applied instead. It is known that latent variable models are only locally identifiable and hierarchical models have complex structures, which lead to some consequences, i.e., local optimizer and singular information matrix. In order to obtain valid inference, we implemented our model within Bayesian framework via *WinBUGS*, since Bayesian inference is always feasible as long as the MCMC algorithm converges. Meanwhile, Bayesian makes it much easier to specify the hierarchial model than under frequentist's framework. It is also easy to incorporate missing data in *WinBUGS* through replacing $y_{missing}$ by the posterior sample from $p(y_{missing}|y_{observed};\theta)$. The implement of the model is within Bayesian framework via *WinBUGS* with noninformative conjugate proper priors.

Without an obvious multivariate distribution for the hierarchically spatially correlated binary response variables, multilevel correlated latent variables can be used to model the wide heterogeneity of the outcomes. Specifically, the dependence structure among the Gaussian spatial latent variables, at the higher level, that are used to induce dependence among four quadrants, is given by UGGMs with zero mean vector and unstructured covariance matrix. Similarly, the dependence structure for the Gaussian spatial latent vectors, at the intermediate level, i.e., the four spatial latent vectors accounting for the heterogeneity of teeth within the same quadrant, is given by UGGMs with zero mean vectors and covariance matrix that is either unstructured or structured under CAR model assumption, i.e., a Markovian type of covariance structure with taking spatial configuration into account. For the identifiability of the model, the two levels of spatial latent vectors are mutually independent with one another. The model is specified as below:

At the higher level, for the ith in the study, there exists a spatial latent vector $Q_i =$

$(Q_{i1}, ..., Q_{ij}, ..., Q_{iJ})'$ that is used to induce the dependence structure among quadrants and generate flexible multivariate distributions, $f_j(\cdot), j = 1, ..., J$, for quadrant-wise response vectors $y_i = (y'_{i1}, ..., y'_{ij}, ..., y'_{iJ})'$ with $y_{ij} = (y_{ij1}, ..., y_{ijk}, ..., y_{ijK})'$. The conditional joint multivariate distribution for the response vectors is specified as

$$f(y_i|Q_i; \theta, \Sigma_Q) = \prod_{j=1}^{J} f_j(y_{ij}|Q_{ij}; \theta, \Sigma_Q),$$

where the associations among the elements of Q_i are used to induce the associations among the four quadrants..

At the intermediate level, for each quadrant i, there exists a spatial latent vector $T_{ij} = (T_{i,1(j)}, ..., T_{i.k(j)}, ..., T_{i.K(j)})'$ that is used to induce the dependence structure among teeth nested within the jth quadrant and generate flexible distributions, $\{f_{k(j)}(\cdot) : k = 1, ..., K\}$,for binary response variable y_{ijk}. The conditional joint distribution for the binary response variables is specified as

$$f_i(y_{ij}|Q_{ij}, T_{ij}; \theta, \Sigma_T^j) = \prod_{k=1}^{K} f_{k(j)}(y_{ijk}|Q_{ij}, T_{i,k(j)}; \theta, \Sigma_T^j).$$

At the lower level, conditional on the higher level spatial latent vector $\{Q_i : i = 1, ..., n\}$ and intermediate level spatial latent vectors $\{T_{ij} : j = 1, ..., J, i = 1, ..., n\}$. The binary response variable y_{ijk} is mutually independent and from Bernoulli family with probability of success $\pi_{ijk} = P(Y_{ijk} = 1)$. That is,

$$(y_{ijk}|Q_{ij}, T_{ijk}; \alpha, \beta, \gamma) \sim Bernoulli(\pi_{ijk}) \models f_{j(i)}(y_{ijk}|Q_{ij}, T_{i,k(j)}; \theta),$$

where

$$logit(\pi_{ijk}|Q_{ij}, T_{i,k(j)}; \alpha, \beta, \gamma) = \alpha + \beta_j + \gamma_{k(j)} + Q_{ij} + T_{i,k(j)},$$

and $\theta = (\alpha, \beta', \gamma')'$ with constraints $\sum_{j=1}^{J} \beta_j = 0$ and $\sum_{k=1}^{K} \gamma_{k(j)} = 0$ for $j = 1, ..., J$.

Let $Q = \{Q_i : i = 1, ..., n\}$ with $Q_i = \{Q_{ij} : j = 1, ..., J\}$ and $T = \{T_i : i = 1, ..., n\}$ with $T_i = \{T_{ij} : j = 1, ..., J\}$ and $T_{ij} = \{T_{i,k(j)} : k = 1, ..., K\}$. If the model formulation is viewed as missing data problem where we treat Q and T as missing

covariates that are used to explain the wide heterogeneity of dental caries experience outcomes,then the complete likelihood is,

$$\begin{aligned} f(y,Q,T|\theta,\Sigma_Q,\{\Sigma_T^j\}) &= f(y|Q,T;\theta)p(Q|\Sigma_Q)p(T|\{\Sigma_T^j\}) \\ &= \prod_{i=1}^n \left\{ f(y_i|Q_i,T_i;\theta)p(Q_i|\Sigma_Q) \prod_{j=1}^J p(T_{ij}|\Sigma_T^j) \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^J \left\{ f_j(y_{ij}|Q_i,T_{ij};\theta)p(T_{ij}|\Sigma_T^j) \right\} p(Q_i|\Sigma_Q) \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^J \left\{ \prod_{k=1}^K \left\{ f_{k(j)}(y_{ijk}|Q_i,T_{i,k(j)};\theta) \right\} p(T_{ij}|\Sigma_T^j) \right\} p(Q_i|\Sigma_Q) \right\}. \end{aligned} \tag{4.1}$$

The distributions for Q_i and T_{ij} are given by UGGMS correspondingly as below:

$$Q_i|\Sigma_Q \sim N_J(\mathbf{0},\Sigma_Q);\ \ i=1,...,n,$$

and

$$T_{ij}|\Sigma_T^j \sim N_K(\mathbf{0},\Sigma_T^i).\ \ j=1,...J, i=1,...,n,$$

where Σ_Q is unstructured and Σ_T^j can be either unstructured or CAR model based.

Other consideration for parameterizations of the fixed effects θ and the probabilistic descriptions about the spatial latent vectors $\{T_{ij} : j=1,...,J, i=1,...,n\}$ may be chosen differently. However, as it can be expected, the results of the inference would not be affected substantially(Agresti(1997)). The model selection is based on DIC for missing data problems(Celeux, *et al.*(2006)). The optimal model selection needs to be based on *RJMCMC*(Green *et al.*(1995)) or *BDMCMC* (Stephens (2000)), which is essential a simultaneous model selection at each iteration of the MCMC posterior sampling algorithm.

4.2 Bayesian mixture of generalized latent variable models

Besides the generalized latent variable models, a finite mixture of distributions is another way to model response variables with wide heterogeneity. Finite mixtures of distributions are mathematical-based approaches to the statistical modeling of a

wide variety of random phenomena. They have been known as an extremely flexible method of modeling. The usefulness of finite mixture distributions in the modeling of heterogeneity in cluster analysis context is obvious. Mixture model provides a convenient semiparametric framework in which to model unknown distributional shapes, whatever the objective, whether it is density estimation or the flexible construction of Bayesian priors. Mixture model is also able to model quite complex distributions through an appropriate choices of its components and number of mixture components to represent accurately the local areas of support of the true distribution. It can handle situations where a single parametric family is unable to provide a satisfactory model for local variations in the observed data. In our approach, we assumed that each of the four quadrant-wise response vectors was from one of a certain number, say, $1 \leq M \leq 4$, of multivariate distributions with corresponding probability. The M multivariate distributions are characterized by M different situations which can accurately represent the corresponding local heterogeneity of observed binary vector. A convenient semiparametric way to incorporate the variability among these four observed quadrant-wise response vectors is to formulate their distributions uniformly in the form of a mixture of these M multivariate distributions. Specifically, the M multivariate distributions correspond to M underlying subgroups or subpopulations that where the four quadrant-wise response vectors are supposed to be able to identify if the subgroups actually exist; and each of the M multivariate distribution is corresponding to one component in the mixture model.

Mixture model can be viewed as missing data problem where the mixture component allocation process is latent. The latent process allocates each of the quadrant-wise response vector, y_{ij} to one of the mixture components, say, the mth component, which means y_{ij} can be characterized by the local situation, i.e., in terms of heterogeneity of the observed vector, associated with the mth underlying cluster. Hierarchically, at higher level, for the ith subject, there exists a mixture component allocation

latent process, $Q_i = (Q'_{i1}, ..., Q'_{ij}, ..., Q'_{iJ})'$ with

$$Q_{ij} = (Q_{ij1}, ..., Q_{ijm}, ..., Q_{ijM})' \sim Multi_M(1, (\pi_{j1}, ..., \pi_{jM})'), \quad j = 1, ..., J, i = 1, ..., n,$$

which means

$$(y_{ij}|Q_{ijm} = 1) \sim f_m(y_{ij}|\theta).$$

The complete distribution can be given as below;

$$f(y_i|Q_i; \pi, \theta) = \prod_{j=1}^{J} \left\{ \prod_{m=1}^{M} \{\pi_{jm} f_m(y_{ij}|Q_{ijm} = 1; \theta)\}^{Q_{ijm}} \right\}. \tag{4.2}$$

At the intermediate level, for the mth component that is a multivariate distribution, there exists a Gaussian spatial latent vector $T_{im} = (T_{i,1(m)}, ..., T_{i,k(m)}, ..., T_{i,K(m)})' \sim N_K(\mathbf{0}, \Sigma_T^m)$, which is used to generate flexible distribution for the K binary response variables that is from the exponential family (McCullagh and Nelder *et al.*) and induce the dependence among the J variables. At lower level, conditional on the allocation process and Gaussian spatial latent vectors, the conditional distribution for the binary caries experience outcome y_{ijk} is given by

$$(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)}; \theta) \sim Bern(logit^{-1}(\eta_{imk})) \models f_{k(m)}(y_{ijk}|Q_{ijm} = 1, T_{i,k(m)}; \theta),$$

where $\eta_{imk} = \alpha_m + \beta_k + T_{i,k(m)}$ and $\theta = (\alpha', \beta')'$, with constraints $\alpha_1 \leq \alpha_2 \leq, \cdot, \cdot, \cdot, \leq \alpha_M$ and $\sum_{k=1}^{K} \beta_k = 0$.

Let $Q = (Q'_1, ..., Q'_i, ..., Q'_n)'$ and $T = (T'_1, ..., T'_i, ..., T'_n)'$, then the complete likelihood is specified as

$$\begin{aligned} f(y, Q, T|\pi, \theta, \{\Sigma_T^m\}) &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \prod_{m=1}^{M} \{\pi_{jm} f_m(y_{ij}, T_{im}|\theta, \{\Sigma_T^m\})\}^{Q_{ijm}} \right\} \\ &= \prod_{i=1}^{n} \prod_{j=1}^{J} \left\{ \prod_{m=1}^{M} \left\{ \pi_{jm} \left\{ \prod_{k=1}^{K} f_{k(m)}(y_{ijk}, T_{i,k(m)}|\theta) \right\} p(T_{im}|\Sigma_T^m) \right\}^{Q_{ijm}} \right\} \end{aligned} \tag{4.3}$$

The model structure has two uncertainties from both mixture model at the higher level and generalized latent variable models at intermediate level. At the higher level, the number of mixture components is left unknown. At intermediate level, the

covariance matrix, $\{\Sigma_T^m : m = 1, ..., M\}$,for the generalized latent variable models can be either unstructured or CAR model based. The appropriate model needs to be determined by formal model selection criterion based DIC for missing data problem (Celeux, *et al.*(2006)). The implement of the model is within Bayesian framework via *WinBUGS* with noninformative conjugate proper priors.

Other consideration for parameterizations of the fixed effects θ and the probabilistic descriptions about the spatial latent vectors $\{Q_i : i = 1, ..., n\}$ and $\{T_{i,j} : j = 1, ..., J, i = 1, ..., n\}$ may be chosen differently. However, as it can be expected, the results of the inference would not be affected substantially(Agresti(1997)). The optimal model selection needs to be based on *RJMCMC*(Green *et al.*(1995)) or *BDMCMC* (Stephens (2000)), which is essential a simultaneous model selection at each iteration of the MCMC posterior sampling algorithm.

4.3 Missing data

In biomedical research, missing data problem is common and there are lots of literatures with different approaches discussed in this area but still the methods are not mature enough yet to handle general situations. Our model were built from the features of the dental data at hand, they have general applications to situations where multilevel discrete data recorded were spatially. The models were implemented via *WinBUGS* that allows missing values in the data set. What *WinBUGS* does to missing values is to replace the missing data by the random sample from its posterior distribution $p(y_{missing}|y_{observed}; \theta)$, which is essentially assumed that the missing is at random, i.e., the missing mechanism is noninformative. However, the missing data is very likely informative, since the teeth within the mouth share the same biological environment. In the presence of the informative missing data, our models need to be extended accordingly. In the future's work, we need to extend the model by incorpo-

rating the informative missing mechanism in a dropout process that is a parametric model for making inference about the missing values in the data set. The process for modeling the dropout pattern is problematic because the parameter that relates the measurement and the dropout process, say, λ, is always unidentifiable from the data at hand. Non-identifiability of the model always yields difficulties in the numerical optimization because of either flat or multimodal likelihood and singular information matrix, which makes the statistical inference infeasible in the frequentist's frame work. Under the Bayesian frame work, the statistical inference is always available as long as the MCMC algorithm converges that are used to sample the posterior samples of the the quantities from their proper posterior distributions that are related to the data at hand.

Bayesian approach for dealing with the informative missing data is known as the selection model (Arminger *et al.*, 1995), which requires the terms representing the non-response mechanism be included explicitly in the likelihood. Best *et al.*(1996) discussed the selection model for informative non-responses in a study of dementia and cognitive decline in the elders. They viewed the full model as two submodels; one representing the substantive relationship of interest and one reflecting the missing data process, with the possibly unobserved response variable representing the common link between the two submodels. Such a model may be readily expressed as a directed conditional independence graph, thus leading itself to Bayesian inference using MCMC approach. However, there is considerable current interest in the topic of informative drop-out(Diggle and Kenward (1994)) in which some argue that any attempt to learn about the selection mechanism will be heavily dependent on modeling assumptions, and that it is preferable to conduct sensitivity analysis to alternative plausible mechanisms. Meanwhile, the MCMC approach can easily provide predictive distributions for any variable of interest and, unlike approaches based on maximum likelihood or empirical Bayes, the MCMC predictions fully account for uncertainty in

both the model and the parameter estimations. Since the data often can not provide much information for estimating the parameters of the models for non-response mechanism, informative prior distributions for the parameters of interest in the selection models are used to facilitate the posterior sampling algorithm based on MCMC. So sensitivity analysis for the priors in the selection model is essential for the validity of the Bayesian analysis for model the non-response mechanism that is incorporated explicitly in the likelihood.

The future work will intend to develop a more general statistical procedure for assessing the sensitivity for both the non-response mechanism learning process and the informative priors used in the selection models. The procedure may be based on either different model selection criteria, for instance, DIC for missing data problem and posterior predictive checking, or dynamic algorithms based on *RJMCM* (Green(1995)) and *BDMCMC* (Stephens (2000)) for simultaneous model selection and parameter estimations.

4.4 Comparison between frequentist and Bayesian

It is well known that many standard statistical methods can be justified by both Bayesian and frequentist arguments. However, even when there is only one unknown parameter, there is a wide class of problems for which no Bayesian method can be found which satisfies the basic frequentist criterion (Bartholomew (1965)). Bartholomew raised two important questions in the comparisons between Bayesian and frequentist when discrepancy arose. The first one is the practical question of whether the discrepancy between the two approaches is ever such as to lead to widely differing conclusions. The second is concerned with the reason for the two approaches to inference giving different results in some cases but not in the other. The two questions is also of great interest to be addressed in our future work. The starting point

for this work will be considering the differences in the statistical thinking of the two statistical schools. For instance, suppose the observations $y = (y_1, ..., y_i, ..., y_n)'$ on a continuous random variable with density function $f(y|\theta)$ and consider the Bayesian and frequentist solution to the problem about making an inference about θ.

The Bayesian first specifies a prior for θ then combining this with the likelihood to obtain a posterior distribution which enable people to make a probability statement about θ of the form

$$P(\hat{\theta} \leq \theta_\alpha(y)|y) = \alpha_b \tag{4.4}$$

where α_b denotes a degree of belief. The major problem for Bayesian is to select a prior density $\pi(\theta)$ to express his ignorance about θ. Kass *et al.*(1995) reviewed several methods for determining a suitable prior distributions for the parameters of interest. For instance, based on Jerreys's rule, if we are ignorant about θ then we are ignorant of about any function of θ. This leads to him to formulate the invariant principle, i.e., $\pi(\theta) \propto \sqrt{I(\theta)}$ where $I(\theta)$ is the Fisher's information function.

The frequentist who wishes to make a statement of the form (6.10) is precluded from treating θ as a random variable as it was treated by Bayesian. He must try to find a statistics $\theta(y)$ such that

$$P(\theta \leq \widehat{\theta_\alpha(y)}|\theta) = \alpha_f \tag{4.5}$$

where α_f indicates the probability is to be interpreted in a frequency sense. The frequentist's ignorance about θ is expressed by the fact that α_f is independent of θ. The statement $\theta \leq \theta_\alpha(y)$ is thus true in the long run with probability α_f for any sequence of θ's. In general, there are many functions $\theta(y)$ satisfying (6.11) and the frequentist's problem is to choose one of them. It may be possible to choose $\theta(y)$ such that

$$\int_{-\infty}^{\theta(y)} p(\theta|y)d\theta = \alpha$$

Where $p(\theta|y)$ is the posterior density of θ for some prior $\pi(\theta)$. If the statistics $\theta(y)$ is

chosen in the way (6.11) is true, we say the Bayesian inference in (6.10) has *frequency* or *confidence* property. Under these circumstances, the Bayesian and frequentist approaches are said to *agree*. Welch *et al.* (1963) gave the following necessary and sufficient conditions for agreement.

(1) It must be possible to write $f(y|\theta)$ in the form $f(s-\tau)$ where s and τ are monotonic functions of y respectively and with $-\infty < \tau, s < \infty$.

(2) The prior density of τ must be uniform over the real line.

In large sample size, it is known that the influence of the prior $\pi(\theta)$ for parameter θ on the form of the posterior density $p(\theta|y)$ diminishes as $n \to \infty$. This means that, under very general conditions, Bayesian statement of (6.10) has the confidence property in the limit as $n \to \infty$ and the approach to agreement is more rapid with n if $\pi(\theta) \propto \sqrt{I(\theta)}$. Gelman *et al.* (2004) discussed the asymptotic normality and consistency of the posterior mean and median. Under some regularity conditions, i.e., the likelihood is a continuous function of θ and that θ_0, the true value of the parameter, is not on the boundary of the parameter space, as $n \to \infty$, the posterior distribution of θ approaches normality with mean θ_0 and variance $(nI(\theta_0))^{-1}$, where $I(\theta_0)$ is the Fisher information evaluated at θ_0. In the limit of large n, the posterior mode, $\hat{\theta}$, approaches θ_0, and the curvature (observed information) approaches $nI(\theta_0)$. When the truth is included in the family of models being fitted, the posterior mode, the posterior mean and median, are consistent, asymptotically unbiased and efficient under mild regular conditions (Gelman *et al.*(2004)).

When sample sizes are small, the prior distribution is a critical part of the model specification. It can only be a serious discrepancy between Bayesian and frequentist methods if the density $f(y|\theta)$ does not satisfy Welch's condition (1), if the sample size is small, or possible, if it is determined sequentially. Bartholomew raised two objectives for the comparison between the two approaches. The first object is about the extent to which Bayesian and frequentist statement of the form (6.10) and (6.11)

may differ in small samples. The second object is the reason for the differences which occur and how they may be avoided. Lee and Song (2006) did a simulation study, which showed Bayesian inference for hierarchial models with small to moderate sample size has a better performance than frequentist's.

Bartholomew (1965) pointed three conclusions in terms of the agreement between Bayesian and frequentist. (a) For shape parameter in gamma distribution, Bayesian interval estimates gave good agreement even if sample size is one; for restricted location parameter and exponential mean the agreement was not so good, but can be improved by an appropriate chosen confidence interval, i.e, either "shortest interval" or "equal tails". (b) Coverage probability of a two-tailed Bayesian interval estimate depends on not only prior but also the way that the interval is chosen. (c) Agreement may be achieved by using a sequential rather than a fixed sample size experiment design. The numerical magnitude of differences between frequentist and Bayesian methods of inference can be practically related to (a) and (b). The reason for the discrepancy is given by (c). He also conjectured that agreement can be always obtained if a correspondence is established between the Bayesian's appropriate choice of prior distributions and the frequentist's choice of sampling rules.

APPENDICES

APPENDIX A

The First Appendix

A.1 *WinBUGS* code one for BGLVM

(with unstructured covariance matrix at intermediate level) for overall spatial symmetry assessment.

```
model{
  ### Gaussian Graphical Models at Quadrant level ###
        InvSigamaQ[1:I,1:I] ~ dwish(IQ[,],(I+1))
           muQ[1]<- 0
           muQ[2]<- 0
           muQ[3]<- 0
           muQ[4]<- 0
  ### Gaussian Graphical Models  at Tooth level ###
           InvSigamaT[1:J,1:J] ~ dwish(IT[,],(J+1))
           muT[1]<- 0
           muT[2]<- 0
           muT[3]<- 0
           muT[4]<- 0
```

```
        muT[5]<- 0
### Generalized Latent Variable Models ###
for(k in 1:N){
   Q[k,1:I] ~ dmnorm(muQ[1:I],InvSigamaQ[1:I,1:I])
   for( i in 1:I){
       T[k,i,1:J] ~ dmnorm(muT[1:J],InvSigamaT[1:J,1:J])
                 }
for( i in 1:I){
   for( j in 1:J){
       Lat[j,i,k]<-alpha+(beta[i]-mean(beta[]))+
                   (gamma[j,i]-mean(gamma[,i]))+Q[k,i]+T[k,i,j]
                 }
              }
for( i in 1:I){
      for( j in 1:J){
             logit(p[j,i,k])<-Lat[j,i,k]
             y[(k-1)*20+(i-1)*5+j] ~ dbin(p[j,i,k],1)
                      }
               }
             }
### Priors ###
    alpha ~ dnorm(0,0.001)
    for(i in 1:I){
        beta[i] ~ dnorm(0,0.001)
                  }
    for( i in 1:I){
          for( j in 1:J){
```

```
                    gamma[j,i] ~ dnorm(0,0.01)
                        }
                    }
  ### Spatial association assessment ###
    ## Spatial association assessment between Left and Right ##
       Tlam12 <- -InvSigamaQ[1,2]/sqrt(InvSigamaQ[1,1]*InvSigamaQ[2,2])
       Tlam34<- -InvSigamaQ[3,4]/sqrt(InvSigamaQ[3,3]*InvSigamaQ[4,4])
    ## Spatial association assessment between Upper and Down ##
       Tlam23<- -InvSigamaQ[2,3]/sqrt(InvSigamaQ[2,2]*InvSigamaQ[3,3])
       Tlam14<- -InvSigamaQ[1,4]/sqrt(InvSigamaQ[1,1]*InvSigamaQ[4,4])
    ## Spatial association assessment between Across quadrants ##
       Tlam13<- -InvSigamaQ[1,3]/sqrt(InvSigamaQ[3,3]*InvSigamaQ[1,1])
       Tlam24<- -InvSigamaQ[2,4]/sqrt(InvSigamaQ[2,2]*InvSigamaQ[4,4])
  ### Hypothesis Testing Overall Spatial Symmetry ###
   LRvsUD<-1/2*(Tlam12+Tlam34)-1/2*(Tlam23+Tlam4)
   LRvsA<-1/2*(Tlam12+Tlam34)-1/2*(Tlam13+Tlam24)
   AvsUD<-1/2*(Tlam13+Tlam24)-1/2*(Tlam23+Tlam14)
}
```

A.2 *WinBUGS* code two for BGLVM

(with CAR model based covariance matrix at intermediate level) for overall spatial symmetry assessment.

```
model{
 ### Gaussian Graphical Models at Quadrant level ###
       InvSigamaQ[1:I,1:I] ~ dwish(IQ[,],(I+1))
          muQ[1]<- 0
```

```
            muQ[2]<- 0
            muQ[3]<- 0
            muQ[4]<- 0
### Gaussian Graphical Models  at Tooth level ###
### with CAR assumption for precision matrix ###
            num[1]<- 1
            num[2]<- 2
            num[3]<- 2
            num[4]<- 2
            num[5]<- 1
            m[1]<- 1
            m[2]<- 1/2
            m[3]<- 1/2
            m[4]<- 1/2
            m[5]<- 1
            cumsum[1]<- 0
             for( i in 2:6){
                 cumsum[i]<-sum(num[1:(i-1)])
                 }
            for(k in 1:8){
                for(i in 1:5){
                pick[k,i]<- step(k-cumsum[i]-esp)*step(cumsum[i+1]-k)
                             }
                 C[k]<- 1/inprod(num[],pick[k,])
                        }
            esp<- 0.0001
            adj[1]<- 2
```

```
        adj[2]<- 1
        adj[3]<- 3
        adj[4]<- 2
        adj[5]<- 4
        adj[6]<- 3
        adj[7]<- 5
        adj[8]<- 4
        muT[1]<- 0
        muT[2]<- 0
        muT[3]<- 0
        muT[4]<- 0
        muT[5]<- 0
 ### Generalized Latent Variable Models ###
for(k in 1:N){
Q[k,1:4] ~ dmnorm(muQ[1:4],InvSigamaQ[1:4,1:4])
T[k,1,1:5] ~ car.proper(muT[],C[],adj[],num[],m[],prec,spat1)
T[k,2,1:5] ~ car.proper(muT[],C[],adj[],num[],m[],prec,spat2)
T[k,3,1:5] ~ car.proper(muT[],C[],adj[],num[],m[],prec,spat3)
T[k,4,1:5] ~ car.proper(muT[],C[],adj[],num[],m[],prec,spat4)
for( i in 1:I){
    for( j in 1:J){
        Lat[j,i,k]<-alpha+(beta[i]-mean(beta[]))+
                    (gamma[j,i]-mean(gamma[,i]))+Q[k,i]+T[k,i,j]
                }
            }
    for( i in 1:I){
         for( j in 1:J){
```

```
                logit(p[j,i,k])<-Lat[j,i,k]
                y[(k-1)*20+(i-1)*5+j] ~ dbin(p[j,i,k],1)
                    }
                }
            }
  ### Priors ###
      alpha ~ dnorm(0,0.01)
      for(i in 1:I){
                 beta[i] ~ dnorm(0,0.01)
                 }
      for( i in 1:I){
          for( j in 1:J){
                     gamma[j,i] ~ dnorm(0,0.01)
                        }
                    }
          prec ~ dgamma(0.005,0.001)
          spatmax<- 0.35
          spatmin<- -0.95
          spat1 ~ dunif(spatmin,spatmax)
          spat2 ~ dunif(spatmin,spatmax)
          spat3 ~ dunif(spatmin,spatmax)
          spat4 ~ dunif(spatmin,spatmax)
### Spatial association assessment ###
   ## Spatial association assessment between Left and Right ##
      Tlam12 <- -InvSigamaQ[1,2]/sqrt(InvSigamaQ[1,1]*InvSigamaQ[2,2])
      Tlam34<- -InvSigamaQ[3,4]/sqrt(InvSigamaQ[3,3]*InvSigamaQ[4,4])
   ## Spatial association assessment between Upper and Down ##
```

```
        Tlam23<- -InvSigamaQ[2,3]/sqrt(InvSigamaQ[2,2]*InvSigamaQ[3,3])
        Tlam14<- -InvSigamaQ[1,4]/sqrt(InvSigamaQ[1,1]*InvSigamaQ[4,4])
      ## Spatial association assessment between Across quadrants ##
        Tlam13<- -InvSigamaQ[1,3]/sqrt(InvSigamaQ[3,3]*InvSigamaQ[1,1])
        Tlam24<- -InvSigamaQ[2,4]/sqrt(InvSigamaQ[2,2]*InvSigamaQ[4,4])
    ### Hypothesis Testing Overall Spatial Symmetry ###
     LRvsUD<-1/2*(Tlam12+Tlam34)-1/2*(Tlam23+Tlam4)
     LRvsA<-1/2*(Tlam12+Tlam34)-1/2*(Tlam13+Tlam24)
     AvsUD<-1/2*(Tlam13+Tlam24)-1/2*(Tlam23+Tlam14)
}
```

APPENDIX B

The Second Appendix

B.1 *WinBUGS* code one for BMGLVM

(with 3 components and unstructured covariance matrix at intermediate level) for overall spatial symmetry assessment.

```
model{
    for( n in 1:N){
      ### Mixture models (for  "mth" mixture( with M components)) ###
      ### at Quadrant level ###
      for( i in 1:I){
      ### Mixture models (for  "kth" mixture( with K components)) ###
        ### at Tooth level ###
           for( j in 1:J){
           y[((n-1)*20+(i-1)*5+j)] ~ dbern(p[n,AQ[n,i],j])
                        }# End of positions index #
            APQ[n,i,1:M] ~ ddirch(alphaQ[])
            AQ[n,i] ~ dcat(APQ[n,i,] )
                      }# End of quadrants index #
```

```
      Q12[n]<- equals(AQ[n,1],AQ[n,2])
      Q13[n]<- equals(AQ[n,1],AQ[n,3])
      Q14[n]<- equals(AQ[n,1],AQ[n,4])
      Q23[n]<- equals(AQ[n,2],AQ[n,3])
      Q24[n]<- equals(AQ[n,2],AQ[n,4])
      Q34[n]<- equals(AQ[n,3],AQ[n,4])
              }# End of Subjects index #
    ### Mixture Components Specification via ###
    ### GLVMs with Unstructured Covariance ###
      theta[1:J] ~ dmnorm(mu[],invR[,])
      alpha1 ~ dnorm(0,tau)
      loca1 ~ dnorm(0,tau)I(0,)
      loca2 ~ dnorm(0,tau)I(,0)
      alpha[1]<- alpha1
      alpha[2]<- alpha1+loca1
      alpha[3]<- alpha1+loca2
  for( n in 1:N){
    for( m in 1:M){
      T[n,m,1:5] ~ dmnorm(muT[1:5],InvSigamaT[1:5,1:5])
      for(j in 1:J){
         logit(p[n,m,j])<-alpha[m]+theta[j]-mean(theta[])+T[n,m,j]
                  }
               }
            }
### Priors ###
  InvSigamaT[1:5,1:5] ~ dwish(IT[,],6)
  tau ~ dgamma(0.01,0.01)
```

```
    muT[1]<- 0
    muT[2]<- 0
    muT[3]<- 0
    muT[4]<- 0
    muT[5]<- 0
 ### Similarity Assessment ###
   MQ12<- mean(Q12[])
   MQ13<- mean(Q13[])
   MQ14<- mean(Q14[])
   MQ23<- mean(Q23[])
   MQ24<- mean(Q24[])
   MQ34<- mean(Q34[])
 ### Hypothesis Testing Overall Spatial Symmetry ###
  LRvsUD<- 1/2*(MQ12+MQ34)-1/2*(MQ23+MQ14)
  LRvsA<- 1/2*(MQ12+MQ34)-1/2*(MQ13+MQ24)
  AvsUD<- 1/2*(MQ13+MQ24)-1/2*(MQ23+MQ14)
  }
```

B.2 *WinBUGS* code two for BMGLVM

(with 3 components and CAR model based covariance matrix at intermediate level) for overall spatial symmetry assessment.

```
model{
    for( n in 1:N){
      ### Mixture models (for  "mth" mixture( with M components)) ###
      ### at Quadrant level ###
      for( i in 1:I){
```

```
      ### Mixture models (for  "kth" mixture( with K components)) ###
      ### at Tooth level ###
        for( j in 1:J){
        y[((n-1)*20+(i-1)*5+j)] ~ dbern(p[n,AQ[n,i],j])
                    }# End of positions index #
            APQ[n,i,1:M] ~ ddirch(alphaQ[])
            AQ[n,i] ~ dcat(APQ[n,i,] )
                     }# End of quadrants index #
         Q12[n]<- equals(AQ[n,1],AQ[n,2])
         Q13[n]<- equals(AQ[n,1],AQ[n,3])
         Q14[n]<- equals(AQ[n,1],AQ[n,4])
         Q23[n]<- equals(AQ[n,2],AQ[n,3])
         Q24[n]<- equals(AQ[n,2],AQ[n,4])
         Q34[n]<- equals(AQ[n,3],AQ[n,4])
                }# End of Subjects index #
### Mixture Components Specification via GLVMs under CAR Model ###
        theta[1:J] ~ dmnorm(mu[],invR[,])
        alpha1 ~ dnorm(0,tau)
        loca1 ~ dnorm(0,tau)I(0,)
        loca2 ~ dnorm(0,tau)I(,0)
        alpha[1]<- alpha1
        alpha[2]<- alpha1+loca1
        alpha[3]<- alpha1+loca2
   for( n in 1:N){
     for( m in 1:M){
     T[n,m,1:5] ~ car.proper(muT[],C[],adj[],num[],invm[],prec,spat[m])
     for(j in 1:J){
```

```
 logit(p[n,m,j])<- alpha[m]+theta[j]-mean(theta[])+T[n,m,j]
              }
             }
           }
### CAR models specification ###
        num[1]<- 1
        num[2]<- 2
        num[3]<- 2
        num[4]<- 2
        num[5]<- 1
        invm[1]<- 1
        invm[2]<- 1/2
        invm[3]<- 1/2
        invm[4]<- 1/2
        invm[5]<- 1
        cumsum[1]<- 0
       for( i in 2:6){
             cumsum[i]<- sum(num[1:(i-1)])
                    }
        for(k in 1:8){
             for(i in 1:5){
        pick[k,i]<- step(k-cumsum[i]-esp)*step(cumsum[i+1]-k)
                         }
             C[k]<- 1/inprod(num[],pick[k,])
                    }
        esp<-0.0001
        adj[1]<- 2
```

```
            adj[2]<- 1
            adj[3]<- 3
            adj[4]<- 2
            adj[5]<- 4
            adj[6]<- 3
            adj[7]<- 5
            adj[8]<- 4
            muT[1]<- 0
            muT[2]<- 0
            muT[3]<- 0
            muT[4]<- 0
            muT[5]<- 0
 ### Priors ###
  prec ~ dgamma(0.005,0.001)
  spatmax<- 0.35
  spatmin<- -0.95
  spat1 ~ dunif(spatmin,spatmax)
  spat2 ~ dunif(spatmin,spatmax)
  spat3 ~ dunif(spatmin,spatmax)
  spat[1]<- spat1
  spat[2]<- spat2
  spat[3]<- spat3
  tau ~ dgamma(0.001,0.001)
### Similarity Assessment ###
  MQ12<- mean(Q12[])
  MQ13<- mean(Q13[])
  MQ14<- mean(Q14[])
```

```
  MQ23<- mean(Q23[])
  MQ24<- mean(Q24[])
  MQ34<- mean(Q34[])
### Hypothesis Testing Overall Spatial Symmetry ###
 LRvsUD<- 1/2*(MQ12+MQ34)-1/2*(MQ23+MQ14)
 LRvsA<- 1/2*(MQ12+MQ34)-1/2*(MQ13+MQ24)
 AvsUD<- 1/2*(MQ13+MQ24)-1/2*(MQ23+MQ14)
 }
```

BIBLIOGRAPHY

Aitkin, M. and Rubin, D., B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the royal Statistical Society* B 47, 67-75.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6, 251-262.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood princ- iple. In *Proc. 2nd Int. Symp. Information Theory* (eds B. N. Petrov and F.Csμaki), pp. 267 -281.

Alan Agresti. (1997). A model for repeated measurements of a multivariate binary respo-nses. *Journal of American Statistical Association*, Vol. 92, No. 437, 315-321.

Alan Agresti. (2002). *Categorical Data Analysis*. Second Edition. New York: Wiley.

Alan Agresti and and D., Hitchcock. (2005). Bayesian inference for categorical data ana- lysis. *Statistical Methods and Application (Journal of the Italian Statistical Society)*.

Alan E. Gelfand and Penelope Vounatsou. (2003). Proper Multivariate Conditional Auto- regressive Models for Spatial Data Analysis. *Biostatistics* 4, 1, pp. 11-25.

Alan E. Gelfand and Sujit K. Sahu. (1994). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, Vol. 94, No. 445, pp. 247-253.

Anderson, T., W. (1971). *An Introduction to Multivariate Statistical Analysis*, 2nd edition. New York: Wiley.

Anders Ekholm, Peter W., F., Smith and John W. McDonald. (1995). Marginal regression analysis of a multivariate binary response *Biometrika*, Vol. 82, No. 4.,pp. 847-854.

Anders Skrondal and Sophia Rabe-Hesketh. (2004). *Generalized Latent Variable Modeling*. New York: Chapman and Hall.

Andrew, Gelman, John, B. Carlin, Hal, S. Stern and Donald, B. Rubin. (2004). *Bayesian Data Analysis* Chapman & Hall/CRC.

Andrew Gelman and Francis Tuerlincks. Type S error rates for classical and Bayesian single and multiple comparison procedures. *Working paper*.

Arminger, G., Clogg, C., C. and Sobel, M., E. (eds). (1995). *Handbook of Statistical Modeling for the Social and Behavioral Science.* New York: Plenum.

Bartholomew, D., J. (1965). A comparison of some Bayesian and frequentist inferences. *Biometrika*, 52, 1 and 2, 19-35.

Bartholomew, D., J. (1984a). The Foundations of Factor Analysis, *Biometrika*, 71, 221-232.

Bartholomew, D., J. (1984b). Scaling Binary Data using a Factor Model, *Journal of the Royal Statistical Society, Series B*, 46, 120-123.

Bartholomew, D., J. (1988). The sensitivity of latent trait analysis to the choice of prior distribution. *British Journal of Mathematical and Statistical Psychology*, 41,101-107.

Bartholomew, D., J. (1994). Bayes's theorem in latent variable modeling. *Aspects of uncertainty: A tribute to D.V. Lindley.* Freeman,P.R. & Smith, A., F., M., Ed. London: Wiley.

Bartholomew, D., J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis.* Edward Arnold Publishers Ltd.

Bayes, T., R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53, 370- 418. Reprinted in *Biometrika*, 45, 243-315,1958.

Bedrick, E., J., Christensen, R., and Jonson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association,* 91, 1450-1460.

Best, N., G., Spiegelhalter, D., J., Thomas, A. and Brayne, C., E., G. (1996). Bayesian analysis of realistically complex models. *Journal of Royal Statistical Association*, A, 159, part 2, 323-342.

Bettina GrÄun and Friedrich Leisch. (2006). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics and Data Analysis.*

Box, G., E., P., and Tiao, G., C. (1973). Bayesian inference in statistical analysis, Reading, MA: Addison-Wesley.

Breslow, N., E. and Clayton, D., G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

Brook, D. (1964). On the distinction between the conditional probability and joint probability approaches in the specification of the nearest neighbor systems. *Biometrika* 51, 481-489.

Bruce Rannala. (2002). Identifiability of parameters in MCMC Bayesian inference of phylogeny. *systematic biology*, Vol.51, No.5, pp. 754-760.

Carey, V., Zeger, S. and Diggle, P. (1993). Modeling multivariate binary data with alternating logistic regressions. *Biometrika*, 80, 3, pp. 517-526.

Carmen Fernandez and Peter, J. Green. (2002). Modeling spatially correlated data via mixture: A Bayesian approach.

Celeux, G., Forbes, F., Robert, C., Titterington, D.M. (2006). Deviance Information Criteria for missing data models with discussion. *Bayesian Analysis*, in print.

Chuan zhou and Jon Wakefield. (2006). A Bayesian mixture model for partitioning gene expression data. *Biometrics* 62, 515-525.

Collett, D. and Stephiewska, K. (1999). Some practical issues in binary data analysis. *Statist. Med.* 18, 2209-2221.

Coull, B., A. and Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, 56, 73-80.

Cox, D., R. (1970). *The Analysis of Binary Data*. London: Methuen.

Cox, D., R. (1972). The analysis of multivariate binary data. *Applied Statistics*, Vol. 21, No. 2, pp. 113-120.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: Wiley. pp: 61-63.

Dale, J., R. (1986). Global cross-ratio models for bivariate discrete ordered responses. *Biometrics* 42, 909-917.

David, B., Dunson. (2007). Bayesian methods for latent trait modeling of longitudinal data. *Statistical Methods in Medical Research*, 16: 399-415.

Davidian, M. and Giltinan D., M. (2003). Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural, Biological, and Environmental Statistics* 8, 387-409.

Dawid, A.,P. (1979). Conditional Independence in Statistical Theory. (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 41,1-31.

Dempster, A. (1972). Covariance selection. *Biometrics* 28, 157-175.

Diebolt, J. and Robert, C., P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society* B 56, 363-375.

Diggle, P. (1992). Discussion of paper by K.-Y. Liang, S., L., Zeger and B., Qaqish. *J. R. Statist. Soc.* B 45, 28-39.

Diggle, P., Liang, L. and Zeger, S. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford.

Diggle, P. and Kenward, M., G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, 43, 49-93.

Dunson, D., B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society B* 62, 355-366.

Edwards, D. (1990). Hierarchical interaction models. *Journal of the Royal Statistical Society*. Series B, 52, 3-20.

Emily, L. Webb and Jonathan, J. Forster. (2006). Bayesian model determination for multivariate ordinal and binary data.

Escobar, M., D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association* 90, 577-588.

Everitt, B., S. and Hand, D., J. (1981). *Finite mixture distributions*. London: Chapman and Hall.

Eva Tzala and Nicky Best. (2007). Bayesian latent variable modeling of multivariate spatial-temporal variable in cancer mortality. *Statistical Methods in Medical Research* 2007; 1-22.

Fitzmaurice, G., M. & Laird, N., M. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika*, 80, 141-51.

Fisher, R., A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transitions of the Royal Society of London*, Series A, 222, 309-368.

Geman, S. and D. Geman. (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." IEEE Trans. Pattern Analysis and Machine Intelligence 6: 721-741.

Green, Peter. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.

Hobert, J., P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in Hierarchical linear mixed models. *Journal of the American Statistical Association 91*, 1461-1473.

Ibrahim, J., Chen, M. and Lipsitz, S. (2002). Bayesian methods for generalized linear models with covariates missing at random. *Canadian Journal of Statistics*, 30, 55-78.

Jansen R., C. (1993). Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics* 49, 227-231.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London.* Series A, 186, 453-461.

Jeroen K. Vermunt and Jay Magidson. (2004). Hierarchical mixture models for nested data structure.

Julian Besag, Peter Green, David Higdon and Kerrie Mengersen. (1995). Bayesian computation and stochastic systems. *Statistical Science* Vol. 10, No. 1, 3-66.

Kass, R. and Raftery, A. (1995). Bayes factors and model uncertainty. *Journal of American Statistical Association*, 90, 773-795.

Laplace, P., S. (1820). English translation: Philosophical essay on Probabilities (1951). New York: Dover.

Lauritzen, S., L. and Wermuth, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, 17, 31-57.

Lee S., Y. and Song X., Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* 2004; 39: 653-686.

Leroux, B. (2006). Analysis of Correlated Dental Data: Challenges and Recent Developments. *Statistical Methods for Oral Health Research, JSM 2006.*

Lesaffre, E. and Bogaerts, K. Spatial Correlations in Caries Attack Patterns in the Deciduous Dentition. *Biostatistics, K.U.Leuven.*

Liang, K., Y. and Zeger, S., L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

Liang, K., Y., Zeger, S., L. (1989). A class of logistic regression models for multivariate binary time series. *Journal of American Statistical Association*, 84, 447-451.

Liang, K., Y., Zeger, S., L. and Qaqish, B. (1992). Multivariate regression analysis of categorical data (with discussion). *Journal of the Royal Statistical Society*, Series B. 45, 3-40.

Lipsitz, R., Laird, N., M. & Harrington, P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* 78, 153-160.

Lipsitz, S. and Ibrahim, J. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, 83, 916-922.

Little, R. and Rubin, D. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Manski, C., F. and McFadden, D. (1981). *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: Massachusetts Institute of Technology Press.

Mardia, K. V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* 24, 265-284.

Mathias, Drton and Micheal Perlman. (2004). Model selection for Gaussian concentration graphs. *Biometrika,* 91,3, pp. 591-602.

McClachlan, G., J. and Basford, K., E. (1988). *Mixture models: inference and applications to clustering.* New York: Marcel Dekker.

McLachlan, G and Peel, D. *Finite Mixture Models.* Wiley (2001).

McCullagh, P. and Nelder, J., A. (1989). *Generalized Linear Models,* 2nd edition. New York: Chapman and Hall.

Morton, R. (1987). A generalized linear model with nested strata of extra-Poisson variation. *Biometrika*, 74, 247-257.

Moustaki, I. (1996). A Latent Trait and a Latent Class Model for Mixed Observed Variables, *British Journal of Mathematical and Statistical Psychology*, 49, 313-334.

Neuthaus, J., M., Hauck, W., W. and Kalbfleisch, J., D. (1992). The effects of mixture distribution misspecification when fitting mixed effects logistics models. *Biomatrika*, 79, 755-762.

O'Malley, A., James and Zaslavsky, Alan M. (2006). Domain-level covariance analysis for survey data with structured nonresponse. *Working paper.*

Paolo Giudici and Peter J. Green. (1999). Decomposable graphical Gaussian model determination. *Biometrika* 86, 4, pp. 785-801.

Pettitt, A., N., Tran, T., T., Haynes, M., A. and Hay, J., L. (2006). A Bayesian hierarchical model for categorical longitudinal data from a social survey of immigrants. *Journal of the Royal Statistical Society* A (2006), 169, Part 1, pp. 97-114.

Prentice, R., L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033-1048.

Richardson, S. and Green. P. (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society*, Series B. Vol. 59, No. 4. (1997), pp. 731-792.

Robert, C., P. (1996). Mixture of distributions: inference and estimation. In *Markov Chain Monte Carlo in Practice*, W.R. Gilks , S. Richardson, and D.P. Spiegelhalter (Eds.). London: Chapman & Hall, pp. 441-464.

Robert E. Kass; Larry Wasserman. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, Vol. 91, No. 435. pp. 1343-1370.

Roeder, K. and Wasserman, L. (1997). Practical density estimation using mixtures of normals. *Journal of the American Statistical Association* 92, 894-902.

Roy, J. (2006). Statistical Approaches for Dealing with Missing Tooth- and Surface Level Data in Caries Research. *Statistical Methods for Oral Health Research, JSM 2006.*

Sammel, M., D., Ryan, L., M., and Legler, J., M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society B* 59, 667-678.

Samuel, M. Manda, Rebecca, E., Walls and Mark, S., Gilthorpe. (2007). A full Bayesian hierarchical mixture model for the variance of gene differential expression. *BMC Bioinformatics.*

Scott L. Zeger and M. Rezaul Karim. (1991). Generalized Linear models with random effects; A Gibbs Sampling Approach. *Journal of the American Statistical Association, theory and Methods* Vol. 86, No. 413, 79-86.

Seam, M. O'Brien and David B. Dunson. (2004). Bayesian multivariate logistics regression. *Biometrics* 60, 739-746.

Spiegelhalter, D., J., Best, N., G., Carlin, B., P. and van der Linde A. (2002). Bayesian measures of model complexity and fit (with discussion) *Journal of Royal Statistical Society.* B. 64, 583-640.

Spiegelhalter, D., J., Thomas, A. and Best, N. (2003). *WinBUGS* Version1.4 User Manual. www.hrc-bsu.cam.ac.uk bugs.

Stephens, M. (2000). Bayesian analysis of mixtures with an unknown number of componentsan alternative to reversible jump methods, *The Annals of Statistics.* Volume 28, Number 1 (2000), 40-74.

Stiratelli, R., Laird, N. and Ware, J., H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, 40, 961 -971.

Stuart, R., Lipsitz and Garrett Fitzmaurice. (1994). An extension of Yule's Q to multivariate binary data. *Biometrics*, 50, 847-852.

Titterington, D., M., Smith, A., F. and Makov, U., E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.

Thompson T., J., Smith P., J. and Boyle J., P. (1998). Finite mixture models with concomitant information: assessing diagnostic criteria for diabetes. *Applied Statistics,* 47, 393-404.

Thomas, A., Best, N., Lunn, D., Arnold, R. and Spiegelhalter, D., J. (2004). *GeoBUGS* Version1.2 User Manual. www.hrc-bsu.cam.ac.uk/bugs.

Van Duijn Maj and Bockenholt U. (1995). Mixture models for the analysis of repeated count data. *Applied Statistics* 44, 47385.

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

Vanobbergen, J., Lesaffre, E., Garca-Zattera, M., J., Jara, A., Martens, L. and Declerck, J. (2007). Caries Patterns in Primary Dentition in 3-, 5- and 7-year-old Children: Spatial Correlation and Preventive Consequences.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer, (Springer series in statistics).

Vincent Carey, Scott L. Zeger and Peter Diggle. (1993). Modeling multivariate binary data with alternating logistic regressions. *Biometrika*, 80, 3, pp. 517-526.

Wang, P., Puterman, M. L., Cockburn, I. and Le, N.D. (1996). Mixed Poisson regression models with covariate rates. *Biometrics* 52, 381-400.

Wang, P. and Puterman M., L. (1998). Mixed logistic regression models. *Journal of Agricultural, Biological, and Environmental Statistics* 3, 175-200.

Wang, F., J. and Wall, M., M. (2003). Generalized common spatial factor model. *Biostatistics* 4, 569-582.

Welch, B., L. and Peers, H., W. (1963). On formulae for confidence points based on integrals of weighted likelihood. *Journal of Royal Statistical Association*, B, 25, 318-329.

West, M., Muller, P., and Escobar, M., D. (1994). Hierarchical priors and mixture models with application in regression and density estimation. In *aspects of Uncertainty: A tribute to D.V. Lindley,* A.F.M. Smith and P. Freeman (Eds.). New York: Wiley, pp. 363-386.

Wu, M., C. and Carroll, R., J. (1988). Estimation and comparison of change in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44,175-188.

Zabell, S., L. (1992). R. A. Fisher and the fiducial argument. *Statistical Science*, 7, 369-387.

Zhao, L., P. and Prentice, R., L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* 77, 642-648.

Zhao, Y., Staudenmayer, J., Coull, B., A. andWand, M., P. (2006). General Design Bayesian Generalized Linear Mixed Models. *Statistical Science* Vol. 21, No.1, 35-51.

Zhu, J., Erickhoff, J., C. and Yan, P. (2005). Generalized Linear Latent Variable Models for Repeated Measures of Spatially Correlated Multivariate Data. *Biometrics*, 61, 674-683.