Applications of Circular Distributions and Spatial Point Processes to the Analysis of Periodontal Data

by

Samopriyo Maitra

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Thomas M. Braun, Chair
Professor William V. Giannobile
Professor Peter Xuekun Song
Assistant Professor Veronica Berrocal

To my family & my advisor

ACKNOWLEDGEMENTS

Over the course of my graduate study I have been fortunate to have received the guidance and have worked under the supervision of various faculty members and researchers. I would like to express my heartfelt gratitude to my advisor Dr. Thomas Braun for all his generous advice, patience and guidance over the course of my PhD career. He has been a true inspiration and has been instrumental in my growth as a biostatistician and researcher. I have learnt a lot from taking his courses, working as a research assistant under his supervision as well as working with him towards my dissertation. He introduced me to the research topic of novel approaches for analyzing periodontal data, which provided me the opportunity to work on applied statistics, an area I had always wanted to pursue my research on. His guidance, feedback and patience has influenced everything I have done and will do as a researcher.

I am grateful to the committee members for their thoughtful insights and feedback on the dissertation. I would like to thank Dr. Berrocal, Dr. Giannobile and Dr. Song for kindly serving on my committee. Dr. Veronica Berrocal had encouraged me to apply spatial statistics in my research and thus introduced me to a branch of statistics I was previously unaware of. She has been extremely patient in helping me get an indepth understanding of spatial statistics and making me confident enough so that I could apply it to my research. I am grateful to Dr. William Giannobile for giving us access to the periodontal dataset motivating our research and for giving me the opportunity to work as a Research Assistant at the School of Dentistry. It helped me understand the motivation behind my research, formulate my research problems and

gave an opportunity to work with periodontists and clinicians. I appreciate Dr. Peter Song for his feedback regarding my dissertation. I have taken several courses under him and his lectures on longitudinal data analysis were extremely useful. His ideas on applications of various statistical techniques to analyze periodontal data helped me immensely in my research.

I owe my deepest gratitude to many faculty members and researchers at the University who had helped me immensely in my growth as a biostatistician. Dr. Rod Little was my mentor in the first couple of years and he helped me select my courses and made me well prepared so that I could be an able researcher in future. Dr. Mousumi Banerjee and Dr. Bhramar Mukherjee had helped me take many informed decisions throughout my PhD career and their encouragement, feedback and advice in deciding my future career path has been highly commendable. I am grateful to all the faculty and statistical consultants at the Center for Statistical Consultation and Research where I have worked as a Research Assistant over the last few years. Working and collaborating with them has appreciably increased my analytic skills as well as honed my skills in explaining statistics to non-statisticians. I am also thankful to my fellow students and friends at the Department for helping me throughout my graduate study and thus making my stay at the Department both fruitful and memorable.

I would like to acknowledge my family, particularly my parents Dr. Saumen Kumar Maitra, Dr. Tapati Maitra and brother Tapapriyo, for their love, encouragement and support in every walk of life. A special note of appreciation to my wife Kohinoor for always being there for me and motivating me in my gloomy days. My life as a graduate student has been extremely rewarding and I owe it to all the support and encouragement from my advisor, mentors, family members as well as the company of friends and well wishers both in and outside the department.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Applications of Circular Distributions and
Spatial Point Processes to the Analysis of
Periodontal Data

by

Samopriyo Maitra

Chair: Thomas M. Braun

The dissertation focuses on three novel approaches for analyzing data arising from studies of periodontal disease, a common cause of tooth loss in adults, in order to provide periodontists with a better understanding of periodontal disease and improving the prevention and treatment of the disease. Our first two methods focus on identifying regions of the mouth that are most susceptible to periodontal disease and thus determining locations of the mouth where localized treatments for the disease can be best applied. First, we assume the directions of diseased teeth to be observations from a unimodal von Mises distribution, the mean of which is a function of mouth-level covariates. Because multiple teeth from a subject are correlated, we use a bias-corrected generalized estimating equation approach to obtain robust variance estimates for our parameter estimates. Second, we extend our methods to model asymmetry and multimodality by assuming the directions of diseased teeth follow a Generalized von Mises distribution. We use generalized estimating equations to model periodontally diseased locations and use a model selection criterion in order to determine the appropriate number of modes. As applied to our motivating set of data,

we find that periodontal disease tends to be located at the back of both sides of the upper jaw, as well as at the middle of the lower jaw. Third, we propose using point pattern data analysis methods to study the association between clinical attachment level (CAL) and bone level (BL) and how the association varies among different types of teeth. Applying these methods to our motivating data, we find that CAL and BL are similar for molars and bicuspids, with the similarity stronger for bicuspids than molars. We also found a substantial similarity between the CAL and BL of molars with the CAL and BL of bicuspids. The results suggest that measurements on a single tooth can be considered to be representative of the measurements obtained from the other teeth. Thus, fewer teeth would need to be part of a periodontal exam, thereby reducing the time and effort devoted to patient exams in future periodontal studies.

CHAPTER I

Introduction

Periodontal disease, commonly called gum disease, is the most common cause of tooth loss among adults and has a moderately high prevalence in the United States (Brown and Loe, 1993). It is an inflammation and infection that destroys the tissues that support the teeth, including the gums and the periodontal ligaments. The most mild form of periodontal disease is gingivitis or inflammation of the gingiva, which with poor dental hygiene and left untreated leads to loosening or loss of teeth and thus develops into periodontitis. Periodontal disease is primarily caused by the deposition of bacterial plaque, a sticky, colorless film that is usually present on the surface of teeth. Regular brushing and flossing helps reduce the plaque biofilms. However, in the absence of routine oral care, the plaque can harden to form calculus or tartar that can irritate and inflate the gingiva. This results in the gingiva getting tender or swollen and thus the gums pull away from the teeth ultimately resulting in tooth loss.

The presence of a number of pathogens in the plaque has been linked with the severity of periodontal disease. Though a single pathogen has not been identified to be most predictive of periodontal health, Socransky et al. (1998) found that the bacterial species *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola*, commonly referred to as the "red complex," had a high association with the

severity of periodontal disease. Various studies have shown that smoking, tobacco use are detrimental for periodontal health and the risk of periodontal disease also increases with aging or having diabetes mellitus. A general overview of risk factors for periodontitis are stated in Genco (1996) and Timmerman and Van der Weijden (2006).

The development and progression of periodontal disease is commonly quantified via longitudinal collection of several clinical parameters, most common of which are clinical attachment level (CAL), pocket depth (PD), bleeding on probing (BOP) and alveolar bone level (BL). The point where the crown (enamel) of the tooth connects with the root (cementum) of the tooth is known as the cementoenamel junction (CEJ). Any detachment of the gingiva from the cementum produces a gap between the gum and the tooth, commonly referred to as a pocket. PD quantifies the depth (in whole millimeters) of the pocket, while CAL quantifies the vertical distance (in whole millimeters) from the base of the pocket to the CEJ. Figure 1.1, presented in Arora et al. (2009), illustrates the clinical parameters and compares the measures between a healthy tooth and a periodontally diseased tooth.

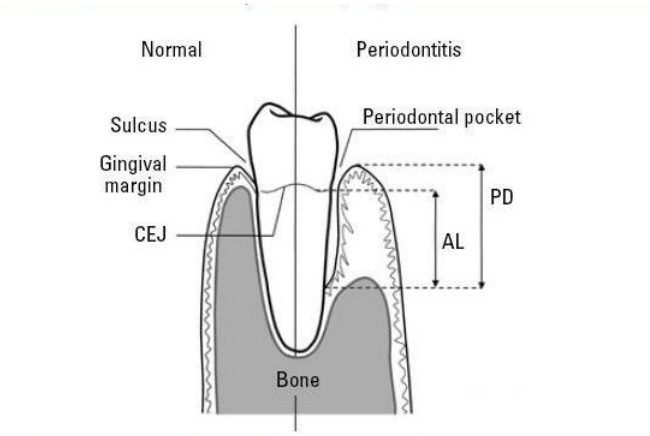

Figure 1.1: Diagram comparing clinical parameters in healthy (left) and periodontally diseased (right) tooth. AL and PD refers to Clinical Attachment Level and Pocket Depth respectively while CEJ refers to cementoenamel junction. This diagram is presented in Arora et al. (2009).

In clinical studies of periodontal disease, all periodontal parameters are typically measured at six sites around every tooth. At each probe, bleeding of the gums or the absence of it gives an indication of active periodontal disease, leading to the binary indicator of bleeding on probing (BOP). BL is calculated using radiographs and is the length (in millimeters) from the CEJ to the crest of alveolar bone and is measured on two surfaces of the tooth. Thus, a high value of CAL, PD, BOP or BL is detrimental to periodontal health and ideally periodontists would expect all the clinical measures to convey similar implications. Although no conclusive definition of periodontitis exists, the American Academy of Periodontology has a site-specific, three-category definition of periodontitis as being mild, moderate or severe according to whether CAL is less than 3 mm, greater than 3 mm but less than 5 mm, and greater than 5 mm respectively (Wiebe and Putnins, 2000).

Various researchers and statisticians have been interested in understanding the orientation of affected teeth in the mouth, as well as identifying the mechanism and determining the risk factors for periodontal disease progression. Such research would create a better understanding of the disease and hopefully yield therapeutic benefits leading to prevention of periodontal disease. Since the measures from multiple teeth are correlated, the association of periodontal disease with various risk factors could be studied using either a marginal model (Hoffman et al., 2001) or a subject specific model (Gillthorpe et al., 2003). In addition to measurements from multiple teeth being correlated, there is an association between different types of teeth depending on their position and function, and researchers have been interested in incorporating these various kinds of associations into modeling of periodontal data. Reich et al. (2007) proposed methods in which CAL data were modeled using a conditionally autoregressive (CAR) prior model. Reich and Bandyopadhyay (2010) used spatial factor analysis methods in order to model the number and location of missing teeth, assuming the number and location of missing teeth to be indicative of a subject's

periodontal health.

There is also considerable research interest in studying the locations of affected teeth and determining locations of the mouth that are most prone to periodontal disease. Loe and Brown (1991) and Thomson et al. (2000) observed that teeth at the back of the mouth, and particular to one side of the mouth showed the greatest indications for periodontal disease. Using basic summary statistics and simplistic statistical models, it has been shown that intraoral distribution of periodontal disease is associated with smoking (Torrungruang et al., 2011), race (Loe et al., 1978), age (Spalj and Plancak, 2003) and pathogen levels (Riviere et al., 1995). However these studies required prior knowledge of which locations and specific types of teeth to examine, and failed to incorporate multiple measurements from subjects. Furthermore, the majority of the results were anecdotal, descriptive, and lacked formal statistical inference. All these studies motivated us to pursue research addressing these issues.

Thus in Chapter II via a unified statistical model, we aim to address the limitations of the studies cited previously. We propose methods for the identification of periodontally diseased locations as well as quantify their dependence with mouth-level characteristics using a regression approach that incorporates multiple measurements from a subject. Our statistical model not only allows us to estimate where periodontal disease is occurring, but also ascribe inferential concepts i.e. standard errors and p-values to our estimates. To the best of our knowledge, ours is the first attempt to specifically model where periodontal disease occurs and how the location varies according to different patient characteristics incorporating multiple measurements from a subject.

However a drawback of our approach is that we assume the distribution of affected teeth is unimodal i.e. there is a single region in the mouth that is most likely to be affected, which might be debatable. Mombelli and Meier (2001) and Darby et al. (2012) studied the distribution of periodontal disease and have noted a symmetry

among the left and right sides of the mouth with respect to the distribution of periodontal disease in patients suffering from severe periodontal disease. Various dental studies (Tomasi et al., 2007; Reich and Bandyopadhyay, 2010) have suggested that periodontal disease is most common particularly at the back of the mouth and thus it is possible that the distribution of affected teeth is multimodal. However, there have been no formal attempts to determine the exact distribution of affected teeth nor methods to verify if the distribution of diseased teeth is multimodal. Thus in Chapter III we propose methods to study the distribution of affected teeth in the mouth and to quantify the number and location of regions in the mouth that are most likely to be affected with periodontal disease. In both Chapters II and III we assume the mouth to be a circle with teeth as points located on the circumference and thus use circular statistical methods to study the association.

Furthermore, periodontists are interested in studying the correspondence between the different clinical measures and how they vary between different teeth. This is essential in order to determine the periodontal status as well as to validate their measures. Of particular interest is the comparison between clinical attachment level (AL) and alveolar bone level (BL) since both of them are used to determine the periodontal status but are evaluated in different ways, namely by periodontal probes and radiographs respectively. Renvert et al. (1981) found a high correlation between attachment level and radiographic bone length measures. Goodson et al. (1984) proposed that attachment loss precedes radiographic bone loss particularly during periods of periodontal disease activity. Jeffcoat (1992) and Hausmann et al. (1994) showed that there was a significant concordance between changes in bone level and attachment level. The results from the above-cited studies were based on either Pearson's correlation coefficients or two-way contingency tables and were based on subject-level data. To complement these existing approaches, we propose in Chapter IV to assess the association of CAL and BL using spatial point pattern data analysis methods,

and our approach allows for the incorporation of multiple tooth-level measurements from each subject. Our methods also allow for direct estimation of the association as a function of a binary covariate that indicates whether a tooth is a molar or biscupid, and also assess the difference in the association between the two types of teeth.

The clinical dataset motivating our methods was obtained from a clinical trial of periodontitis studied in Ramseier et al. (2009) and Kinney et al. (2011). It was a non-randomized observational study, conducted at the Michigan Center for Oral Health Research involving 50 periodontally healthy and 50 periodontally diseased subjects, based upon entry criteria specified in Ramseier et al. (2009). In the study, investigators gave periodontal exams and collected saliva samples from each of the 100 subjects at baseline, as well as six and twelve months after enrollment. There were various types of data collected on each subject. We have information about their demographic characteristics, clinical measures like CAL, BOP, BL and also various salivary biomarkers and pathogen levels. The main motivation behind the study was to assess the association between salivary biomarkers and red complex pathogens in the presence of periodontal disease. For our endeavor, we have analyzed the data in all the chapters and it also was the motivating dataset on which we based all our simulation examples.

CHAPTER II

Analysis of Periodontal Data using Unimodal Circular Statistics

2.1 Introduction

Periodontal disease is the most common cause of tooth loss among adults and has a prevalence of about 30-50 % in the United States (Brown and Loe, 1993). The most mild form of periodontal disease is gingivitis, or inflammation of the gingiva, which, in the absence of routine oral care to reduce plaque, frequently develops into periodontitis, which, if left untreated, leads to loosening and loss of teeth. The American Academy of Periodontology and National Institute of Dental and Craniofacial Research state that the primary cause of periodontal disease is the presence of bacterial plaque, a sticky, colorless film that is usually present on the surface of teeth.

The presence of a number of pathogens in the plaque has been linked with the severity of periodontal disease. Though a single pathogen has not been identified to be most predictive of periodontal health, Socransky et al. (1998) found that the bacterial species *Porphyromonas gingivalis*, *Tannerella forsythia* and *Treponema denticola*, commonly referred to as the "red complex," had a high association with the severity of periodontal disease. Recently, Ramseier et al. (2009) showed that the salivary biomarkers matrix metalloproteinase (MMP)-8 and MMP-9, together with red complex pathogens, were indicative of the presence of periodontal disease. A general overview of mouth-level risk factors for periodontitis are stated in Timmerman and

Van der Weijden (2006).

The development and progression of periodontal disease is commonly quantified via longitudinal collection of several clinical parameters, including clinical attachment level (CAL), pocket depth (PD), and a binary indicator of bleeding on probing (BOP). The point where the crown (enamel) of the tooth connects with the root (cementum) of the tooth is known as the cementoenamel junction (CEJ). Any detachment of the gingiva from the cementum produces a pocket. PD quantifies the depth (in whole millimeters) of the pocket, while CAL quantifies the vertical distance (in whole millimeters) from the base of the pocket to the CEJ. In clinical studies of periodontal disease, all periodontal parameters are typically measured at six sites around every tooth. Although no definitive definition of periodontitis exists, the American Academy of Periodontology has a site-specific, three-category definition of periodontitis as being mild, moderate, or severe according to whether CAL is less than 3 mm, greater than 3 mm but less than 5 mm, and greater than 5 mm respectively (Wiebe and Putnins, 2000).

Because the measures collected from multiple teeth of a subject are correlated, the association of periodontal disease with mouth-level risk factors could be determined using either a marginal model or a subject-specific model. For example, Hoffman et al. (2001) analyzed periodontal data using marginal methods, while Gillthorpe et al. (2003) used a random coefficient model in order to model periodontal disease progression. Furthermore, as clinical measures are collected on sites and teeth with a specific orientation in the mouth, existing statistical methods have focused on how to incorporate this orientation into the analysis of periodontal data. Reich et al. (2007) proposed methods in which CAL data were modeled using a conditionally autoregressive (CAR) prior model, and Reich and Hodges (2008) extended that idea to a spatiotemporal model to monitor the progression of CAL. Reich and Bandyopadhyay (2010) used spatial factor analysis methods in order to model the number and location

of missing teeth, assuming the number and location of missing teeth to be indicative of a subject's periodontal health. Recently, Zhang et al. (2011) used latent variable models for analyzing tooth level caries data.

Quantifying where periodontal disease occurs in the mouth would assist in identifying locations in the mouth most susceptible to periodontal disease and lead to better prevention of and treatment for the disease. For example, the work of Tiwari (2010) presents a method for delivering the antimicrobial agent metronidazole through the use of microspheres to localize the delivery of the drug. Such a localized delivery might also be used in subjects who are at higher risk of developing periodontal disease, but have yet to show symptoms of the disease, as a method for reducing the prevalence of periodontal disease. However, such an approach is only feasible once we are able to accurately determine the locations where periodontal disease is most likely to occur.

Although much published research exists on the location of periodontal disease, the majority of the results are anecdotal and descriptive and lack formal statistical inference. For example, Loe and Brown (1991) and Thomson et al. (2000) studied specific locations of the mouth and concluded that periodontal disease is most common at the back of the mouth. However, their work required *a priori* knowledge of which locations and specific types of teeth to examine, and lacked a mouth-wide assessment of all locations in the mouth. Additionally, various researchers have been interested in detecting variations in the intraoral distribution of periodontal disease associated with smoking (Torrungruang et al., 2011), race (Loe et al., 1978), age (Spalj and Plancak, 2003), pathogen levels (Riviere et al., 1995), and other patient characteristics (Tomasi et al., 2007). However, the statistical methods used in these examples were simplistic, relying upon summary statistics of clinical measures and comparing these statistics across different regions of the mouth and differing patient characteristics using t-tests and ANOVA, without simultaneously modeling all lo-

cations in the mouth. Hirotomi et al. (2010) used multilevel logistic regression to identify tooth-level and mouth-level factors affecting periodontal disease progression, and found that multi-rooted and maxillary teeth were at higher risk for periodontal disease progression, demonstrating continuing interest in identifying specific locations in the mouth that are more susceptible to periodontal disease.

Through a single statistical model, we aim to address the limitations of the studies cited previously. Our proposed methods will allow for the identification of periodontally diseased locations as well as quantify their dependence with mouth-level characteristics. In contrast to existing methods, we propose a regression approach that incorporates the multiple measurements from a subject and does not require any prior knowledge about which specific mouth locations to include. Furthermore, our methods allow us to identify the mean location of affected teeth, which quantifies the region in the mouth where one would expect periodontal disease to be pronounced in the study population and how the location varies with patient characteristics. Most importantly, through our use of a statistical model, we are able to not only estimate where periodontal disease is occurring, but ascribe inferential concepts, i.e. standard errors and p-values to our estimates. To the best of our knowledge, ours is the first attempt to specifically model where periodontal disease occurs and how the location varies according to different patient characteristics.

We apply circular statistics, a specific area of directional statistics, for analyzing data arising from periodontal studies. The primary motivation of our approach arises from the orientation of teeth in the mouth. We assume the mouth to be a circle and the teeth to be points lying on the circumference of the circle. The periodontal disease status of an individual tooth is determined on the basis of its tooth-averaged CAL. Using a circular-linear regression model (Fisher and Lee, 1992), we model the direction of diseased teeth in the mouth as a function of mouth-level covariates. Using this model, we can determine the mean location of periodontal disease in the

mouth and also identify important predictors of where periodontal disease occurs in the mouth. Section 2.2 describes the details of our proposed methods, Section 2.3 contains the results of simulation studies of our proposed methods as well as application of our methods to an actual dataset and Section 2.4 contains concluding remarks.

2.2 Methods

2.2.1 Circular Location Model

In order to incorporate the spatial orientation of teeth in the mouth, we envision the mouth to be a circle and assume the teeth are distributed on the circumference as shown in Figure 2.1. We use the Universal Numbering System adopted by the American Dental Association, in which the 32 teeth in an adult are numbered sequentially from 1 to 32, with the numbers 1-16 referring to the teeth in the upper jaw (maxillary) and the numbers 17-32 referring to the sixteen teeth in the lower jaw (mandibular). As wisdom teeth (teeth 1, 16, 17 and 32) are often removed in most adults even when healthy, these teeth are usually omitted from periodontal studies, leading to a maximum of 28 teeth measured in each subject.

These numbers allow us to designate a specific direction in radians for each tooth corresponding to its location in the mouth. If we assume that all 32 teeth are uniformly distributed among the circumference of a circle covering 2π radians ranging from $-\pi$ to π, tooth number $t, t = 1, 2, \ldots 32$, is associated with a direction of $\pi(2t - 33)/32$ radians. Thus, a location of zero radian lies between teeth 16 and 17, designating maxillary teeth with negative radians and mandibular teeth with positive radians. We define a tooth to be diseased with periodontal disease if its corresponding six sites have an average CAL of 3 mm or more, as in Hoffman et al. (2001). Thus, the data to be analyzed for each subject are the locations (in radians) of the teeth

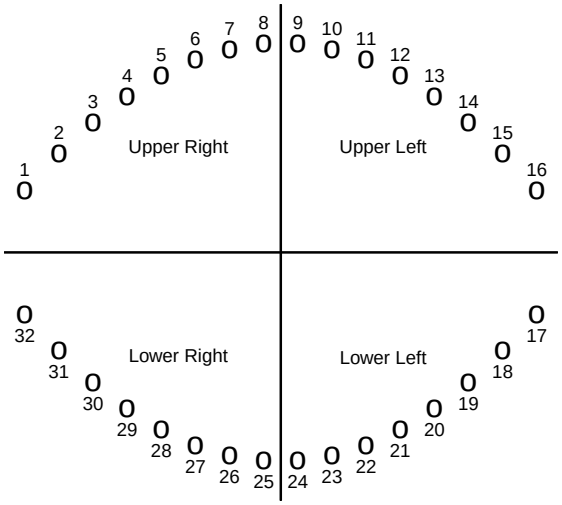

Figure 2.1: Diagram showing orientation of teeth in the mouth including numbering according to the Universal Numbering System.

affected with periodontal disease. Since we are primarily interested in identifying mean locations of diseased teeth, teeth not affected with periodontal disease (mean CAL less than 3 mm) are noninformative for our purpose and therefore will not be a part of the data to be analyzed.

We have m subjects in our study and subject $i, i = 1, 2, \ldots m$ has $1 \leq n_i \leq 28$ diseased teeth. We let Y_{ij} denote the location of diseased tooth $j, j = 1, 2, \ldots n_i$ in subject i, with a corresponding vector of K mouth-level covariates $\boldsymbol{X}_i = \{X_{i1}, X_{i2}, \ldots, X_{iK}\}$. Since the response variable Y_{ij} denotes an angle on a circle, we assume Y_{ij} to be a realization from a von Mises distribution. The von Mises distribution, also known as a Circular Normal distribution, is a continuous probability distribution on a circle (Mardia, 1972; Fisher, 1993). It is a symmetric unimodal distribution widely used in directional statistics to model circular data. The von Mises density for random

variable Z, with mean direction μ and concentration parameter κ, is given by

$$f(z) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(z - \mu)] \qquad -\pi \leq z, \mu < \pi,\ \kappa \geq 0 \tag{2.1}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order zero (Abramowitz and Stegun, 1965). The parameter μ denotes the mean direction while the parameter κ quantifies the concentration (variability) about the mean direction with smaller values of κ indicating greater variability.

For our purposes, we assume the response Y_{ij} has a marginal von Mises distribution with mean μ_i and concentration parameter κ. We denote the vector of locations for subject i as $\boldsymbol{Y}_i = \{Y_{i1}, Y_{i2}, \ldots, Y_{in_i}\}$, which has corresponding mean locations $\boldsymbol{\mu}_i = \mu_i \mathbf{1}_{\boldsymbol{n_i}}$. We note that each response Y_{ij} is a circular random variable while the corresponding vector of mouth-level or mouth-level covariates $\boldsymbol{X}_i$ are not circular, i.e. are continuous or categorical as with standard regression approaches. Hence, in order to study the relationship between Y_{ij} and $\boldsymbol{X}_i$, we use a so-called circular-linear regression model (Fisher and Lee, 1992). This model assumes a monotone link function that maps the explanatory variables to a circle. Though a variety of choices of the link function can be used, as discussed in Fisher and Lee (1992), we use the link function $h(x) = 2\arctan(x)$. Hence, the circular-linear regression equation for modeling the marginal response is

$$\mu_i = 2\arctan(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ...\beta_K X_{iK}) = 2\arctan(\eta_i) \tag{2.2}$$

where $\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ...\beta_K X_{iK}$.

We note that our methods can incorporate covariates measured at both mouth-level (equal for all teeth in the same mouth) and tooth-level (possibly different for each tooth in the same mouth). However, we focus solely on mouth-level covariates as our model (2.2) results in a single mean location per subject of where periodontal

disease exists in a mouth. Tooth-level predictors in our regression model would output a vector of mean locations for each subject, which though statistically correct, would not be easily interpretable in our setting. This challenge in interpretation is analogous to the use of time-varying covariates with longitudinal data. Our goal is to find mouth-level characteristics that may be associated with disease location. We note that some of these mouth-level covariates can be summary statistics of tooth-level measures, i.e. average pathogen level throughout the mouth or the mean number of locations that bleed when probed.

One crucial assumption of a von Mises distribution is that the location of periodontal disease is unimodal, an assumption that some may view as debatable. We first note that ours is the first attempt at modeling the distribution of periodontal disease locations, and no conclusive proof exists regarding whether unimodality is a valid assumption, as any evidence of multi-modality of periodontal disease locations is based mostly on anecdotal evidence from relatively small samples of data, with no formal attempts to determine whether more than one mode can be detected relative to the amount of variability inherent in the data. Mombelli and Meier (2001) and Darby et al. (2012) studied the distribution of periodontal disease and have noted a symmetry among the left and right sides of the mouth with respect to the distribution of periodontal disease in patients suffering from severe periodontal disease. Similarly, Thomson et al. (2000) and Loe and Brown (1991) observed that teeth at the back of the mouth, and particular to one side of the mouth showed the greatest indications for periodontal disease. When affected teeth mostly occur on one side and at the back of the mouth, our methods should give a valid estimate of the mean direction. We further note that for bimodal data where there similarity between opposing teeth in both jaws, our algorithm would result in a mean direction that is the average direction between the two modes, but closer to the the part of the mouth having the maximum number of affected teeth (see simulation results in Section 2.3).

2.2.2 Computing Mean Parameter Estimates and Their Standard Error Estimates

Our goal is to estimate the regression coefficients $\boldsymbol{\beta}=(\beta_0,\beta_1,...,\beta_K)^T$ in equation (2.2) and derive consistent variance estimates of the parameter estimates. Our data is comprised of multiple locations of diseased teeth per person, the number of which can vary from person to person, that are likely to be correlated due to the presence of some unobservable (latent) subject characteristics. Although this correlation could be incorporated through the inclusion of random effects in equation (2.2), such an approach changes the interpretation of the coefficients to be conditional on the values of the random effects. We instead focus on a population-averaged (marginal) approach for determining where periodontitis in the mouth. Given that we intend to model the marginal effects of the covariates, we have chosen to account for the correlation of locations from the same subject using generalized estimating equations (GEE) (Liang and Zeger, 1986). The advantage of GEE is that it does not require the correlation structure to be correctly specified in order to produce consistent estimates of the regression parameters, assuming that mean structure of the locations has been correctly specified. Moreover, GEE can produce consistent variance estimates for the regression parameter estimates even if the correlation structure is misspecified, although incorporating a correctly-specified correlation structure leads to improved efficiency (Lipsitz et al., 1994).

This last fact is important in our setting, because it is difficult to quantify the correct correlation structure with periodontal outcomes. The within-subject correlation of locations is not only related to the proximity of teeth to each other, but is also related to the biological functioning of the teeth, i.e. two molars, although in different regions of the mouth, are expected to be more correlated than a molar and incisor that are in different regions of the mouth. Hence, the assumption of any standard correlation structure, e.g. exchangeable or autoregressive, seems unreasonable

in our setting, while an unstructured correlation matrix will require the estimation of far too many parameters. Thus, our methods will simply assume independence of locations from the same subject when estimating the regression parameters, i.e. use an independence "working" correlation matrix, and then use the resulting residuals to produce "robust", or so-called "sandwich", variance estimates.

We note that Artes et al. (2000) and Song (2007) have both discussed the application of GEE to correlated circular outcomes. However, both of those approaches focused upon the large-sample properties of GEE and did not examine settings with a small or moderate number of subjects. Through our work, we have found that the robust variance estimates produced by GEE are quite biased in the sample sizes seen with most periodontal studies and lead to inflated Type I error rates. Thus, we propose the use of a bias-corrected robust variance estimator based on the methods of Mancl and DeRouen (2001). We now present details of GEE and the bias-correction methods that are pertinent to outcomes with marginal von Mises distributions.

The von Mises distribution, specified by equation (2.1), falls under the general family of dispersion models having unit deviance function $d(z;\mu) = 2[1 - \cos(z - \mu)]$ (Jorgensen, 1997). The deviance score and the unit variance function are related to the unit deviance function as $\delta(z;\mu) = -\frac{1}{2}\frac{\partial(d(z;\mu))}{\partial(\mu)} = \sin(z - \mu)$ and $V(\mu) = 2[\frac{\partial^2}{\partial\mu^2}d(z;\mu)|_{z=\mu}]^{-1} = 1$ respectively. For dispersion models, residuals are generally defined as the scaled deviance scores and are formulated as $r(z;\mu) = V(\mu)\delta(z;\mu)$. Since the von Mises distribution has a unit variance function $V(\mu) = 1$, the residuals are equal to the deviance scores.

We define the vector of score residuals for subject i as $\boldsymbol{r_i} = (r_{i1}, r_{i2}, ..., r_{in_i})^T$ where the residual corresponding observation j is defined as $r_{ij} = \sin[y_{ij} - 2\arctan(\eta_i)]$. It can be shown that $\text{Var}(r_{ij}) = \kappa^{-1}A_1(\kappa)$, where $A_1(\kappa) = I_1(\kappa)/I_0(\kappa)$ is the mean resultant length and $I_p(\kappa)$ is the modified Bessel function of the first kind of order p (Abramowitz and Stegun, 1965). As we assume that the observations, and hence

the residuals, from a subject are independent, we use the working covariance matrix $\boldsymbol{\Sigma}_i = A_1(\kappa)\kappa^{-1}\boldsymbol{I}_{n_i}$, where $\boldsymbol{I}_{n_i}$ is an $n_i \times n_i$ identity matrix. Based upon the basic methodology for GEE, we estimate $\boldsymbol{\beta}$ from the equation:

$$\Psi(\boldsymbol{\beta}) = \sum_{i=1}^{m} \boldsymbol{D}_i^T \boldsymbol{r}_i = 0 \tag{2.3}$$

where $\boldsymbol{D}_i^T = \boldsymbol{X}_i\boldsymbol{Z}_i$, $\boldsymbol{X}_i$ is a design matrix of order $p \times n_i$ and $\boldsymbol{Z}_i$=$(1+\eta_i^2)^{-1}\boldsymbol{I}_{n_i}$. The working covariance matrix $\boldsymbol{\Sigma}_i$ being a constant multiple of an identity matrix does not directly appear in the estimating equation (2.3).

The model based variance estimator of $\hat{\boldsymbol{\beta}}$, i.e. the variance estimator of $\hat{\boldsymbol{\beta}}$ assuming the correlation has been correctly specified is given by $Var_{MB}(\hat{\boldsymbol{\beta}}) = S(\hat{\boldsymbol{\beta}})^{-1}$ where $S(\boldsymbol{\beta}) = E[\Psi'(\boldsymbol{\beta})] = A_1(\kappa)\sum_i \boldsymbol{D}_i^T\boldsymbol{D}_i$. However, because $\boldsymbol{\Sigma}_i$ is misspecified, making $Var_{MB}(\hat{\boldsymbol{\beta}})$ inconsistent, an alternative is the robust, or so-called "sandwich" variance estimator $Var_R(\hat{\boldsymbol{\beta}}) = S(\hat{\boldsymbol{\beta}})^{-1}V(\hat{\boldsymbol{\beta}})S(\hat{\boldsymbol{\beta}})^{-1}$, where $V(\boldsymbol{\beta}) = E[\Psi(\boldsymbol{\beta})\Psi(\boldsymbol{\beta})^T] = \sum_i \boldsymbol{D}_i^T cov(\boldsymbol{r}_i)\boldsymbol{D}_i$. Usually $cov(\boldsymbol{r}_i)$ is estimated by $\boldsymbol{r}_i\boldsymbol{r}_i^T$ and thus $V(\boldsymbol{\beta}) = \sum_i \boldsymbol{D}_i^T\boldsymbol{r}_i\boldsymbol{r}_i^T\boldsymbol{D}_i$.

However, the robustness property of $Var_R(\hat{\boldsymbol{\beta}})$ to the misspecification of the working covariance matrix is a large-sample concept, and the working covariance matrix may still negatively influence $Var_R(\hat{\boldsymbol{\beta}})$ in finite samples, which is what we discovered in our simulations. As result, we propose the use of the bias-corrected robust variance estimator, computed using a first order Taylor series expansion of the residual vector $\boldsymbol{r_i}$, as in Mancl and DeRouen (2001). If we denote $V_1(\boldsymbol{\beta}) = \sum_i \boldsymbol{D}_i^T(\boldsymbol{I}_i - \boldsymbol{H}_{ii})^{-1}\boldsymbol{r}_i\boldsymbol{r}_i^T(\boldsymbol{I}_i - \boldsymbol{H}_{ii})^{-1}\boldsymbol{D}_i$, in which $\boldsymbol{H}_{ii} = \boldsymbol{D}_i\Big(\sum_i \boldsymbol{D}_i^T\boldsymbol{D}_i\Big)^{-1}\boldsymbol{D}_i^T$, the bias-corrected robust variance estimator of $\hat{\boldsymbol{\beta}}$ is given by $Var_{BC}(\hat{\boldsymbol{\beta}}) = S(\hat{\boldsymbol{\beta}})^{-1}V_1(\hat{\boldsymbol{\beta}})S(\hat{\boldsymbol{\beta}})^{-1}$. In comparison to the uncorrected variance estimator, the bias-corrected variance estimator uses an additional term $(\boldsymbol{I}_i - \boldsymbol{H}_{ii})^{-1}$ which serves to inflate the robust variance, making $Var_{BC}(\hat{\boldsymbol{\beta}})$ more likely to yield a consistent variance estimate than $Var_R(\hat{\boldsymbol{\beta}})$.

We note that in our setting, the methods of Mancl and DeRouen (2001) require

the computation of the first derivative of the deviance function with respect to the mean function. Due to the computational complexity of this first derivative, we chose to use the expected value of the corresponding derivative as a very accurate approximation that was much simpler to compute. Existing literature (Paik, 1998; Lipsitz et al., 1994) has suggested the use of Student's t or F-distribution as a reference distribution instead of the asymptotic normal or chi-square distribution applied to a Wald statistic. However, the choice of the degrees of freedom to be used is rather subjective. As suggested by Mancl and DeRouen (2001), we have chosen to use a t-distribution with degrees of freedom equal to the difference between the number of subjects and the number of parameters in the regression model, as it leads to a Wald test with approximately nominal size in our simulations.

In order to get starting parameter values for the GEE algorithm, we assume that the data are uncorrelated. We fit the von Mises distribution (2.1) to the data and obtain the maximum likelihood estimates (MLE) of μ and κ, as described in Fisher (1993) and Jammalamadaka and SenGupta (2001). We choose the MLE of κ as its starting value for our algorithm, and the MLE of μ is not used further. From equations (2.1) and (2.2), the value of $\boldsymbol{\beta}$ that maximizes the log-likelihood, assuming independent observations, is equal to the value that maximizes

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \cos\{\theta_{ij} - 2\arctan(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ...\beta_K X_{iK})\}$$

Over a fixed grid of values, we find the value of β_0 that maximizes $l(\boldsymbol{\beta})$ when we restrict $\beta_1, \beta_2, ..., \beta_K$ to be zero. We use this estimate of β_0 to find the value of β_1 that maximizes $l(\boldsymbol{\beta})$, over a fixed grid of values, when we restrict $\beta_2, \beta_3, ..., \beta_K$ to be zero. We carry on this procedure until we get initial estimates for all the elements of $\boldsymbol{\beta}$.

Once the starting values are computed, estimates of $\boldsymbol{\beta}$ and κ are computed re-

cursively. Since the estimating equation (2.3) cannot be solved analytically, we use a standard Newton Scoring algorithm for estimating $\boldsymbol{\beta}$. We obtain the updated $\hat{\boldsymbol{\beta}}_{\boldsymbol{new}}$ from the previous estimate $\hat{\boldsymbol{\beta}}_{\boldsymbol{old}}$ using

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{new}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{old}} - S(\hat{\boldsymbol{\beta}}_{\boldsymbol{old}})^{-1}\Psi(\hat{\boldsymbol{\beta}}_{\boldsymbol{old}})$$

The updated estimate, $\hat{\kappa}$, is computed using the inverse of the relation

$$\mathrm{A}_1(\hat{\kappa}) = \left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} cos(\theta_{ij} - 2\arctan(\hat{\eta}_i))\right)/(\sum_{i=1}^{m} n_i)$$

2.3 Numerical Examples

2.3.1 Simulation of Data

We now examine the performance of our methods using simulations of periodontal data motivated by a study described in Ramseier et al. (2009). We are interested in determining the ability of our methods to identify mouth-level covariates that are significantly associated with the location of periodontal disease. Our simulations focus upon two mouth-level covariates: (i) the proportion of sites in the mouth that experience bleeding on probing (BOP), and (ii) the percentage of total plaque bacterial pathogen load (PL) belonging to the red complex pathogen group. We assume that the number of diseased teeth in a person follows a Poisson distribution with mean $\lambda = 5.81$, which is the mean number of diseased teeth i.e. teeth having average CAL greater than 3mm, at baseline for the data described in Ramseier et al. (2009). The 99th percentile of this Poisson distribution is 12, which is also a realistic upper bound on the number of diseased teeth expected to occur in a person.

For a given subject $i, i = 1, 2, \ldots m$, there are n_i diseased teeth. We denote B_i as the proportion of sites experiencing BOP for subject i and P_i as the mouth level PL for subject i. Without loss of generality, we have assumed that if subject i has n_i

diseased teeth, those diseased teeth occur at teeth numbered $1, 2, \ldots n_i$, which means that in our simulations periodontal disease is occurring more frequently on one side of the upper jaw. We simulate B_i and P_i using the linear mixed effects models: $\text{logit}(B_i) = \mu_B + \gamma_1 n_i + u_i + e_i^B$ and $P_i = \mu_P + \gamma_2 n_i + u_i + e_i^P$ where μ_B is the mean log odds of proportion of sites experiencing BOP and μ_P is the mean PL in subjects having at least one diseased tooth. We incorporate a random subject effect u_i that has a normal distribution with mean 0 and variance τ^2, denoted $u_i \sim \mathcal{N}(0, \tau^2)$, to create correlation among measures from the same subject. The error terms $e_i^B \sim \mathcal{N}(0, \sigma_B^2)$, and $e_i^P \sim \mathcal{N}(0, \sigma_P^2)$ are mutually independent of each other and the random subject effect u_i.

In our simulation algorithm the parameter τ^2 denotes the random subject effect variance. The strength of the association of the two mouth-level covariates with the location of periodontal disease is quantified by the parameters γ_1 and γ_2. Thus when $\tau = \gamma_1 = \gamma_2 = 0$ there is no dependence between the covariates and the affected teeth locations, and the observations from a subject are independent. In this setting, mouth level BOP and PL values are distributed randomly throughout the mouth and neither is associated with where periodontal disease is occurring. We refer to this one as our "null setting". Non-zero values of γ_1 and γ_2 indicate that BOP and PL have an association with the location of periodontal disease. Furthermore, the dependence of the mouth level covariates on the location of periodontal disease is incorporated using the number of diseased teeth as we assume the diseased teeth to be occurring at consecutive tooth numbers starting with tooth numbered one. The values of the parameters presented above were chosen such that the distribution of the simulated data were similar to the baseline data in Ramseier et al. (2009), resulting in $\mu_B = 0.04$, $\mu_P = 3.6$, $\tau = 0.2$, $\sigma_B = 0.4$, $\sigma_P = 0.6$.

Although our approach for simulating data does not match the model we chose to analyze the data, we chose to simulate the data with consecutive teeth as diseased

for several reasons. First, our approach was a simple way to generate unimodal periodontal data. Second, as we know that affected teeth occur on only one side of the mouth, we can get an idea about average direction of affected teeth in our simulated datasets. Hence we will be able to calculate the mean direction of affected teeth in the simulated datasets and compare it to the estimated direction at the mean value of the covariate(s) using our fitted model. Thus, our approach provides a way to assess the performance of our estimation method and check if we are able to predict the mean location well when we correctly assume the unimodality of periodontal disease. Third, we are able to examine the performance of our methods in settings where the data is generated from a model different from that used to analyze the data.

2.3.2 Simulation Results

We first focus on modeling the marginal association between the proportion of sites with BOP and location of periodontal disease. We simulated 2,000 sets of data in four different settings and three different sample sizes of $m \in \{30, 50, 100\}$. The first setting is our "null" setting in which both $\gamma_1 = 0$ and $\tau^2 = 0$, so that BOP occurs randomly throughout the mouth. The remaining three settings examine increasing magnitudes of association between BOP and the location of periodontal disease. In each simulated set of data, we fit the regression model $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i)$ and record $\hat{\beta}_1$, the estimate of β_1, as well as the bias-uncorrected and bias-corrected standard error estimates of $\hat{\beta}_1$. Among the 2,000 simulation results, we computed the empirical mean and standard deviation of the estimates of β_1. For each simulation, we also performed a Wald test for $\beta_1 = 0$ using the bias-corrected standard error estimate and recorded whether or not the test indicated rejection based upon a t-distribution with $(m - 2)$ degrees of freedom. For the "null" setting, we expect the Wald test to have nominal size of 0.05, indicated by the observed proportion of simulations with

rejection lying in the interval $(0.040, 0.060)$. The results are shown in Table 2.1.

Table 2.1 Table showing the mean of estimated parameters and corresponding test sizes obtained on fitting the model: $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i)$ for 2,000 simulated datasets. The test sizes were calculated using bias-corrected standard error estimates assuming that the resulting Wald statistic followed a t-distribution with $(m-2)$ degrees of freedom; E=empirical; UC=uncorrected; BC=bias-corrected.

	Sample		SE of $\hat{\beta}_1$			
Setting	Size (m)	Mean $\hat{\beta}_1$	E	UC	BC	TestSize
$\gamma_1 = 0.00; \tau = 0.00$	30	0.05	2.51	2.09	2.55	0.05
	50	0.02	1.85	1.68	1.90	0.05
	100	-0.01	1.30	1.22	1.30	0.05
$\gamma_1 = 0.03; \tau = 0.2$	30	1.70	2.26	1.87	2.30	0.11
	50	1.68	1.69	1.49	1.69	0.17
	100	1.64	1.16	1.09	1.17	0.27
$\gamma_1 = 0.05; \tau = 0.2$	30	2.74	2.23	1.84	2.28	0.21
	50	2.72	1.67	1.47	1.68	0.36
	100	2.69	1.17	1.08	1.16	0.63
$\gamma_1 = 0.07; \tau = 0.2$	30	3.71	2.20	1.80	2.27	0.36
	50	3.69	1.65	1.43	1.66	0.60
	100	3.67	1.16	1.05	1.14	0.90

For the null model, we see that the bias-corrected standard error estimate is close to the empirical standard error of $\hat{\beta}_1$. However, the uncorrected standard error estimate is negatively biased, with the amount of underestimation being greater for smaller sample sizes. Thus, inference should be based on the bias-corrected variance estimator instead of the robust variance estimator. Histograms and quantile-quantile plots of the Wald statistics, calculated using the bias-corrected standard error, showed that the distribution was symmetric but had heavier tails than expected for a standard normal distribution. The empirical variance and kurtosis of the Wald statistics were close to that of a t-distribution with $(m-2)$ degrees of freedom, thereby supporting the proposal of Mancl and DeRouen (2001).

In the three settings where an association exists between proportion of sites experiencing BOP and location of periodontal disease, we continue to see that the bias-corrected standard error estimate is close to the empirical standard error of $\hat{\beta}_1$,

with the uncorrected standard error continuing to underestimate the true standard error. We note that γ_1 values of 0.03, 0.05 and 0.07 correspond to average Spearman rank correlations of 0.07, 0.11 and 0.15 between the proportion of sites experiencing BOP and the direction of diseased teeth, respectively. Thus, our proposed algorithm has good power for moderately sample sizes even when the correlation between the proportion of sites experiencing BOP and the location of periodontal disease is modest.

In order to study the marginal association between bacterial pathogen load (PL) in the mouth and location of periodontal disease, we fit the regression model $\mu_i = 2\arctan(\beta_0 + \beta_2 P_i)$. We simulated data using the same approach that we used for assessing the marginal association of mouth-level BOP with location of periodontal disease. The results are presented in Table 2.2 where γ_2 values of 0.05, 0.07, 0.09 correspond to average Spearman rank correlations of 0.08, 0.11, 0.14 between PL and direction of diseased teeth respectively. As with mouth-level BOP, the results in Table 2.2 show the improvement of the bias-corrected standard error in comparison to the bias-uncorrected standard error in all the four scenarios.

Furthermore, in order to study the association between proportion of sites experiencing BOP and location of periodontal disease conditional on the value of mouth-level PL we consider the model $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i + \beta_2 P_i)$. Inference using simulation results, presented in Table 2.3, are analogous to those presented in Tables 2.1 and 2.2. However, the test sizes are now calculated in reference to a t-distribution with $(m-3)$ degrees of freedom. Here, the three settings which examine increasing association between the covariates and the location of diseased teeth correspond to average Spearman rank correlations of 0.16, 0.19, 0.23 between the covariates BOP and PL respectively.

We also note that we found that across all simulation settings and both covariates, our fitted regression models tended to predict the mean direction of diseased teeth to

Table 2.2 Table showing the mean of estimated parameters and corresponding test sizes obtained on fitting the model: $\mu_i = 2\arctan(\beta_0 + \beta_2 P_i)$ for 2,000 simulated datasets. The test sizes were calculated using bias-corrected standard error estimates assuming the resulting Wald statistic followed a t-distribution with $(m-2)$ degrees of freedom; E=empirical; UC=uncorrected; BC=bias-corrected.

Setting	Sample Size (m)	Mean $\hat{\beta}_2$	SE of $\hat{\beta}_2$			Test Size
			E	UC	BC	
$\gamma_2 = 0.00; \tau = 0.00$	30	0.01	0.40	0.34	0.42	0.05
	50	0.00	0.30	0.27	0.31	0.05
	100	-0.01	0.21	0.20	0.21	0.05
$\gamma_2 = 0.05; \tau = 0.2$	30	0.31	0.36	0.30	0.38	0.12
	50	0.31	0.27	0.24	0.28	0.19
	100	0.32	0.19	0.18	0.19	0.37
$\gamma_2 = 0.07; \tau = 0.2$	30	0.42	0.35	0.29	0.37	0.19
	50	0.42	0.26	0.23	0.27	0.33
	100	0.43	0.19	0.17	0.19	0.61
$\gamma_2 = 0.09; \tau = 0.2$	30	0.51	0.34	0.28	0.36	0.29
	50	0.52	0.25	0.22	0.26	0.48
	100	0.52	0.18	0.16	0.18	0.82

be -2.50 radians which corresponds to tooth number four. This is to be expected, as we had simulated the data assuming that the number of diseased teeth in a subject follows a Poisson distribution with mean 5.81, with diseased teeth numbered in an ascending order starting from tooth number one. Since the circular mean of the directions corresponding to tooth numbers 1 to 6 is -2.55 radians, our regression models give a good prediction of the mean direction of diseased teeth.

We also fitted our methods to bimodal data such that when a tooth is affected, the tooth on the contralateral side, i.e. on the opposite jaw, is also affected. We simulated 2,000 sets of data of 30 subjects each, similar to the data summarized in Tables 1-3, such that the affected teeth occur around modes at teeth 15 and 18, i.e. on the left side of the mouth, and there are more affected teeth at and adjoining tooth numbered 18 along the lower left jaw. Our algorithm resulted in a mean direction corresponding to edge between teeth 17 and 18 (results not shown). This suggests that with bimodal data, our algorithm will produce a mean direction that is close to the mean of the

Table 2.3 Table showing the mean of estimated parameters and corresponding test sizes obtained on fitting the model: $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i + \beta_2 P_i)$ for 2,000 simulated datasets. The test sizes were calculated using bias-corrected standard error estimates assuming the resulting Wald statistic to follow a t-distribution with $(m-3)$ degrees of freedom; E=empirical; UC=uncorrected; BC=bias-corrected.

	Sample		SE of $\hat{\beta}_1$			
Setting	Size (m)	Mean $\hat{\beta}_1$	E	UC	BC	Test Size
$\gamma_1 = 0.00, \gamma_2 = 0.00; \tau = 0.00$	30	0.06	2.62	2.09	2.74	0.05
	50	0.02	1.92	1.67	1.97	0.04
	100	-0.02	1.31	1.22	1.32	0.05
$\gamma_1 = 0.03, \gamma_2 = 0.05; \tau = 0.2$	30	1.42	2.35	1.86	2.47	0.07
	50	1.39	1.73	1.48	1.76	0.11
	100	1.34	1.18	1.09	1.19	0.19
$\gamma_1 = 0.05, \gamma_2 = 0.07; \tau = 0.2$	30	2.30	2.31	1.83	2.46	0.14
	50	2.28	1.72	1.45	1.74	0.26
	100	2.21	1.18	1.06	1.18	0.45
$\gamma_1 = 0.07, \gamma_2 = 0.09; \tau = 0.2$	30	3.08	2.28	1.79	2.43	0.23
	50	3.05	1.69	1.42	1.72	0.41
	100	2.98	1.16	1.04	1.16	0.73

	Sample		SE of $\hat{\beta}_2$			
Setting	Size (m)	Mean $\hat{\beta}_2$	E	UC	BC	Test Size
$\gamma_1 = 0.00, \gamma_2 = 0.00; \tau = 0.00$	30	0.01	0.42	0.34	0.45	0.04
	50	0.001	0.31	0.27	0.32	0.05
	100	-0.01	0.21	0.20	0.21	0.05
$\gamma_1 = 0.03, \gamma_2 = 0.05; \tau = 0.2$	30	0.27	0.38	0.30	0.41	0.10
	50	0.28	0.28	0.24	0.29	0.15
	100	0.28	0.19	0.18	0.20	0.29
$\gamma_1 = 0.05, \gamma_2 = 0.07; \tau = 0.2$	30	0.35	0.36	0.29	0.40	0.13
	50	0.35	0.27	0.23	0.28	0.23
	100	0.35	0.19	0.17	0.19	0.45
$\gamma_1 = 0.07, \gamma_2 = 0.09; \tau = 0.2$	30	0.40	0.35	0.28	0.38	0.18
	50	0.40	0.26	0.22	0.27	0.31
	100	0.40	0.18	0.16	0.18	0.60

two modes and skewed towards the mode with a greater propensity of the data. We confirmed these conclusions by simulating data in three additional settings with the modes located at two different teeth and found that our conclusions held with these settings as well (results not shown). However, we emphasize in settings like these

that a simple plot of the resulting residuals (as presented in the next section) would indicate that an assumption of a unimodal distribution of periodontal disease would be suspect.

2.3.3 Data Analysis

We applied our methods to the data motivating our simulations, which came from the clinical trial described in Ramseier et al. (2009) and Kinney et al. (2011). This non-randomized observational study, conducted at the Michigan Center for Oral Health Research, involved 50 periodontally healthy and 50 periodontally diseased subjects, based upon entry criteria specified in Ramseier et al. (2009). Investigators gave periodontal exams and collected saliva samples from each of the 100 subjects at baseline, as well as six and twelve months after enrollment. Among the data recorded for each subject, we focus upon the tooth-level mean CAL, proportion of sites in the mouth experiencing BOP, and the total percentage of the bacterial pathogens *P. gingivalis*, *T. denticola*, and *T. forsythia* present in the saliva.

We wish to use our methods to examine if the mean location of diseased teeth is associated with mouth level BOP and/or pathogen levels at baseline or six months. At baseline, there were 42 subjects with at least one diseased tooth (mean CAL > 3 mm) contributing a total of 244 teeth. At six months, there were 33 subjects with at least one diseased tooth contributing a total of 199 teeth. The decrease in sample size is due to the loss of a few subjects during the first six months of the study, as well as the Hawthorne effect (Braunholtz et al., 2001) common in observational studies, i.e. patients in the study naturally improved their oral health behaviors slightly once enrolled in the study. The maximum likelihood estimate of the mean location of periodontal disease at baseline is -0.33 radians, which corresponds to tooth number 15 in Figure 2.1. These results supported our cursory exam of the data, in which we saw periodontal disease occurring at molars and other teeth at the back of the mouth.

We then applied our algorithm to study both the marginal and joint associations of mouth-level BOP and PL on the mean location of periodontal disease. We fit the models $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i)$ and $\mu_i = 2\arctan(\beta_0 + \beta_2 P_i)$ in order to study the marginal association with proportion of BOP and percentage of red complex pathogen level separately. We also fit the model $\mu_i = 2\arctan(\beta_0 + \beta_1 B_i + \beta_2 P_i)$ which quantifies the joint dependence of the predictors with the diseased locations. As in earlier sections, we consider inference using the bias-corrected standard error estimate and assuming the Wald statistic follows a null t-distribution. The results are tabulated in Table 2.4. We find that neither proportion of sites with BOP nor percentage of red complex pathogens at baseline have an association with location of periodontal disease at the beginning of the study (baseline). However, we see a much stronger association of both factors at six months with location of periodontal disease at six months, with a slightly stronger association for percentage of red complex pathogens than proportion of sites with BOP whether modeled alone or jointly.

Table 2.4 Estimated parameters and bias-corrected standard error estimates obtained on fitting the models: $(i)\mu_i = 2\arctan(\beta_0 + \beta_1 B_i)$ $(ii)\mu_i = 2\arctan(\beta_0 + \beta_2 P_i)$ $(iii)\mu_i = 2\arctan(\beta_0 + \beta_1 B_i + \beta_2 P_i)$. ' $*\,*'$ denotes significance with p-value < 0.05 and '$*$' denotes significance with $0.05 \leq$p-value ≤ 0.10. The p-values are based on a t-distribution with $(m - p)$ degrees of freedom where 'm' denotes the number of subjects and 'p' the number of parameters in the regression model.

	Affected teeth locations at baseline ($m = 42$)		Affected teeth locations at month 6 ($m = 33$)	
Predictors	$\hat{\beta}_1(se)$	$\hat{\beta}_2(se)$	$\hat{\beta}_1(se)$	$\hat{\beta}_2(se)$
(i) BOP only	-0.33(1.54)	n/a	-1.31(1.25)	n/a
(ii) PL only	n/a	0.08(0.12)	n/a	-0.34(0.21)*
(iii) BOP and PL	0.06(1.55)	0.08(0.13)	-1.61(1.15)*	-0.39(0.19)**

We examined the standardized residuals from our model to assess our assumption of a marginal von Mises distribution using the approach of Song (2007). For $Y_{ij} \sim$ von Mises (μ_i, κ), the standardized residual for observation j from the subject i is defined as $\hat{e}_{ij} = [A_1(\hat{\kappa})/\hat{\kappa}]^{-1/2} \sin(Y_{ij} - \hat{\mu}_i)$. A plot of the standardized residuals $\hat{e}_{ij}$ against

the estimated mean direction $\hat{\mu}_i$ can be used to check for the validity of the marginal distributional assumption. Ideally we would expect all the points to be randomly scattered around zero and about 95% of the points to be between -2 and 2. For our baseline data, the residual plots obtained from modeling the predictors separately, as shown in Figure 2.2, and using the joint modeling of the two predictors show that there is no obvious violation of our assumption.

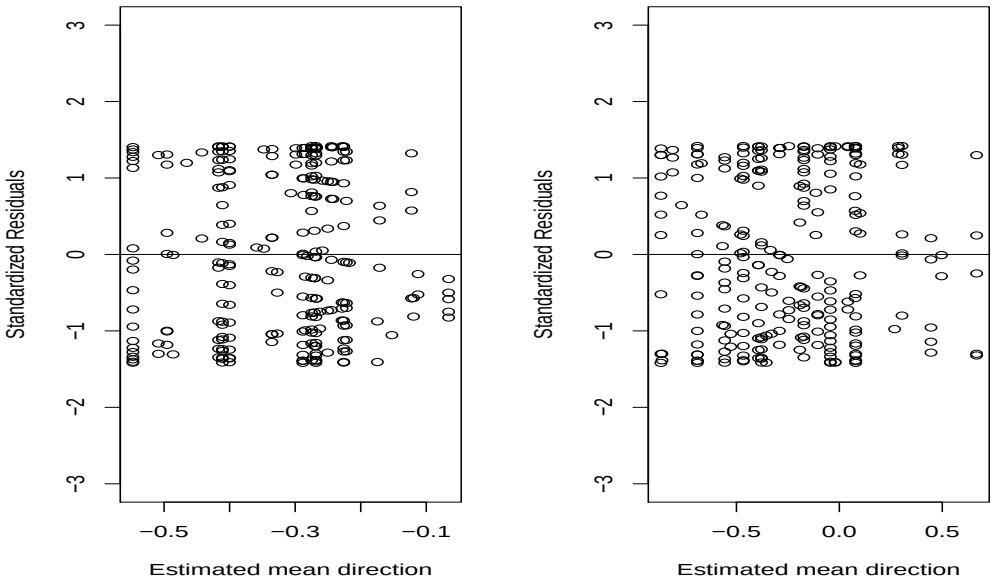

Figure 2.2: Standardized residual plots corresponding to the baseline data analysis in Section 2.3.3. Left plot displays the residuals calculated using average BOP and right plot displays the residuals calculated using average PL as predictors.

2.4 Discussion

Our methodology provides an improvement over the current state of art as there is no formal statistical approach designed for specifically determining the location of periodontal disease. Existing literature usually considers observations on *a priori* selected specific regions of the mouth and then compares the frequency of affected

teeth across these regions in the mouth without any formal modeling. Furthermore, most of the existing work fails to account for multiple measurements from a subject and considers summary information aggregated over the selected mouth regions. Nonetheless, there are issues with our methods that motivate further research. One could consider fitting mixtures of unimodal von Mises distributions or a generalized von Mises distribution (Gatto and Jammalamadaka, 2007) to allow for the analysis of multimodal periodontal data and methods for determining the appropriate number of modes. However, such generalization is not immediately straightforward, as Gatto (2008) states a number of theoretical challenges with using a mixture of von Mises distributions. A major issue is that the likelihood function of the mixture of von Mises(μ_1, κ_1) and von Mises(μ_2, κ_2) distributions is unbounded. Thus an overall supremum of the likelihood of a von Mises mixture does not give a consistent estimate although some other local supremum do so.

Although we chose to use the bias-corrected variance estimator of Mancl and DeRouen (2001), alternate approaches have been proposed by Fay and Graubard (2001) and Kauermann and Carroll (2001). It can be shown that the methods of Mancl and DeRouen (2001) and Fay and Graubard (2001) both lead to the same bias-corrected standard error estimator with von Mises marginal distributions. Via simulations, we also found that the degrees of freedom estimators of Fay and Graubard (2001) led to an underestimated size for the Wald test (results not shown). We also applied the methods of Kauermann and Carroll (2001) to the simulation settings presented in Section 2.3.1 and found slight improvement to the uncorrected robust standard errors, but not as much as the methods of Mancl and DeRouen (2001), which is expected given the findings of Lu et al. (2007). Thus, we recommend the use of the approach of Mancl and DeRouen (2001) in our methods.

In our motivating dataset, subjects had few missing teeth, and of the few teeth for which measures were missing, there was no discernible pattern in the missing-

ness at baseline nor at six months, and no tooth that was diseased at baseline was missing at six months. In reality, subjects with existing periodontal disease may previously had other teeth removed, some of which may have been due to chronic periodontal disease. Such missing data poses a challenge, as the missing teeth are not missing at random (MAR), which is a necessary assumption for the validity of the robust standard errors produced by marginal approaches such as GEE (Kenward and Molenberghs, 1998). Thus, the investigation of the performance of our methods for data with greater numbers of non-randomly missing teeth is certainly an important future area of research.

CHAPTER III

Analysis of Periodontal Data using Multimodal Circular Statistics

3.1 Introduction

In the previous chapter, we modeled the mean direction of affected teeth in the mouth as a function of mouth-level covariates using the symmetric, unimodal von Mises distribution. Recently various dental studies (Tomasi et al., 2007; Reich and Bandyopadhyay, 2010) have suggested that periodontal disease is most common particularly at the back of the mouth and thus it is possible that the distribution of affected teeth is multimodal. However, there have been no formal attempts to determine the exact distribution of affected teeth nor methods to verify if the distribution of diseased teeth is multimodal. We propose methods to address these questions.

We consider the same setting as that in the previous chapter. We assume the mouth as a circle and the teeth as points on the circumference of a circle as shown in Figure 3.1. We use the Universal Numbering System adopted by the American Dental Association, in which the 32 teeth in an adult are numbered sequentially from 1 to 32, with the numbers 1-16 referring to the teeth in the upper jaw (maxillary) and the numbers 17-32 referring to the sixteen teeth in the lower jaw (mandibular). As wisdom teeth (teeth 1, 16, 17 and 32) are often removed in most adults even when healthy, these teeth are usually omitted from periodontal studies, leading to a maximum of 28 teeth measured in each subject. These numbers allow us to designate a specific

direction for each tooth, in radians, corresponding to its location in the mouth. If we assume that all 32 teeth are uniformly distributed among the circumference of a circle covering 2π radians ranging from 0 to 2π, tooth number $t, t = 1, 2, \ldots 32$, is associated with a direction of $\pi(2t-1)/32$ radians. Hence, the corresponding angle (in radians) associated with a tooth are higher for higher tooth numbers. A location of zero radian lies between teeth 1 and 32, designating maxillary teeth with radians ranging from 0 to π and mandibular teeth with radians ranging from π to 2π.

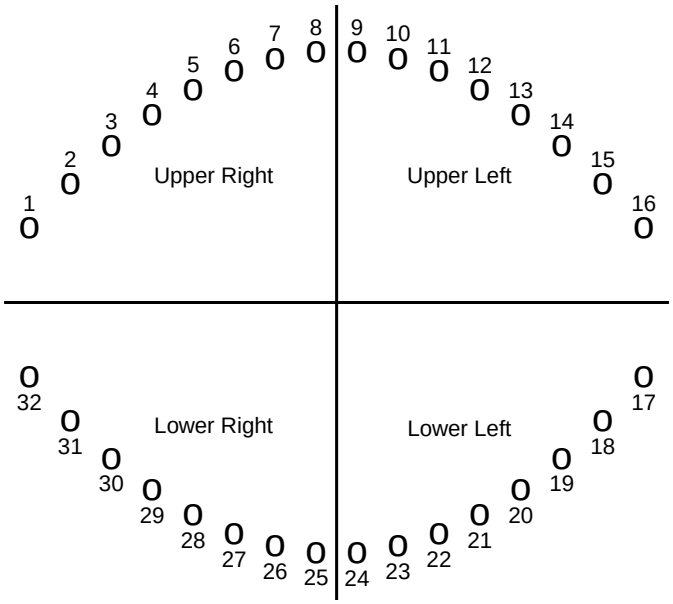

Figure 3.1: Diagram showing orientation of teeth in the mouth including numbering according to the Universal Numbering System

We classify the disease status of a tooth using its mean clinical attachment level (CAL) value. We consider a tooth to be diseased if its mean CAL value, averaged over six sites of the tooth, is 3 mm or more, as in Hoffman et al. (2001). Since we are solely interested in studying the distribution of diseased teeth, teeth not diseased with periodontal disease (mean CAL less than 3 mm) are noninformative for our purpose

and therefore will not be a part of the data to be analyzed.

3.2 Statistical Methods

Suppose we have m subjects in our study and subject $i, i = 1, 2, \ldots m$ has $1 \leq n_i \leq 28$ diseased teeth. Let Y_{ij} denote the location of diseased tooth $j, j = 1, 2, \ldots n_i$ in subject i. We assume that the responses Y_{ij} have the same marginal distribution and observations from the same subject are correlated while observations from different subjects are uncorrelated. Since the response variable Y_{ij} denotes an angle on a circle, we assume Y_{ij} to be a realization from the Generalized von Mises distribution of order k, denoted as GvM_k, where the order k reflects the maximum number of modes possible for the distribution.

The Generalized von Mises distribution (GvM) is a probability distribution on a circle flexible enough to accommodate multiple modes and asymmetry. It is a generalization of the von Mises distribution (Mardia, 1972; Fisher, 1993), a symmetric unimodal distribution widely used in directional statistics to model circular data. The Generalized von Mises distribution originates from Maksimov (1967) and a bivariate analogue of the distribution was proposed by Yfantis and Borgman (1982). Recently, Gatto and Jammalamadaka (2007) and Gatto (2008) have studied the distribution in details, stating theoretical properties, simulation algorithms and applications to an actual dataset.

The density for a random variable Z following a Generalized von Mises distribution of order k, denoted as $Z \sim GvM_k(\mu_1, ..., \mu_k, \kappa_1, ..., \kappa_k)$ is

$$f(z) = \frac{1}{2\pi G_0^{(k)}(\delta_1, ..., \delta_{k-1}, \kappa_1, ..., \kappa_k)} \exp\big[\sum_{j=1}^{k} \kappa_j \cos j(z - \mu_j)\big] \tag{3.1}$$

where $\kappa_1, .., \kappa_k > 0$, $\mu_\ell \in [0, 2\pi/\ell)$, $\ell = 1, 2, ...k$ and $G_0^{(k)}(\delta_1, ..., \delta_{k-1}, \kappa_1, ..., \kappa_k) =$

$\frac{1}{2\pi}\int_0^{2\pi}\exp\left[\kappa_1\cos(z)+\sum_{\ell=2}^{k}\kappa_\ell\cos\ell(z+\delta_{\ell-1})\right]$. The shape, skewness and modality of the distribution depends on the values of the parameters and the GvM_k distribution can have a maximum of k modes. The location parameters $(\mu_1,\mu_2,...,\mu_k)$ determine the location of the modes while the concentration parameters $(\kappa_1,\kappa_2,...,\kappa_k)$ determine the peakedness of the distribution around the modes. From equation (3.1), we note that $k=1$ corresponds to a von Mises distribution.

An important feature of the GvM_k distribution is that it can be reparameterized, leading to a canonical exponential family form. If we define $\lambda_{2\ell-1}=\kappa_\ell\cos(\ell\mu_\ell)$ and $\lambda_{2\ell}=\kappa_\ell\sin(\ell\mu_\ell), \ell=1,2,..k$, the density of a GvM_k distribution can be expressed as

$$f(z)=\exp[\boldsymbol{\lambda}^T\boldsymbol{T(z)}-K(\boldsymbol{\lambda})] \tag{3.2}$$

in which

$$\begin{aligned}\boldsymbol{\lambda} &= (\lambda_1,\lambda_2,...,\lambda_{2k})^T\\ \boldsymbol{T(z)} &= (\cos z,\sin z,\cos 2z,\sin 2z,...,\cos kz,\sin kz)^T\\ K(\boldsymbol{\lambda}) &= \log(2\pi)+\log G_0^{(k)}(\delta_1,...,\delta_{k-1},\|\boldsymbol{\lambda}^{(1)}\|,...,\|\boldsymbol{\lambda}^{(k)}\|),\end{aligned}$$

where $\|\boldsymbol{\lambda}^{(\ell)}\|$ denotes the Euclidean norm of $\boldsymbol{\lambda}^{(\ell)}=(\lambda_{2\ell-1},\lambda_{2\ell})^T$ and $\delta_{\ell-1}=(\arg\boldsymbol{\lambda}^{(1)}-\arg\boldsymbol{\lambda}^{(\ell)}/\ell) \mod (2\pi/\ell)$. The reparameterization of the GvM_k density corresponds to a $2k$ parameter exponential family. Thus, in this setting $\boldsymbol{T(z)}$ is a complete sufficient statistic for $\boldsymbol{\lambda}$.

As computing the normalizing constant $G_0^{(k)}(\delta_1,...,\delta_{k-1},\kappa_1,...,\kappa_k)$ for higher values of k is tedious, the aforementioned manuscripts have considered inference based on a Generalized von Mises distribution of second order. Gatto (2008) derived the score function and calculated the covariance matrix of the parameters of a GvM_2 distribution. Furthermore, they proposed Akaike's Information Criterion (AIC) and

other model selection criteria in order to determine the appropriate order of GvM distribution which gives the best fit to the data. Their inference was based on independent observations.

A complication arises in our setting as observations from a subject are correlated due to various subject-level characteristics some of which may be latent or unobservable. Thus in our situation, the challenges are twofold. We need to calculate valid parameter and standard error estimates to a dataset of correlated angular responses. We also need to compute a model selection criterion in order to determine the appropriate order of the Generalized von Mises distribution to be fitted. From the fitted GvM distribution we can calculate the modes of the fitted distribution and thus determine the number and locations of teeth which are most likely to have periodontal disease.

3.2.1 Parameter estimation

White (1982) and Boos (1992) derived the "robust" variance estimator for an unbiased estimating function based on independent observations having the same marginal distribution. Liang and Zeger (1986) proposed inference based on generalized estimating equations (GEE) for correlated observations having the same marginal distribution. The advantage of GEE is that it does not require the correlation structure to be correctly specified in order to produce consistent parameter estimates, assuming that mean structure of the locations have been correctly specified. Moreover, GEE can produce consistent variance estimates for the parameter estimates even if the correlation structure is misspecified, although incorporating a correctly-specified correlation structure leads to improved efficiency (Lipsitz et al., 1994). The variance estimators for the parameter estimates are called "robust" or "sandwich" variance estimators as they are robust to misspecification of the correlation structure. Since it is difficult to assess the correct correlation structure to be used for periodontal outcomes, for

simplicity, we assume an independence working correlation structure i.e. we assume all the observations to be uncorrelated. We derive the corresponding likelihood, referred to henceforth as quasi-likelihood, and the quasi-score function for independent GvM_k observations. Based upon the "robust" variance estimator proposed by Liang and Zeger (1986), we obtain a valid covariance matrix of our parameters estimates.

Assume that the angular direction of affected tooth j in subject i follows a marginal GvM_k distribution i.e. $Y_{ij} \sim GvM_k(\lambda_1, \lambda_2, ..., \lambda_{2k})$; $j = 1, 2, ..., n_i$ and $i = 1, 2, ..., m$. The density of Y_{ij} can be expressed as $f(Y_{ij}; \boldsymbol{\lambda})$ as in equation (3.2) where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_{2k})$ is the vector of parameters.

$$f(Y_{ij}) = \exp[\boldsymbol{\lambda}^T \boldsymbol{T}(Y_{ij}) - K(\boldsymbol{\lambda})]$$

We are interested in calculating valid estimates and covariance matrix of $\boldsymbol{\lambda}$. Assuming the observations to be independent, the quasi-likelihood function is given by $L(\boldsymbol{\lambda}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} f(Y_{ij}; \boldsymbol{\lambda})$ and the log-quasilikelihood function is $l(\boldsymbol{\lambda}) = \log L(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log f(Y_{ij}; \boldsymbol{\lambda})$. Hence the corresponding score function is $S(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} s(Y_{ij}; \boldsymbol{\lambda})$, where $s(Y_{ij}; \boldsymbol{\lambda}) = \frac{\partial}{\partial \boldsymbol{\lambda}} \log f(Y_{ij}; \boldsymbol{\lambda}) = \frac{\partial}{\partial \boldsymbol{\lambda}} [\boldsymbol{\lambda}^T \boldsymbol{T}(Y_{ij}) - K(\boldsymbol{\lambda})]$. Thus the unbiased estimating equation is given by

$$S(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} s(Y_{ij}; \boldsymbol{\lambda}) = \sum_{i=1}^{m} S_i(\boldsymbol{\lambda}) = \mathbf{0}$$

where $S_i(\boldsymbol{\lambda}) = \sum_{j=1}^{n_i} s(Y_{ij}; \boldsymbol{\lambda})$.

The model based covariance matrix of $\boldsymbol{\lambda}$ is obtained as the inverse of the observed Fisher information matrix and is thus expressed as $Var_{MB}(\hat{\boldsymbol{\lambda}}) = \big(\sum_{i=1}^{m} A_i(\boldsymbol{\lambda})\big)^{-1}$, where $A_i(\boldsymbol{\lambda}) = -\frac{\partial}{\partial \boldsymbol{\lambda}} S_i(\boldsymbol{\lambda})$ is the negative of the hessian matrix for observations from the i th subject. Note that

$A_i(\boldsymbol{\lambda}) = -\frac{\partial}{\partial \boldsymbol{\lambda}} S_i(\boldsymbol{\lambda}) = \sum_{j=1}^{n_i} (-\frac{\partial}{\partial \boldsymbol{\lambda}} s(Y_{ij}; \boldsymbol{\lambda})) = \sum_{j=1}^{n_i} \big(\frac{\partial}{\partial \boldsymbol{\lambda} \boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big) = n_i \big(\frac{\partial}{\partial \boldsymbol{\lambda} \boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big)$.

Hence the Fisher information matrix for the quasi-likelihood function is given by $\sum_{i=1}^{m} A_i(\boldsymbol{\lambda}) = \sum_{i=1}^{m} n_i\big(\frac{\partial}{\partial \boldsymbol{\lambda}\boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big) = n\big(\frac{\partial}{\partial \boldsymbol{\lambda}\boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big)$ where $n = \sum_{i=1}^{m} n_i$ is the total number of observations from all subjects in the dataset. Hence, the model based covariance matrix of $\boldsymbol{\lambda}$ is $Var_{MB}(\hat{\boldsymbol{\lambda}}) = \big(n\frac{\partial}{\partial \boldsymbol{\lambda}\boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big)^{-1}$.

However, the model based estimator $Var_{MB}(\hat{\boldsymbol{\lambda}})$ fails to give a valid standard error estimate for correlated data. Hence we come up with a "robust" variance estimator along the lines of Liang and Zeger (1986) denoted as

$$Var_R(\hat{\boldsymbol{\lambda}}) = Var_{MB}(\hat{\boldsymbol{\lambda}})\big(\sum_{i=1}^{m} S_i(\boldsymbol{\lambda}) S_i(\boldsymbol{\lambda})^T\big) Var_{MB}(\hat{\boldsymbol{\lambda}})$$

For uncorrelated data, the variance estimator $Var_R(\hat{\boldsymbol{\lambda}})$ is equivalent to the variance estimator stated in White (1982) and Boos (1992) which is robust to model misspecification.

In order to determine the parameter estimate $\hat{\boldsymbol{\lambda}}$ we follow the following procedure. Ideally the gradient of the log-likelihood function evaluated at $\hat{\boldsymbol{\lambda}}$ should be close to zero i.e. $S(\hat{\boldsymbol{\lambda}}) \approx 0$ and it should yield a valid covariance matrix i.e. $\det(Var_{MB}(\hat{\boldsymbol{\lambda}})) > 0$. Usually for the fitted GvM_2 distributions studied by Gatto and Jammalamadaka (2007) and Gatto (2008) the parameter estimates were obtained using the "fminsearch" function in Matlab which yields the unconstrained minimum of the negative log-likelihood function. We also adopted a similar approach in order to obtain parameter estimates. However, a caveat of the approach is that the algorithm is a direct search algorithm, i.e. an algorithm which does not use the derivative of the likelihood function. Hence we calculated the gradient of the log-likelihood function and the determinant of the information matrix at the estimated value of the parameters and chose those parameter estimates to be valid for which the gradient of the log-likelihood function were close to zero i.e. $S(\hat{\boldsymbol{\lambda}}) \approx 0$ and the determinant of the information matrix was positive i.e. $\det\big(\frac{\partial}{\partial \boldsymbol{\lambda}\boldsymbol{\lambda}^T} K(\boldsymbol{\lambda})\big) > 0$ (so as to yield a

valid covariance matrix). Furthermore, we used Newton Scoring algorithm to update the parameter estimates obtained using the "fminsearch" algorithm. However, the results obtained using both the methods were very similar.

3.2.2 Model Selection Criterion

Akaike (1973) proposed a likelihood-based quantity for model selection, commonly referred to as Akaike's Information Criterion (AIC). Gatto (2008) had proposed the use of AIC in order to determine the appropriate order of the Generalized von Mises distribution to be fit to uncorrelated angular responses. However in our setting, we do not formally have the joint likelihood of the data due to their correlation, making AIC unsuitable as a measure of model fit. Pan (2001) proposed a modification of AIC, termed as QIC, which is appropriate for quasi-likelihood based inference as in GEE. Due to our use of an independence working correlation matrix, we use the criterion

$$QIC = -2l(\hat{\boldsymbol{\lambda}}) + 2\, trace(Var_{MB}(\hat{\boldsymbol{\lambda}})^{-1} Var_R(\hat{\boldsymbol{\lambda}}))$$

where $l(\hat{\boldsymbol{\lambda}}), Var_{MB}(\hat{\boldsymbol{\lambda}}), Var_R(\hat{\boldsymbol{\lambda}})$ are defined as in Section 3.2.1. For different values of k, we fit the corresponding GvM_k and we will choose the GvM_k distribution that has the smallest value of QIC.

3.2.3 Calculation of Modes

Although we will have produced parameter estimates and their corresponding variance estimates for the best-fitting GvM_k distribution, interpretation of these parameters is difficult. A more interesting quantity is what the corresponding value of the density for each angle will be. Thus, after we have determined the appropriate order of the distribution to be fitted and obtained the parameter estimates as described in Sections 3.2.1 and 3.2.2, we can obtain density estimates at each angle.

In our setting, this translates to the density estimates for each tooth and leads to an indication of the relative frequency of periodontal disease at each tooth, with the modes occurring at teeth with the largest density estimates.

3.3 Simulation Examples

3.3.1 Simulation Algorithm

We consider various simulation settings to examine the performance of our methods. We simulate correlated angular observations having marginal GvM_k distribution for different values of the order k. In one particular simulation scenario, we then fit varying orders of Generalized von Mises distribution to the correlated angles and calculate QIC values for each of the fitted distributions. We then choose the best fitting distribution as the one having minimum QIC value. Ideally the fitted distribution having the minimum QIC value should match with the marginal distribution of the simulated angles and thus validate our methods. In our simulation examples, we consider simulation settings where the marginal distribution of the angles follow von Mises (GvM_1), GvM_2, GvM_3 or GvM_4 distributions. We restrict our simulations to Generalized von Mises distributions of fourth order as a mouth has four quadrants which are mirror images of each other and thus it is natural to assume that the distribution of affected teeth is unlikely to have more than four modes. Furthermore, it should be noted that a GvM_k distribution has $2k$ parameters and $2k(2k+1)/2$ distinct elements in the corresponding variance covariance matrix. Hence, given a subject is unlikely to have more than 10 or 12 affected teeth, we might run into the problem of overfitting if we intend to fit a Generalized von Mises distribution of higher order.

Gatto (2008) mentions detailed steps for simulation of independent marginal GvM_2 observations using the ratio-of-uniforms algorithm and also discusses the simulation algorithm for higher order. We extend the approach to simulate correlated

GvM_k; $k = 1, 2, 3, 4$ observations. We simulate observations from a marginal GvM_k distribution and induce correlation among observations from the same subject using a random effects approach. Lets consider a study of m subjects with subject i; $i = 1, 2, ..., m$ having n_i observations. Let Y_{ij} denote the angular response for observation j from the subject i; $i = 1, 2, ..m; j = 1, 2, ...n_i$. Hence for our research problem, we can consider Y_{ij} to be the the angle (in radians) of the affected tooth j; $j = 1, 2, 3..., n_i$ in subject i; $i = 1, 2, 3, ..., m$. We assume that the observations follow a marginal Generalized von Mises distribution i.e. $Y_{ij} \sim GvM_k(\boldsymbol{\lambda_i})$, $k = 1, 2, ..$ where $\boldsymbol{\lambda_i} = \boldsymbol{\lambda} + u_i$ and $u_i \sim N(0, \tau^2)$. Thus observations for subject i; $i = 1, 2, .., m$ are correlated due to the presence of the random subject effect u_i.

Our simulations are motivated by a clinical trial studied in in Ramseier et al. (2009) and discussed in Section 3.4. In our motivating dataset there were data on 287 affected teeth from 46 subjects. Thus on the average there were $287/46 \approx 6$ affected teeth in each subject. Hence in our simulation examples, we consider 1000 simulated datasets comprising of observations from $m = 100$ subjects with $n_i = 6$ observations from each subject correlated using subject-level random effect. However, it should be noted that the primary goal of our study is to determine regions of the mouth that are most susceptible to periodontal disease which a population averaged measure. Hence though we simulate correlated angles using random effect approach, we analyze the correlated responses using marginal modeling approach. In order to analyze the correlated angular observations we fit Generalized von Mises distributions of varying order and use GEE approach to obtain valid standard errors. Furthermore, we calculate the QIC values of the fitted distributions and determine the distribution that provides the best fit to the data.

3.3.2 Simulation Results

We consider three simulation settings to illustrate the performance of our methods. In each of the settings, we examine the performance of our methods on datasets comprising of independent observations i.e. random effect variance parameter $\tau = 0$ as well as datasets comprising of correlated observations from a subject (corresponding to $\tau = 0.6$). We consider 1000 simulations and in each simulated dataset we have six angular observations from each of the 100 subjects. For the uncorrelated case, we would expect both the model based and robust standard errors to be close to the empirical standard errors. For the correlated case we would expect the robust standard errors to provide an improvement to the model based standard errors and be closer to the empirical standard errors. In each simulation setting, we fit varying orders of GvM distributions to the simulated angular observations θ and get the density estimates $f_\theta(\theta)$. Furthermore, we convert the simulated angles θ to affected tooth numbers t ranging from 1 to 32, using the formula $t = (32\theta/\pi + 1)/2$ and rounding it to the nearest integer. Using the jacobian of transformation, the density of t can be calculated as $g_t(t) = f_\theta(h(t))|\partial\theta/\partial t| = f_\theta(h(t))\pi/16$. We plot the simulated angles (in radians), simulated affected tooth numbers and the corresponding estimated densities for the various orders of the fitted GvM distributions.

In our simulations, the location parameters $(\mu_1, \mu_2, ..., \mu_k)$ are chosen such that there are distinct non-overlapping modes and the values of the concentration parameters $(\kappa_1, \kappa_2, ..., \kappa_k)$ are chosen so that none of the modes are too peaked compared to the other modes. We choose a relatively small value of the random effect variance parameter $\tau = 0.6$ as the parameter values for the underlying marginal distribution are also relatively small and we do not want to distort the original underlying marginal distribution by inducing too much variation between subjects. In order to examine the performance of our methods we focus on the following simulation examples.

Example 1: We simulate from marginal GvM_4 distribution having parameters

$\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07, -0.33, -0.38) which corresponds to $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, μ_4=1, κ_1=0.5, κ_2=0.5, κ_3=0.5, κ_4=0.5). We compare the parameter estimates, estimated densities and QIC values while fitting GvM_4, GvM_3, GvM_2, von Mises (vM) distributions. The results for $\tau = 0$ (uncorrelated case) and $\tau = 0.6$ (correlated case) are stated in Tables 3.1 and 3.2 respectively. Figures 3.2 and 3.3 show the plot of the simulated angles, simulated affected tooth numbers and the corresponding estimated densities for the various fitted distributions under $\tau = 0$ and $\tau = 0.6$ respectively. In both the scenarios, from the plots and the QIC values it is seen that GvM_4 distribution gives the best fit to the data. This validates our methods as the angles were also simulated from a marginal GvM_4 distribution. Under independence, the estimated parameter values of the GvM_4 distribution are $\hat{\boldsymbol{\lambda}}$=(0.07, -0.79, 0.78, -0.42, -0.39, 0.39, -0.37, -0.53) or equivalently $\hat{\boldsymbol{\theta}}$=(4.80, 2.89, 0.79, 1.03, 0.79, 0.89, 0.55, 0.65) which is close to the values of the parameters of the marginal distribution from which the data were simulated.

Example 2: We simulate from marginal GvM_3 distribution having parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07) which corresponds to $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, κ_1=0.5, κ_2=0.5, κ_3=0.5). We compare the parameter estimates, estimated densities and QIC values while fitting GvM_3, GvM_2, vM distributions. The results for $\tau = 0$ (uncorrelated case) and $\tau = 0.6$ (correlated case) are stated in Tables 3.3 and 3.4 respectively. Figures 3.4 and 3.5 show the plot of the simulated angles, simulated affected tooth numbers and the corresponding estimated densities for the various fitted distributions under $\tau = 0$ and $\tau = 0.6$ respectively. In both the scenarios, from the plots and the QIC values it is seen that GvM_3 distribution gives the best fit to the data. This is to be expected as the angles were simulated from a marginal GvM_3 distribution. Under independence, the estimated parameter values of the GvM_3 distribution are $\hat{\boldsymbol{\lambda}}$=(0.12, -0.64, 0.58, -0.22, -0.42, 0.30) or equivalently $\hat{\boldsymbol{\theta}}$=(4.90, 2.96, 0.84, 0.65, 0.62, 0.52) which is close to the values of the parameters of the marginal

distribution from which the data were simulated.

Example 3: We simulate from marginal GvM_3 distribution having parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, 0, 0) which corresponds to $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=0, κ_1=0.5, κ_2=0.5, κ_3=0). Note that in this setting we are actually simulating from a GvM_2 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14) or $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, κ_1=0.5, κ_2=0.5). We are interested in seeing if our methods are able to determine the true distribution of the simulated data in this situation. We compare the parameter estimates, estimated densities and QIC values while fitting GvM_3, GvM_2, vM distributions. The results for $\tau = 0$ and $\tau = 0.6$ are stated in Tables 3.5 and 3.6 respectively. Figures 3.6 and 3.7 show the plot of the simulated angles, simulated affected tooth numbers and the corresponding estimated densities for the various fitted distributions under $\tau = 0$ and $\tau = 0.6$ respectively. In both the scenarios, the results obtained on fitting GvM_3 or GvM_2 are similar. The QIC values obtained on fitting the two distributions are close, though the one corresponding to GvM_3 is slightly lower. Also, from the plot it can be seen that there is no appreciable difference in the estimated densities obtained on fitting the two distributions. For the independence case, and for the best fitting GvM_3 distribution, the estimated value of the parameter $\hat{\boldsymbol{\lambda}}$=(0.15, -0.54, 0.53, -0.16, 0.06, -0.08) or equivalently $\hat{\boldsymbol{\theta}}$=(4.98, 2.99, 1.79, 0.56, 0.55, 0.10) which is close to the values of the parameters of the marginal distribution from which the data are simulated. It should be noted that here we are unable to get the correct estimate of μ_3 which was fixed to be zero in our simulations. This is because for all finite orders of j, μ_j is indeterminate if $\kappa_j = 0$ (Gatto, 2008). For the independence case, and for the best fitting GvM_2 distribution, the estimated value of the parameter $\hat{\boldsymbol{\lambda}}$=(0.14, -0.48, 0.48, -0.14) or $\hat{\boldsymbol{\theta}}$=(5.0, 3.0, 0.5, 0.5) which matches the values of the parameters of the marginal GvM_2 distribution from which the data are actually simulated. Note that a GvM_2 distribution requires the estimation of 4 parameters and 10 unique variance, covariance parameters while a GvM_3 distribution requires the

estimation of 6 parameters and 21 unique variance, covariance parameters. In this setting, from both the plots and the QIC values it can be seen that both GvM_2 and GvM_3 distributions give equivalent model fits. However fitting a GvM_2 distribution requires estimation of fewer number of parameters. Hence in this situation we can choose GvM_2 distribution to give the most parsimonious fit.

Table 3.1 Table showing the results obtained on fitting GvM_4, GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_4 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07, -0.33, -0.38) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, μ_4=1, κ_1=0.5, κ_2=0.5, κ_3=0.5, κ_4=0.5)) and $\tau = 0$ (i.e. independence).

	GvM_4	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.07	0.17	0.10	0.18
$\hat{\lambda_2}$	-0.79	-0.45	-0.41	-0.35
$\hat{\lambda_3}$	0.78	0.46	0.38	
$\hat{\lambda_4}$	-0.42	-0.32	-0.34	
$\hat{\lambda_5}$	-0.39	-0.33		
$\hat{\lambda_6}$	0.39	0.14		
$\hat{\lambda_7}$	-0.37			
$\hat{\lambda_8}$	-0.53			
Empirical se ($\hat{\lambda_1}$)	0.11	0.08	0.06	0.07
Empirical se ($\hat{\lambda_2}$)	0.21	0.11	0.07	0.05
Empirical se ($\hat{\lambda_3}$)	0.17	0.07	0.05	
Empirical se ($\hat{\lambda_4}$)	0.15	0.10	0.07	
Empirical se ($\hat{\lambda_5}$)	0.10	0.11		
Empirical se ($\hat{\lambda_6}$)	0.19	0.19		
Empirical se ($\hat{\lambda_7}$)	0.10			
Empirical se ($\hat{\lambda_8}$)	0.12			
Model Based se ($\hat{\lambda_1}$)	0.06	0.06	0.06	0.06
Model Based se ($\hat{\lambda_2}$)	0.11	0.08	0.07	0.06
Model Based se ($\hat{\lambda_3}$)	0.10	0.07	0.06	
Model Based se ($\hat{\lambda_4}$)	0.08	0.07	0.06	
Model Based se ($\hat{\lambda_5}$)	0.06	0.06		
Model Based se ($\hat{\lambda_6}$)	0.07	0.05		
Model Based se ($\hat{\lambda_7}$)	0.05			
Model Based se ($\hat{\lambda_8}$)	0.06			
Robust se ($\hat{\lambda_1}$)	0.09	0.06	0.06	0.06
Robust se ($\hat{\lambda_2}$)	0.20	0.11	0.07	0.05
Robust se ($\hat{\lambda_3}$)	0.17	0.07	0.06	
Robust se ($\hat{\lambda_4}$)	0.14	0.08	0.06	
Robust se ($\hat{\lambda_5}$)	0.09	0.06		
Robust se ($\hat{\lambda_6}$)	0.11	0.06		
Robust se ($\hat{\lambda_7}$)	0.08			
Robust se ($\hat{\lambda_8}$)	0.09			
Mean QIC	1952	2041	2086	2159

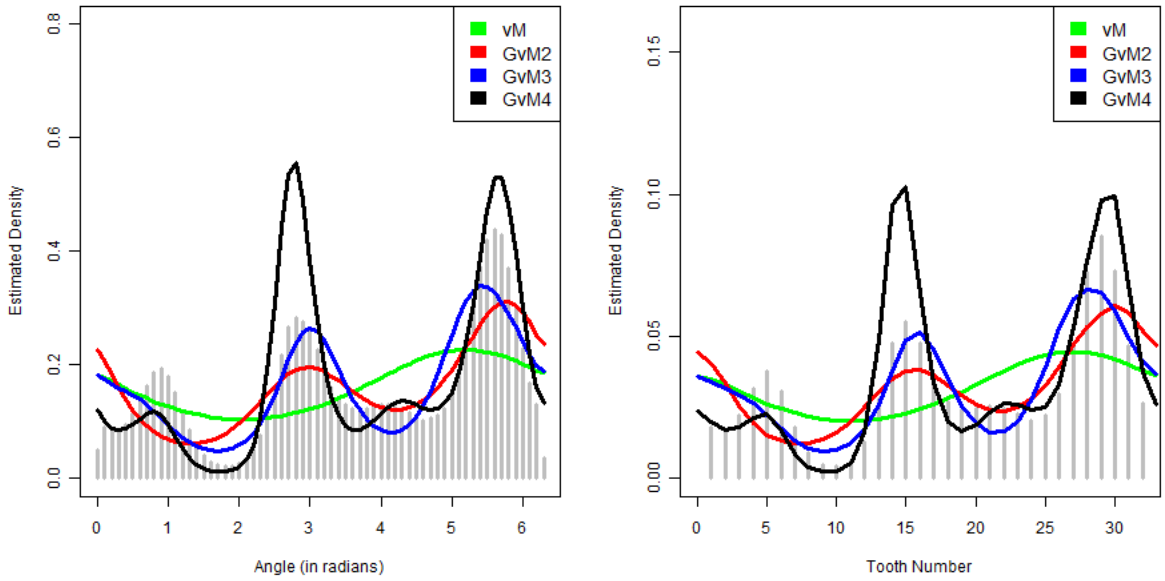

Figure 3.2: Figures showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_4 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07, -0.33, -0.38) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, μ_4=1, κ_1=0.5, κ_2=0.5, κ_3=0.5, κ_4=0.5)) and we consider independence model i.e. $\tau = 0$.

Table 3.2 Table showing the results obtained on fitting GvM_4, GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_4 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07, -0.33, -0.38) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, μ_4=1, κ_1=0.5, κ_2=0.5, κ_3=0.5, κ_4=0.5)) and $\tau = 0.6$.

	GvM_4	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.84	0.66	0.58	0.76
$\hat{\lambda_2}$	-0.59	-0.45	-0.36	-0.33
$\hat{\lambda_3}$	0.79	0.67	0.52	
$\hat{\lambda_4}$	-0.16	-0.26	-0.20	
$\hat{\lambda_5}$	-0.59	-0.20		
$\hat{\lambda_6}$	0.21	-0.05		
$\hat{\lambda_7}$	-0.42			
$\hat{\lambda_8}$	0.26			
Empirical se ($\hat{\lambda_1}$)	0.20	0.09	0.07	0.09
Empirical se ($\hat{\lambda_2}$)	0.32	0.11	0.08	0.07
Empirical se ($\hat{\lambda_3}$)	0.19	0.11	0.07	
Empirical se ($\hat{\lambda_4}$)	0.22	0.13	0.09	
Empirical se ($\hat{\lambda_5}$)	0.33	0.29		
Empirical se ($\hat{\lambda_6}$)	0.30	0.24		
Empirical se ($\hat{\lambda_7}$)	0.19			
Empirical se ($\hat{\lambda_8}$)	0.22			
Model Based se ($\hat{\lambda_1}$)	0.08	0.06	0.06	0.06
Model Based se ($\hat{\lambda_2}$)	0.13	0.09	0.08	0.06
Model Based se ($\hat{\lambda_3}$)	0.10	0.07	0.07	
Model Based se ($\hat{\lambda_4}$)	0.09	0.07	0.06	
Model Based se ($\hat{\lambda_5}$)	0.08	0.05		
Model Based se ($\hat{\lambda_6}$)	0.07	0.05		
Model Based se ($\hat{\lambda_7}$)	0.06			
Model Based se ($\hat{\lambda_8}$)	0.06			
Robust se ($\hat{\lambda_1}$)	0.18	0.08	0.08	0.09
Robust se ($\hat{\lambda_2}$)	0.32	0.12	0.08	0.08
Robust se ($\hat{\lambda_3}$)	0.21	0.09	0.07	
Robust se ($\hat{\lambda_4}$)	0.22	0.13	0.10	
Robust se ($\hat{\lambda_5}$)	0.19	0.06		
Robust se ($\hat{\lambda_6}$)	0.16	0.09		
Robust se ($\hat{\lambda_7}$)	0.15			
Robust se ($\hat{\lambda_8}$)	0.14			
Mean QIC	1829	1899	1939	2020

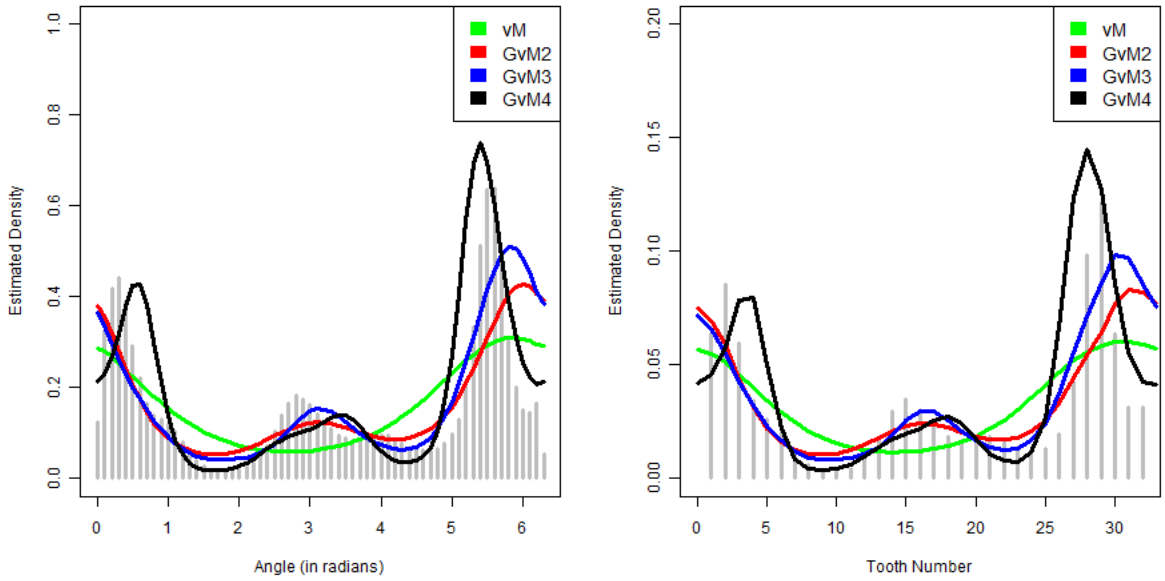

Figure 3.3: Figures showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_4 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07, -0.33, -0.38) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, μ_4=1, κ_1=0.5, κ_2=0.5, κ_3=0.5, κ_4=0.5)) and correlation is induced via random effect variance parameter $\tau = 0.6$.

Table 3.3 Table showing the results obtained on fitting GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, κ_1=0.5, κ_2=0.5, κ_3=0.5)) and $\tau = 0$ (i.e. independence).

	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.12	0.03	0.09
$\hat{\lambda_2}$	-0.64	-0.51	-0.41
$\hat{\lambda_3}$	0.58	0.42	
$\hat{\lambda_4}$	-0.22	-0.25	
$\hat{\lambda_5}$	-0.42		
$\hat{\lambda_6}$	0.30		
Empirical se ($\hat{\lambda_1}$)	0.07	0.06	0.07
Empirical se ($\hat{\lambda_2}$)	0.12	0.07	0.05
Empirical se ($\hat{\lambda_3}$)	0.09	0.06	
Empirical se ($\hat{\lambda_4}$)	0.07	0.06	
Empirical se ($\hat{\lambda_5}$)	0.05		
Empirical se ($\hat{\lambda_6}$)	0.06		
Model Based se ($\hat{\lambda_1}$)	0.06	0.06	0.06
Model Based se ($\hat{\lambda_2}$)	0.09	0.07	0.06
Model Based se ($\hat{\lambda_3}$)	0.07	0.06	
Model Based se ($\hat{\lambda_4}$)	0.07	0.06	
Model Based se ($\hat{\lambda_5}$)	0.06		
Model Based se ($\hat{\lambda_6}$)	0.05		
Robust se ($\hat{\lambda_1}$)	0.06	0.06	0.06
Robust se ($\hat{\lambda_2}$)	0.12	0.07	0.05
Robust se ($\hat{\lambda_3}$)	0.09	0.06	
Robust se ($\hat{\lambda_4}$)	0.08	0.06	
Robust se ($\hat{\lambda_5}$)	0.06		
Robust se ($\hat{\lambda_6}$)	0.06		
Mean QIC	2018	2085	2152

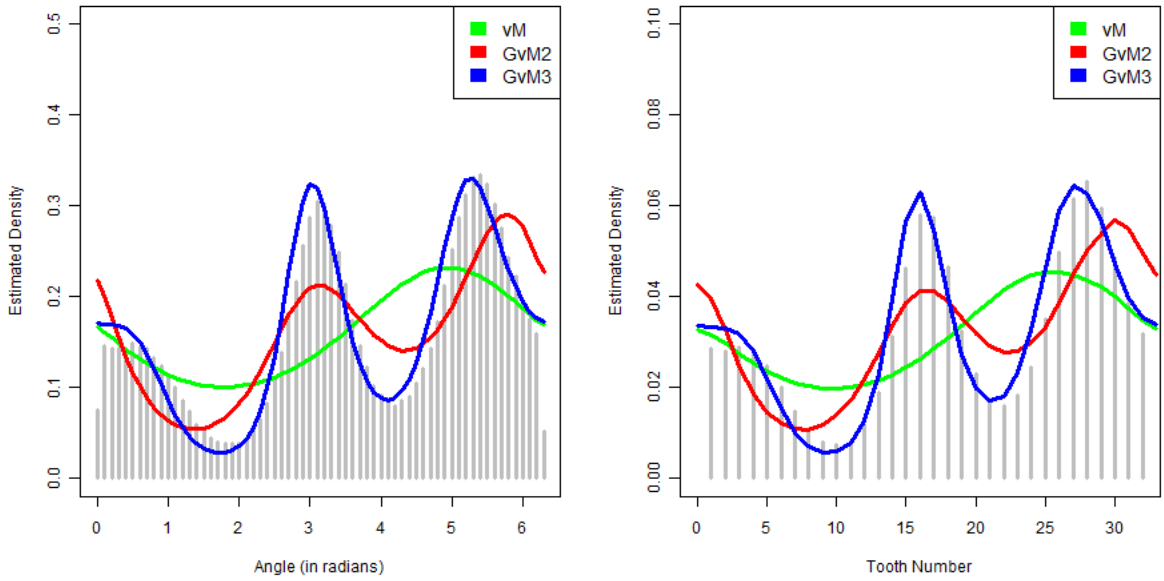

Figure 3.4: Figures showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07) or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, κ_1=0.5, κ_2=0.5, κ_3=0.5) and we consider independence model i.e. $\tau = 0$.

Table 3.4 Table showing the results obtained on fitting GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, κ_1=0.5, κ_2=0.5, κ_3=0.5)) and $\tau = 0.6$.

	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.55	0.41	0.52
$\hat{\lambda_2}$	-0.67	-0.45	-0.36
$\hat{\lambda_3}$	0.74	0.45	
$\hat{\lambda_4}$	-0.06	-0.07	
$\hat{\lambda_5}$	-0.61		
$\hat{\lambda_6}$	0.36		
Empirical se ($\hat{\lambda_1}$)	0.08	0.08	0.09
Empirical se ($\hat{\lambda_2}$)	0.13	0.09	0.08
Empirical se ($\hat{\lambda_3}$)	0.09	0.07	
Empirical se ($\hat{\lambda_4}$)	0.11	0.09	
Empirical se ($\hat{\lambda_5}$)	0.07		
Empirical se ($\hat{\lambda_6}$)	0.06		
Model Based se ($\hat{\lambda_1}$)	0.06	0.06	0.06
Model Based se ($\hat{\lambda_2}$)	0.10	0.07	0.06
Model Based se ($\hat{\lambda_3}$)	0.08	0.06	
Model Based se ($\hat{\lambda_4}$)	0.08	0.06	
Model Based se ($\hat{\lambda_5}$)	0.06		
Model Based se ($\hat{\lambda_6}$)	0.05		
Robust se ($\hat{\lambda_1}$)	0.08	0.08	0.08
Robust se ($\hat{\lambda_2}$)	0.13	0.08	0.08
Robust se ($\hat{\lambda_3}$)	0.09	0.07	
Robust se ($\hat{\lambda_4}$)	0.11	0.09	
Robust se ($\hat{\lambda_5}$)	0.07		
Robust se ($\hat{\lambda_6}$)	0.06		
Mean QIC	1921	2033	2090

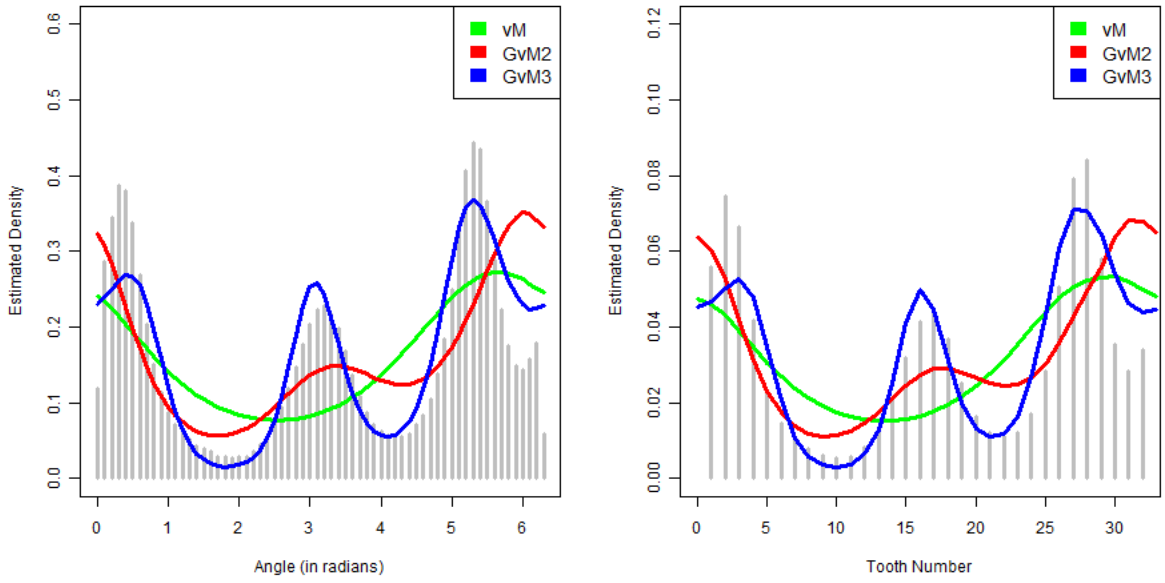

Figure 3.5: Figures showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, -0.49, 0.07) or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=1, κ_1=0.5, κ_2=0.5, κ_3=0.5) and correlation is induced via random effect variance parameter $\tau = 0.6$.

Table 3.5 Table showing the results obtained on fitting GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, 0, 0) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=0, κ_1=0.5, κ_2=0.5, κ_3=0)) and $\tau = 0$ (i.e. independence).

	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.15	0.14	0.21
$\hat{\lambda_2}$	-0.54	-0.48	-0.38
$\hat{\lambda_3}$	0.53	0.48	
$\hat{\lambda_4}$	-0.16	-0.14	
$\hat{\lambda_5}$	0.06		
$\hat{\lambda_6}$	-0.08		
Empirical se ($\hat{\lambda_1}$)	0.06	0.06	0.07
Empirical se ($\hat{\lambda_2}$)	0.09	0.07	0.05
Empirical se ($\hat{\lambda_3}$)	0.08	0.06	
Empirical se ($\hat{\lambda_4}$)	0.07	0.06	
Empirical se ($\hat{\lambda_5}$)	0.08		
Empirical se ($\hat{\lambda_6}$)	0.17		
Model Based se ($\hat{\lambda_1}$)	0.06	0.05	0.06
Model Based se ($\hat{\lambda_2}$)	0.08	0.07	0.06
Model Based se ($\hat{\lambda_3}$)	0.07	0.06	
Model Based se ($\hat{\lambda_4}$)	0.06	0.06	
Model Based se ($\hat{\lambda_5}$)	0.04		
Model Based se ($\hat{\lambda_6}$)	0.04		
Robust se ($\hat{\lambda_1}$)	0.06	0.05	0.06
Robust se ($\hat{\lambda_2}$)	0.09	0.07	0.05
Robust se ($\hat{\lambda_3}$)	0.08	0.06	
Robust se ($\hat{\lambda_4}$)	0.07	0.06	
Robust se ($\hat{\lambda_5}$)	0.04		
Robust se ($\hat{\lambda_6}$)	0.05		
Mean QIC	2069	2080	2150

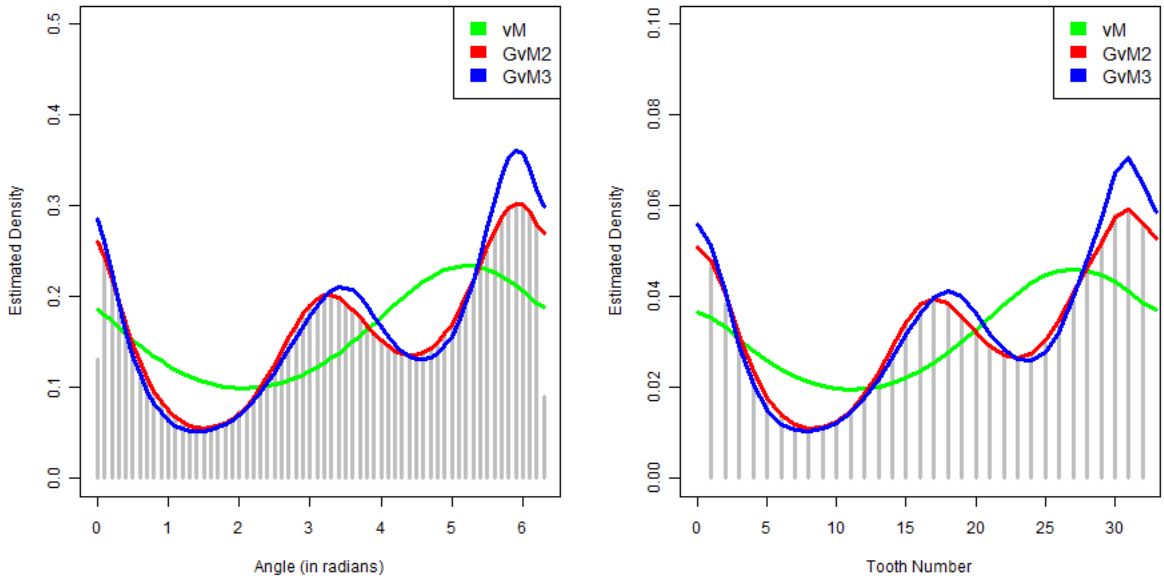

Figure 3.6: Figures showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, 0, 0) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=0, κ_1=0.5, κ_2=0.5, κ_3=0)) and and we consider independence model i.e. $\tau = 0$.

Table 3.6 Table showing the results obtained on fitting GvM_3, GvM_2, vM distributions for 1000 simulated datasets. Each dataset comprises of 100 subjects with each subject having 6 observations and within subject correlation is induced via random effect variance τ^2. Angles are simulated assuming a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14,-0.48,0.48,-0.14,0,0) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=0, κ_1=0.5, κ_2=0.5, κ_3=0)) and $\tau = 0.6$.

	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	0.52	0.48	0.60
$\hat{\lambda_2}$	-0.57	-0.48	-0.37
$\hat{\lambda_3}$	0.63	0.52	
$\hat{\lambda_4}$	-0.01	-0.01	
$\hat{\lambda_5}$	-0.09		
$\hat{\lambda_6}$	0.10		
Empirical se ($\hat{\lambda_1}$)	0.09	0.08	0.09
Empirical se ($\hat{\lambda_2}$)	0.11	0.09	0.08
Empirical se ($\hat{\lambda_3}$)	0.09	0.07	
Empirical se ($\hat{\lambda_4}$)	0.10	0.08	
Empirical se ($\hat{\lambda_5}$)	0.20		
Empirical se ($\hat{\lambda_6}$)	0.16		
Model Based se ($\hat{\lambda_1}$)	0.06	0.06	0.06
Model Based se ($\hat{\lambda_2}$)	0.08	0.07	0.06
Model Based se ($\hat{\lambda_3}$)	0.07	0.06	
Model Based se ($\hat{\lambda_4}$)	0.07	0.06	
Model Based se ($\hat{\lambda_5}$)	0.05		
Model Based se ($\hat{\lambda_6}$)	0.04		
Robust se ($\hat{\lambda_1}$)	0.09	0.08	0.09
Robust se ($\hat{\lambda_2}$)	0.11	0.08	0.07
Robust se ($\hat{\lambda_3}$)	0.09	0.07	
Robust se ($\hat{\lambda_4}$)	0.10	0.08	
Robust se ($\hat{\lambda_5}$)	0.07		
Robust se ($\hat{\lambda_6}$)	0.07		
Mean QIC	1975	1995	2068

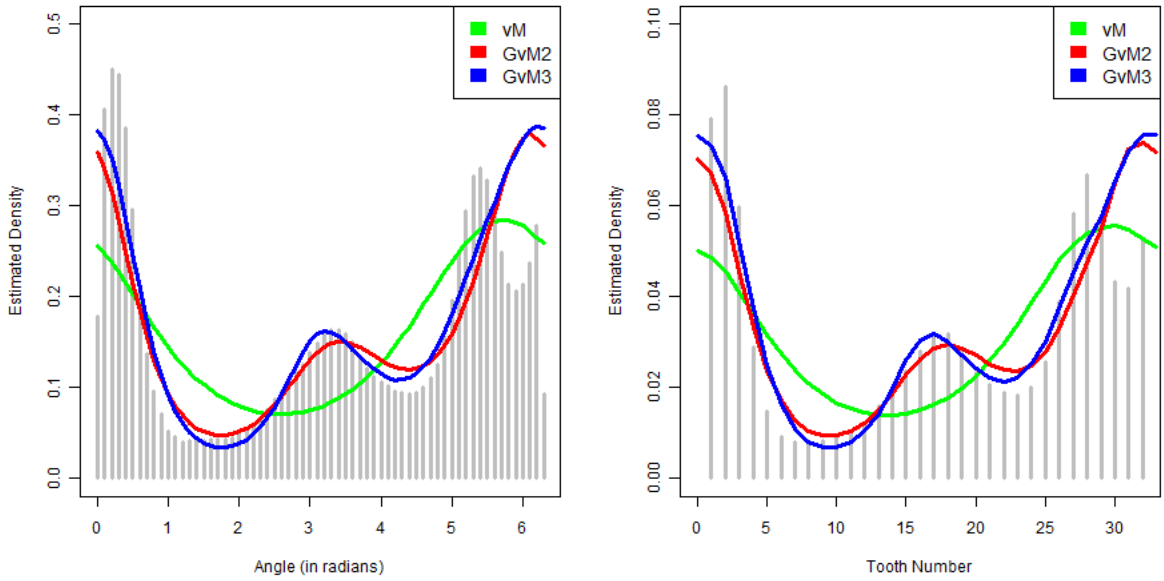

Figure 3.7: Figure showing the simulated angles (left), simulated affected tooth numbers (right) and the corresponding estimated densities for the various fitted distributions. The angles are simulated from a marginal GvM_3 distribution with parameters $\boldsymbol{\lambda}$=(0.14, -0.48, 0.48, -0.14, 0, 0) (or equivalently $\boldsymbol{\theta}$=(μ_1=5, μ_2=3, μ_3=0, κ_1=0.5, κ_2=0.5, κ_3=0)) and correlation is induced via random effect variance parameter $\tau = 0.6$.

In all the simulations, it is seen that the robust standard error estimates are close to the empirical standard errors and provide improvement to the model based standard error estimates for the correlated case. Thus the use of GEE approach for obtaining robust standard error estimates is recommended for analyzing correlated angular observations. From the plots, it should be noted that the QIC value is minimum for that order of the GvM distribution which provides the best fit to the simulated data. In both the first two simulation examples, our algorithm is able to detect the correct order of the GvM distribution from which it is simulated as that yields the minimum QIC value. In the third simulation example when we are simulating from a GvM_3 distribution but with $\lambda_5 = \lambda_6 = 0$, we can see that GvM_3 and GvM_2 distributions provide equivalent fits and the corresponding QIC values

are also similar. Hence our methods are capable of identifying the order of GvM distribution that provides the best fit to the simulated angular data. Furthermore, we are able to obtain valid parameter and standard error error estimates for that distribution.

It should be noted that in Table 3.1 the model based and the robust standard error estimates fail to match the empirical standard error estimates well for certain parameter estimates. We attribute that to the problem of estimating too many parameters relative to the number of observations in the dataset. For a GvM_4 distribution the covariance matrix will have 36 unique elements in the covariance matrix and since there are 600 observations in one simulated dataset we might be estimating too many parameters relative to the number of observations.

Also, we had tried simulating from a lower order GvM distribution and fitting a higher order GvM distribution to the data i.e. simulating from marginal GvM_3 or GvM_2 and fitting GvM_4 distribution to the data. In such situations, we often run into the problem of overfitting i.e. fitting a model which tries to estimate a higher number of modes than what is inherent in the data. In such a situation the results of the direct search algorithm are misleading and thus we run into convergence issues. We feel that in such a situation we are actually fitting an overly complicated model while a relatively simpler model might be more appropriate to that particular dataset. We mostly encounter this issue while fitting GvM_3 or GvM_4 distributions to observations generated from marginal GvM_2 or von Mises distributions.

3.4 Data Analysis

We apply our methods on data obtained from the clinical trial described in Ramseier et al. (2009) and also analyzed in Chapter II. In this non-randomized observational study 100 subjects, 50 periodontally healthy and 50 periodontally diseased, were administered periodontal exams at baseline. Via our methods, we intend to

determine the distribution of affected teeth and thus quantify the locations of teeth that are most susceptible to be periodontally diseased. Hence the dataset to be analyzed should comprise of teeth and their corresponding locations (in radians) that are affected at baseline.

The diseased status of an individual tooth is determined using the tooth averaged clinical attachment level (CAL) value as in Chapter II. Recall that though CAL is measured at six sites of each tooth we are primarily interested in a tooth-level measure i.e. determining locations of affected teeth in mouth. Hence we focus on affected teeth only. We take the mean value of CAL (in mm) at each of the six sites of a tooth and calculate the tooth averaged CAL value. A tooth is considered to be affected if its tooth-level CAL value is greater than 3 mm. In our motivating dataset, there are information on 2646 teeth corresponding to 100 subjects. At baseline, there are 287 affected teeth corresponding to 46 subjects and thus on average there are 6 affected teeth in each subject. Furthermore, since subjects in the study has their wisdom teeth extracted beforehand, each subject has a maximum of 28 teeth and hence we do not have any information about teeth numbered 1, 16, 17, 32.

For our endeavor, we convert the affected tooth numbers into radians assuming all the teeth to be uniformly distributed in a circle of 2π radians as described in Section 3.1. Thus tooth number $t, t = 1, 2, \ldots 32$, is associated with a direction of $\pi(2t-1)/32$ radians. Hence the data to be analyzed comprises of 287 affected teeth and its corresponding locations (in radians) from 46 subjects. We fit various orders of the Generalized von Mises distribution and calculate valid parameter and standard error estimates using the GEE algorithm in order to account for multiple measurements from a subject. We calculate the corresponding QIC values and choose the order of the Generalized Von Mises distribution that has the minimum QIC value. The results are shown in Table 3.7.

From Table 3.7 we see that the QIC values obtained on fitting GvM_3 and GvM_4 are

Table 3.7 Table showing the parameter estimates, robust standard error estimates (in parentheses) and the QIC values obtained on fitting GvM_4, GvM_3, GvM_2, vM distributions for the clinical data described in Ramseier et al. (2009)

	GvM_4	GvM_3	GvM_2	GvM_1
$\hat{\lambda_1}$	-0.10 (0.002)	-0.09 (0.01)	-0.10 (0.11)	-0.09 (0.10)
$\hat{\lambda_2}$	0.09 (0.003)	0.09 (0.01)	0.11 (0.12)	0.11 (0.12)
$\hat{\lambda_3}$	-0.09 (0.02)	-0.08 (0.02)	-0.10 (0.16)	
$\hat{\lambda_4}$	0.01 (0.02)	0.02 (0.003)	0.03 (0.10)	
$\hat{\lambda_5}$	0.04 (0.001)	0.04 (0.004)		
$\hat{\lambda_6}$	0.36 (0.03)	0.35 (0.05)		
$\hat{\lambda_7}$	0.02 (0.001)			
$\hat{\lambda_8}$	-0.01 (0.001)			
QIC	1032.66	1032.76	1050.40	1052.03

very close and are lower than the corresponding QIC values obtained on fitting GvM_2 and GvM_1 distributions. Thus on observing the QIC values, it seems that GvM_3 distribution might provide the best fit to the data. However, the robust standard error estimates corresponding to the GvM_4 and GvM_3 are very small. This might be because in such a situation we might be fitting a model where the number of parameters to be estimated might be too many relative to the number of observations in the dataset. In order to fit a GvM_4 distribution we have to estimate eight parameter values and 36 unique elements in the corresponding covariance matrix. This might be difficult given that we have to base our inference on only 287 observations.

We fit varying orders of GvM distributions to the angular observations corresponding to directions of affected teeth and get the density estimates. We can calculate the density estimates of affected tooth by transforming the density estimates corresponding to the angles. Let us denote t as the affected tooth number and $\theta = \pi(2t-1)/32 = h(t)$ as the corresponding direction. We estimate the density $f_\theta(\theta)$ by fitting a GvM distribution. Using the jacobian of transformation, the density of t can be calculated as $g_t(t) = f_\theta(h(t))|\partial\theta/\partial t| = f_\theta(h(t))\pi/16$.

Figure 3.8 shows the plot of the affected teeth numbers, affected teeth locations (in

radians) and the corresponding estimated densities for the various orders of the fitted GvM distributions. The plot shows that there are three modes in the motivating dataset and GvM_4 or GvM_3 distributions give the best fit to the data. Furthermore, the estimated densities obtained on fitting GvM_4 and GvM_3 distributions are identical. Hence GvM_3 distribution is the most parsimonious model to be fitted to our motivating dataset. The three modes roughly correspond to region of the mouth having teeth numbered 2, 3, 4; teeth numbered 13, 14, 15 and teeth numbered 23, 24, 25. Thus referring to the Figure 3.1 we can infer that in the clinical data stated in Ramseier et al. (2009) there are three regions in the mouth where periodontal disease is most pronounced and they correspond to the upper right corner, upper left corner and middle of the lower jaw. Thus in our motivating dataset we find that the distribution of periodontal disease has a symmetry between the left and right sides of the mouth, as shown in some previous studies (Mombelli and Meier, 2001; Darby et al., 2012).

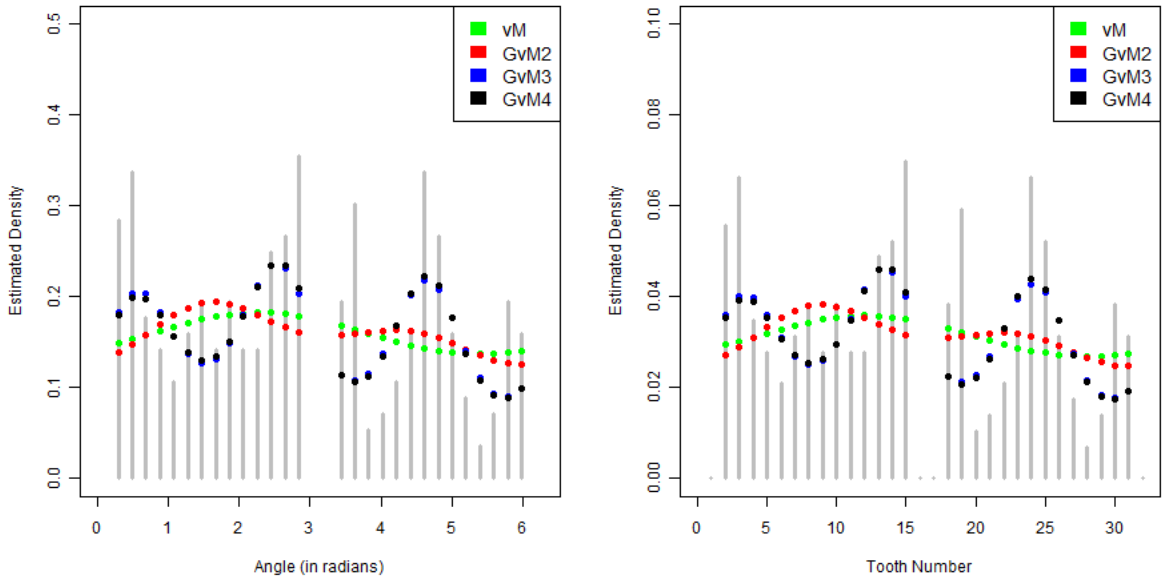

Figure 3.8: Figures showing the affected teeth locations (left), affected teeth numbers (right) and the corresponding estimated densities for the different orders of GvM distributions fitted to the clinical data described in Ramseier et al. (2009).

3.5 Discussion

Our methods are primarily motivated to address the issue of quantifying the number and location of modes of periodontally affected teeth. Based on the Universal numbering system and assuming the mouth as a circle, we are able to convert the affected tooth numbers into directions of affected teeth. Our methods allow us to fit various orders of Generalized von Mises distributions to these angular responses. As an extension to existing literature we propose a methodology to obtain valid parameter and standard error estimates by fitting GvM_k; $k = 1, 2, 3, 4$ distributions to model correlated angular responses. Furthermore, using the QIC measure we are able to identify the order of Generalized von Mises distribution that provides the best fit to the correlated angular data and thus are able to detect the number and locations of periodontally diseased regions of the mouth.

An alternative approach for detecting multiple modes in circular distributions would be to use a mixture of a number of unimodal von Mises distributions. However, Gatto (2008) has stated some disadvantages of adopting such an approach and have recommended the use of Generalized von Mises distribution in such situations. The main disadvantage is that the likelihood of the mixture of $vM(\mu_1, \kappa_1)$ and $vM(\mu_2, \kappa_2)$ distributions might be unbounded and thus the overall supremum of the likelihood of a von Mises mixture does not always yield a consistent estimator.

Fisher (1993) had proposed methods to model marginal von Mises observations as a function of covariates. Artes et al. (2000), Song (2007) had extended that to correlated outcomes. We had followed such an approach in the previous chapter as well. However we are unaware of any such approach for modeling marginal Generalized von Mises observations. Thus we are unable to formally model periodontal disease locations as a function of covariates assuming the angular responses to follow a marginal GvM distribution. The major complication is that though the von Mises distribution belongs to the family of dispersion models the GvM_k distribution for values $k > 1$ do not belong to this family of models. Hence derivation of regression models in such a context presents an interesting research problem.

Our methods allow us to detect the number and location of modes of periodontally affected teeth. Furthermore, our methods can be readily extended to do inference on the density estimates and the resulting modes. It would seem possible to use the Delta method to produce a variance estimate for the density estimate for any angle, although we can also attempt to use resampling methods as an alternative to estimating the variability of the density. Such methods would allow us to calculate confidence interval at each tooth and thus obtain a confidence band around our fitted density estimates. We are developing these extensions in our future work.

CHAPTER IV

Detecting association between clinical periodontal measures using Gibbs point process models

4.1 Introduction

There are 32 teeth in an adult individual. They can be classified into four broad types from the back to the front of the mouth as: molars (grinding teeth), bicuspids, cuspids and incisors (biting teeth). We consider the Universal Numbering System, in which the teeth are numbered sequentially from 1 to 32, with the numbers 1-16 referring to the teeth in the upper jaw (maxillary) and the numbers 17-32 referring to the sixteen teeth in the lower jaw (mandibular). Note that usually the wisdom teeth (teeth 1, 16, 17 and 32) are often removed in most adults and thus periodontal exams usually give us data on 28 teeth. Figure 4.1 shows the orientation of teeth in the mouth with tooth numbers according the Universal Numbering System and the different tooth types classified according to position and functional groupings.

The periodontal status of an individual tooth is determined using a number of clinical measurements most common of which are clinical attachment level (CAL) and alveolar bone level (BL). The point where the crown of the tooth connects with the root of the tooth is known as the cementoenamel junction (CEJ). Any detachment of the gingiva from the cementum produces a periodontal pocket. CAL quantifies the vertical distance (in whole millimeters) from the base of the pocket to the CEJ and represents a measure of disease severity in terms of loss of support for the tooth. Peri-

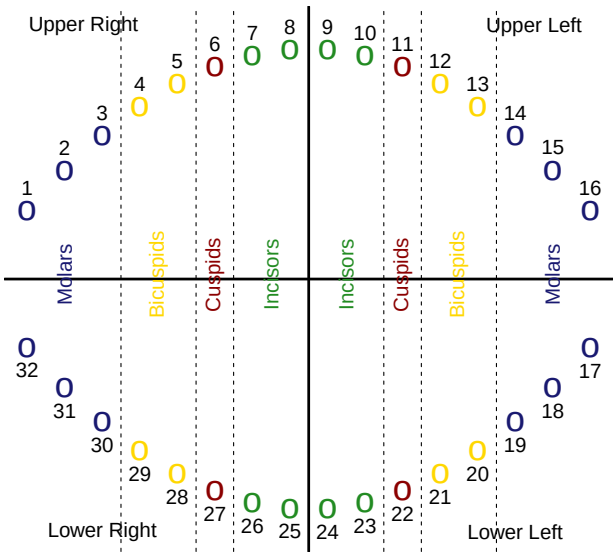

Figure 4.1: Diagram of teeth numbered according to the Universal Numbering System and showing the types of teeth according to position and functional groupings.

odontists detect pockets and thus measure CAL by periodontal probes at each of the mesiobuccal, mid-buccal, distobuccal, mesiolingual, mid-lingual and distolingual sites of every tooth. The bone that surrounds and supports the root of the tooth is called the alveolar bone. Alveolar bone level (BL) quantifies the length (in millimeters) from the CEJ to the crest of alveolar bone.

BL is measured using radiographs usually taken at each of the two surfaces, namely distal (comprising of the distobuccal and distolingual sites) and mesial (comprising of the mesiobuccal and mesiolingual sites) surfaces of the tooth. Because there are no radiographic measurements for buccal and lingual surfaces of a tooth, we discard the CAL values for buccal and lingual sites of each tooth. We then average the BL and CAL values for the distal and mesial sites of each tooth and thus consider tooth-level averaged values. Both these clinical measures are used to determine the

health status of an individual tooth with a high value indicating deterioration of periodontal health which ultimately translates into a high risk of the disease. Thus researchers have been interested in studying the association between these clinical measures and how the association varies according to different types of teeth. It is also of interest to know how the association of BL and CAL measurements vary among different kinds of teeth. Renvert et al. (1981) found high correlation between attachment level and radiographic bone length measures. Goodson et al. (1984) proposed that attachment loss precedes radiographic bone loss particularly during periods of periodontal disease activity. Jeffcoat (1992) and Hausmann et al. (1994) showed that there was a significant concordance between changes in bone level and attachment level. The results from the above-cited studies were based on either Pearson's correlation coefficients or two-way contingency tables and were based on subject-level data. To complement these existing approaches, we propose to assess the association of CAL and BL using spatial point pattern data analysis methods, and our approach allows for the incorporation of multiple tooth-level measurements from each subject.

The primary motivation arises from the distribution of BL and CAL values in different types of teeth as shown in Figure 4.2. The values are obtained from the data for all subjects in the study conducted by Ramseier et al. (2009) where each point represents the tooth-level BL and CAL values, averaged over the distal and mesial sites of each tooth. The data studied in Ramseier et al. (2009) had information on 16 teeth in an individual corresponding to 8 molars and 8 bicuspids. Figure 4.2 shows scatter plot of CAL and BL values for these two types of teeth. We use spatial statistics in order to model the spatial orientation of these points and thus study the association between BL and CAL. Considering the location of the points in the scatter plot as random, we treat the observations as point pattern data.

We are interesting in studying the joint association of the clinical parameters BL

and CAL by modeling the orientation of the points in Figure 4.2. We are particularly interested in seeing if the points in the figure tend to be close to one another or repel each other and how the association varies according to different types of teeth. Our methods have the practical implication of determining if the BL and CAL measures of one tooth provide substantial information about the BL and CAL of other teeth, which our methods would model as strong attraction between points. If the level of attraction is strong enough to suggest that knowing the BL and CAL of one tooth is representative of other teeth, then one might consider not measuring the BL and CAL of every tooth. Doing so would appreciably reduce the time and cost associated with a periodontal exam and could impact the design of future periodontal studies. Such findings would also justify the approach of the National Health and Nutrition Examination Survey (NHANES), which collects periodontal measures on only one-half of each jaw of each subject (Dye et al., 2007a; Dye et al., 2007b).

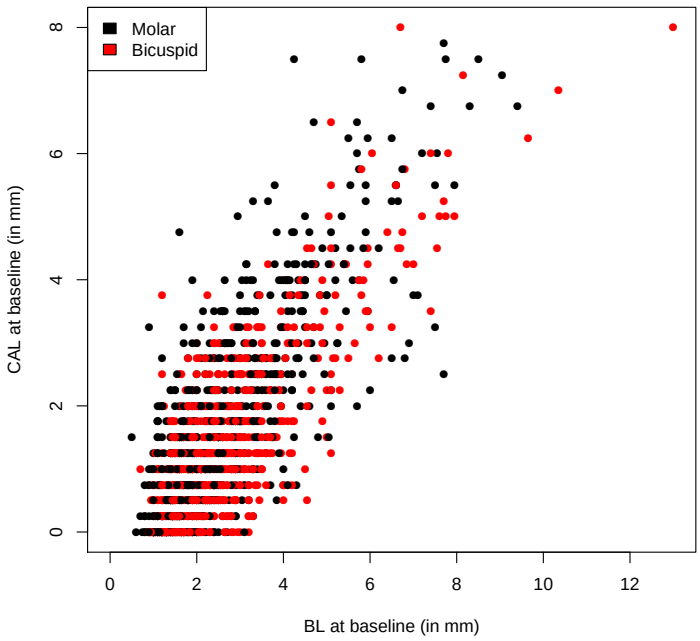

Figure 4.2: Diagram showing association between bone level (BL) and clinical attachment level (CAL) for different types of teeth at baseline in the study of Ramseier et al. (2009)

This chapter is organized as follows. Section 4.2 explains the theory behind point pattern data analysis methods and the modeling framework for studying the overall association and how that association varies between molars and bicuspids. Section 4.3 describes the performance of our methods on simulated data and Section 4.4 shows the application to the data that motivated our analysis. In Section 4.5, we discuss the potential extensions of our methods and the other scenarios to which our modeling approach can be applied.

4.2 Statistical Methods

Spatial statistics usually involve three types of data: geostatistical or point-referenced data, areal or lattice data and point pattern data. Point-referenced data refer to observations collected at finite number of points located in a spatial domain $\mathbf{D}$; lattice data refer to observations collected over a fixed grid or lattice with well defined boundaries. Observations where the locations of the points, the variables of interest, are random are regarded as point pattern data. Banerjee et al. (2004) discuss different methods of analysis of point-referenced and areal data while an overview of current research on spatial statistics is stated in Gelfand et al. (2010).

As shown in Figure 4.2, the locations of the points with coordinates BL and CAL values can be treated as randomly distributed over the window of observations $W = [0, a] \times [0, b]$. As the maximum tooth level BL and CAL values for the baseline data in Ramseier et al. (2009) are 13 mm and 8 mm respectively, we consider points to be randomly distributed in a rectangular window of dimension $a = 13 \times b = 8$.

Moller and Waagepeterson (2007) and Illian et al. (2008) each contain an overview of models and methods for analyzing point pattern data, the most common of which are Cox models and Gibbs process models. Aggregation in point processes can be attributed to two primary types: (i) clustering of the points around points arising from another process and (ii) interactions between points. Broadly, Cox processes

focus on type (i) while Gibbs point processes focus on type (ii). For Cox process models, the observations follow a Poisson process conditional on a random intensity function, so that although the observations arising from a Cox process are aggregated, they are conditionally independent given the intensity function. A difficulty of Cox process models is that the density function typically cannot be expressed in closed form, making evaluation of the likelihood involve a complicated integral.

Gibbs process models are more flexible than Cox process models as they can be used for modeling independence, repulsion and attraction between points. They can model the spatial intensity as a function of covariates and also allow for interaction between points. Gibbs point process models, defined inside a bounded window, have closed form densities. However, given that the density frequently involves an unknown normalizing constant, likelihood inference based on Markov Chain Monte Carlo (MCMC) methods is common for parametric Gibbs point process models. However, an alternative is to directly model the conditional intensity function because it can be easily derived from the corresponding density and does not involve the normalizing constant. The conditional intensity of a point can be loosely interpreted as the conditional probability of the point to be in the point pattern given the rest of the points are already in the point pattern dataset. Hence, the conditional intensity is the preferred modeling tool for Gibbs point processes.

As a result, we focus on modeling point patterns using Gibbs process models. Baddeley and Turner (2000) proposed a simplified approach based on a pseudolikelihood using a loglinear model that has been implemented in the R package "spatstat" (Baddeley and Turner, 2005). However, the usual variance estimates, i.e. information matrix, derived from the pseudolikelihood are inconsistent as they are not based upon an actual likelihood. The only recourse for finding standard error estimates has been using the bootstrap and other resampling techniques in order to obtain consistent standard error estimates. We propose use of a "sandwich" standard error estimator

that is straightforward to implement with loglinear models and is computationally faster than the resampling techniques considered in existing literature.

4.2.1 Gibbs Point Processes

Let $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ be a point pattern of n points, which is a realization of the point process $\mathbf{X}$. Let W denote the window containing all possible realizations of $\mathbf{X}$. The conditional intensity $\lambda(u; \mathbf{x})$ of $\mathbf{x}$ at a location $u \in W$ can be loosely interpreted as the conditional probability that $\mathbf{X}$ has a point at u given the rest of the process coincides with $\mathbf{x}$. For any Gibbs point process on W with density f, the conditional intensity (Papangelou, 1974) at a point $u \in W$ is

$$\lambda(u; \mathbf{x}) = \begin{cases} f(\mathbf{x} \cup \{u\})/f(\mathbf{x}) & (u \notin \mathbf{x}) \\ f(\mathbf{x})/f(\mathbf{x} \setminus \{u\}) & (u \in \mathbf{x}) \end{cases}$$

where $\mathbf{x} \setminus \{u\} = \{v : v \in \mathbf{x}, v \neq u\}$, is the collection of all elements of $\mathbf{x}$ apart from u.

With Gibbs point processes, we typically model the conditional intensity with a loglinear model as

$$\lambda_\theta(u; \mathbf{x}) = \exp(\boldsymbol{\theta}^T \mathbf{S}(u; \mathbf{x})) \tag{4.1}$$

where $\mathbf{S}(u; \mathbf{x})$ is a vector of spatial covariates defined at each point u in W. This includes exponential family likelihoods with canonical parameter $\boldsymbol{\theta}$. The two primary types of Gibbs models are Poisson processes and pairwise interaction models. The inhomogenous Poisson process with intensity function $\lambda(.)$ has conditional intensity $\lambda(u; \mathbf{x}) = \lambda(u)$ at all points u. The fact that this does not depend on $\mathbf{x}$ is a consequence of the independence properties of the Poisson process.

Pairwise interaction models are used to model interpoint interaction or stochastic dependence between points and can be constructed depending on whether the points

tend to be far from one another (repulsion) or tend to be clustered together (attraction). Strauss process (Strauss, 1975) is commonly used to model repulsion between points while Area Interaction process (Baddeley and Van Lieshout, 1995) or Geyer's saturation process (Geyer, 1999) are used to model repulsion or attraction. Note that in our setting, repulsion is biologically implausible because given the BL and CAL value of one tooth, it is unlikely that there cannot be another tooth having a similar BL and CAL value. Furthermore, we want a model flexible enough to determine if there is an attraction, repulsion or independence between the points. Hence, for our purpose we will adopt an Area Interaction point process.

4.2.2 Area Interaction Point Process

The Widom-Rowlinson model (Widom and Rowlinson, 1970) or Area Interaction process (Baddeley and Van Lieshout, 1995) with disc radius r, intensity parameter κ and interaction parameter γ is a point process having probability density

$$f(\mathbf{x}) = \alpha\kappa^{n(\mathbf{x})}\gamma^{-A(\mathbf{x})} \qquad \kappa, \gamma > 0$$

where $\mathbf{x} = (x_1, x_2, ..., x_n)$ represent the points of the pattern, $n(\mathbf{x})$ is the number of points in the pattern, $A(\mathbf{x})$ is the area of the region formed by union of discs of radius r centered at point $\mathbf{x}$ and α is the normalizing constant. Thus an isolated point, i.e. a point lying outside a distance $2r$ from any of the points in the pattern, contributes a factor $\kappa\gamma^{-\pi r^2}$ to the density. Hence the parameters fail to have a ready interpretation. Considering the parameterization $C(\mathbf{x}) = B(\mathbf{x}) - n(\mathbf{x})$, $\beta = \kappa\gamma^{-\pi r^2}$, $\eta = \gamma^{\pi r^2}$ and

$B(\mathbf{x}) = A(\mathbf{x})/\pi r^2$, the above density can be re-expressed as:

$$\begin{aligned} f(\mathbf{x}) &= \alpha\kappa^{n(\mathbf{x})}\gamma^{-A(\mathbf{x})} \\ &= \alpha\kappa^{n(\mathbf{x})}(\gamma^{\pi r^2})^{-A(\mathbf{x})/\pi r^2} \\ &= \alpha\kappa^{n(\mathbf{x})}(\gamma^{\pi r^2})^{-B(\mathbf{x})} \\ &= \alpha(\kappa\gamma^{-\pi r^2})^{n(\mathbf{x})}(\gamma^{\pi r^2})^{-B(\mathbf{x})+n(\mathbf{x})} \\ &= \alpha\beta^{n(\mathbf{x})}\eta^{-C(\mathbf{x})} \end{aligned}$$

When an additional point is added to an existing point pattern $\mathbf{x}$, it contributes a factor $\beta\eta^{\delta}$, $0 \leq \delta \leq 1$, to the density. For a point that is outside a distance $2r$ from each point in the pattern, $\delta = 0$, and its contribution to the density is β. On the other hand, if the new point coincides with any of the existing points in the pattern, then $\delta = 1$. For all other cases, when the new point is within a distance $2r$ of any other point in the pattern, $0 < \delta < 1$. Thus based on the reparameterization $\beta = \kappa\gamma^{-\pi r^2}$ is the intensity parameter, $\eta = \gamma^{\pi r^2}$ is the interaction parameter, $C(\mathbf{x})$ is the interaction potential, and $B(\mathbf{x}) = A(\mathbf{x})/\pi r^2$ is the normalized area covered by the point pattern $\mathbf{x}$.

Thus the corresponding conditional intensity of an Area Interaction process for $u \notin \mathbf{x}$ with interaction radius r is given by:

$$\begin{aligned} \lambda_{\beta,\eta}(u, \mathbf{x}) &= f(\mathbf{x} \cup \{u\})/f(\mathbf{x}) \\ &= \alpha\beta^{n(\mathbf{x}\cup\{u\})}\eta^{-C(\mathbf{x}\cup\{u\})}/\alpha\beta^{n(\mathbf{x})}\eta^{-C(\mathbf{x})} \\ &= \beta\eta^{C(\mathbf{x})-C(\mathbf{x}\cup u)} \\ &= \beta\eta^{-D(u,\mathbf{x})} \qquad \beta, \eta \geq 0 \end{aligned} \tag{4.2}$$

where $D(u, \mathbf{x}) = C(\mathbf{x} \cup u) - C(\mathbf{x}) = (A(\mathbf{x} \cup u) - A(\mathbf{x}))/\pi r^2 - 1$. For an isolated point $u, \parallel u - x_i \parallel \geq 2r$ for all $x_i \in \mathbf{x}$, we get $A(\mathbf{x} \cup u) = A(\mathbf{x}) + \pi r^2$, so that $D(u, \mathbf{x}) = 0$

and $\lambda_{\beta,\eta}(u, \mathbf{x}) = \beta$. For any coincident point $u = x_i$ for any $x_i \in \mathbf{x}$, $A(\mathbf{x} \cup u) = A(\mathbf{x})$, $D(u, \mathbf{x}) = -1$ and $\lambda_{\beta,\eta}(u, \mathbf{x}) = \beta\eta$. For all other points $u, \| u - x_i \| < 2r$ for at least one $x_i \in \mathbf{x}$, $-1 < D(u, \mathbf{x}) < 0$.

Depending upon the parameter values, the Area Interaction process is flexible enough to model a regular or an inhibitive pattern or an attractive or clustered data pattern. If $\eta = 1$, the model reduces to a homogenous Poisson process with intensity β while $\eta < 1$ indicates a regular and $\eta > 1$ indicates clustered point pattern respectively. The interaction radius r is referred to as an "irregular" parameter while the parameters β, η, which are directly estimated based on the data, are referred to as "regular" parameters.

4.2.3 Pseudolikelihood and Parameter Estimation

Assuming data points to be realizations from an Area Interaction process, we use maximum pseudolikelihood method of estimation to estimate β and η. Since this method uses a likelihood that is not the actual likelihood of the data, the resulting standard error estimates derived by taking the inverse of the information matrix using the pseudolikelihood would fail to provide valid standard error estimates. Hence the standard errors derived using our methods should be "robust" to the misspecification of the correct likelihood function. As a result, we use a generalized estimating equation (GEE) approach. Liang and Zeger (1986) proposed the use of GEE to produce consistent parameter and standard error estimates for correlated data assuming the marginal mean model is completely specified. We use the resulting residuals to produce "robust", or so-called "sandwich", variance estimates. The GEE approach, though non-likelihood based, is defined for an unbiased estimating equation and thus should yield valid parameter and standard errors in our setting.

Besag (1977) defined the pseudolikelihood of a point process with conditional intensity $\lambda(u;\mathbf{x})$ over a subset $A \subseteq W$, the window or grid of all observations to be

$$PL_A(\mathbf{x}) = \prod_{x_i \in A} \lambda(x_i;\mathbf{x}) exp(-\int_{u \in A} \lambda(u;\mathbf{x})du) \tag{4.3}$$

Consider m subjects in which the point pattern of subject $k = 1, 2, ..., m$ is $\mathbf{x_k}$=$(x_{k1}, x_{k2}, ..., x_{kn_k})$. Thus subject k has n_k data points. Denote the aggregation of point patterns for all the subjects as $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_m})$. From equation (4.3), taking $A = W$, the pseudolikelihood of a point process for subject k is

$$PL(\boldsymbol{\theta}; \boldsymbol{x_k}) = \prod_{x_{ki} \in W} [\lambda_\theta(x_{ki}; \boldsymbol{x_k}) exp(-\int_W \lambda_\theta(u; \boldsymbol{x_k})du)]$$

Hence the log-pseudolikelihood of the point process for subject k is given by

$$\begin{aligned} logPL(\boldsymbol{\theta}; \boldsymbol{x_k}) &= \sum_{x_{ki} \in W} log\lambda_\theta(x_{ki}; \boldsymbol{x_k}) - \int_W \lambda_\theta(u; \boldsymbol{x_k})du \\ &= \sum_{i=1}^{n_k} log\lambda_\theta(x_{ki}; \boldsymbol{x_k}) - \int_W \lambda_\theta(u; \boldsymbol{x_k})du \end{aligned} \tag{4.4}$$

Assuming m_k quadrature points, including data and dummy points which span the window W, we can approximate the above integral by a finite summation as

$$\int_W \lambda_\theta(u; \boldsymbol{x_k})du \approx \sum_{j=1}^{m_k} \lambda_\theta(u_{kj}; \boldsymbol{x_k}) w_{kj}$$

where w_{kj} is quadrature weight $j = 1, 2, ..., m_k$ for person $k = 1, 2, ..., m$. Thus, we can write the log-pseudolikelihood of the point process for subject k as

$$\begin{aligned} logPL(\boldsymbol{\theta}; \boldsymbol{x_k}) &= \sum_{i=1}^{n_k} log\lambda_\theta(x_{ki}; \boldsymbol{x_k}) - \sum_{j=1}^{m_k} \lambda_\theta(u_{kj}; \boldsymbol{x_k}) w_{kj} \\ &= \sum_{j=1}^{m_k} (y_{kj} log\lambda_{kj} - \lambda_{kj}) w_{kj} \end{aligned}$$

where $\lambda_{kj} = \lambda_\theta(u_{kj}; \boldsymbol{x_k})$ and $y_{kj} = z_{kj}/w_{kj}$ and

$$z_{kj} = \begin{cases} 1 & \text{if } u_{kj} \text{ is a data point in } \boldsymbol{x_k} \; ; u_{kj} \in \boldsymbol{x_k} \\ 0 & \text{if } u_{kj} \text{ is a dummy point in } \boldsymbol{x_k} \; ; u_{kj} \notin \boldsymbol{x_k} \end{cases}$$

For a fixed $\boldsymbol{x_k}$, the above is equivalent to a log-likelihood of weighted independent Poisson random variables Y_{kj} with mean λ_{kj} and weight w_{kj}. Assuming all the point patterns generated from the m subjects are independent, the pseudolikelihood for all the subjects is

$$\begin{aligned} PL_W(\boldsymbol{\theta}; \boldsymbol{X}) = \prod_{k=1}^{m} PL_W(\theta; \boldsymbol{x_k}) &= \prod_{k=1}^{m} \prod_{x_{ki} \in W} [\lambda_\theta(x_{ki}; \boldsymbol{x_k}) exp(-\int_W \lambda_\theta(u; \boldsymbol{x_k}) du)] \\ &= \prod_{k=1}^{m} \prod_{i=1}^{n_k} [\lambda_\theta(x_{ki}; \boldsymbol{x_k}) exp(-\int_W \lambda_\theta(u; \boldsymbol{x_k}) du)] \end{aligned}$$

We can write the log-pseudolikelihood for all the subjects as

$$logPL_W(\boldsymbol{\theta}; \boldsymbol{X}) = \sum_{k=1}^{m} logPL_W(\boldsymbol{\theta}; \boldsymbol{x_k}) = \sum_{k=1}^{m} \sum_{j=1}^{m_k} (y_{kj} log\lambda_{kj} - \lambda_{kj}) w_{kj} \tag{4.5}$$

The above equation represents the weighted log-likelihood of marginally independent Poisson random variables. For our purpose, we apply a GEE approach for inference on the parameter $\boldsymbol{\theta}$ using a loglinear model. The GEE methodology for marginal Poisson data is briefly discussed below.

From equation (4.1) note that $\lambda_{kj} = f(x_{kj}^T\boldsymbol{\theta})$. Thus for subject k, $\boldsymbol{\lambda_k} = \boldsymbol{f_k(\theta)} = (\lambda_{k1}, \lambda_{k2}, ..., \lambda_{km_k}) = [f(x_{k1}^T\boldsymbol{\theta}), f(x_{k2}^T\boldsymbol{\theta}), ..., f(x_{km_k}^T\boldsymbol{\theta})]$. For marginally Poisson distributed random variables, the generalized estimating equation can be written as:

$$\sum_{k=1}^{m} \Delta_k^T \Lambda_k^{-1} W_k [Y_k - \boldsymbol{f_k(\theta)}] = 0$$

where Δ_k is a m_k by p matrix whose (j, s) element is the derivative of λ_{kj} with respect to θ_s, $s = 1, 2, ..., p$. Λ_k is the working covariance matrix and $W_k = diag(w_{k1}, w_{k2}, ..., w_{km_k})$ is the weight matrix having the quadrature weights in the diagonal. It is to be recalled that we use a working independence correlation matrix for applying the GEE methodology i.e. we assume the observations from a subject to be uncorrelated and use GEE to obtain "robust" standard errors. Thus, for marginally Poisson random variables, the "working" covariance matrix is $\Lambda_k = G_k = diag(\lambda_{k1}, \lambda_{k2}, ..., \lambda_{km_k})$. Furthermore, it can be shown that $\Delta_k = G_k X_k$ where X_k is the design matrix of order m_k by p. Hence the generalized estimating equation can be expressed as:

$$\begin{aligned}
\sum_{k=1}^{m} \Delta_k^T \Lambda_k^{-1} W_k [\boldsymbol{Y_k} - \boldsymbol{f_k(\theta)}] &= 0 \\
\sum_{k=1}^{m} X_k^T G_k G_k^{-1} W_k [\boldsymbol{Y_k} - \boldsymbol{f_k(\theta)}] &= 0 \\
\sum_{k=1}^{m} X_k^T W_k [\boldsymbol{Y_k} - \boldsymbol{\lambda_k}] &= 0
\end{aligned}$$

The model based variance estimator of $\boldsymbol{\theta}$ is given by $V_{MB}(\boldsymbol{\theta}) = S(\boldsymbol{\theta})^{-1}$ where

$$S(\boldsymbol{\theta}) = \sum_{k=1}^{m} \Delta_k^T \Lambda_k^{-1} W_k \Delta_k = \sum_{k=1}^{m} X_k^T W_k G_k X_k$$

The robust variance estimator of $\boldsymbol{\theta}$ is given by $V_R(\boldsymbol{\theta}) = S(\boldsymbol{\theta})^{-1}V(\boldsymbol{\theta})S(\boldsymbol{\theta})^{-1}$ where

$$\begin{aligned} V(\boldsymbol{\theta}) &= \sum_{k=1}^{m} \Delta_k^T \Lambda_k^{-1} W_k \boldsymbol{r_k} \boldsymbol{r_k}^T W_k \Lambda_k^{-1} \Delta_k \\ &= \sum_{k=1}^{m} X_k^T W_k \boldsymbol{r_k} \boldsymbol{r_k}^T W_k X_k \end{aligned}$$

and $\boldsymbol{r_k} = \boldsymbol{Y_k} - \hat{\boldsymbol{\lambda}}_{\boldsymbol{k}}$ is the vector of estimated residuals.

There are a number of ways to calculate the quadrature weights. For our purpose, we generate dummy points spanning the entire observation window and divide the rectangular window into tiles of equal area, say squares of unit area each. To each data or dummy point u_{kj} of the k th subject we assign a weight $w_{kj} = a/b_{kj}$ where $a = 1$ is the area of each tile and b_{kj} is the total number of quadrature points, including data and dummy points, of the k th subject in the same tile as u_{kj}.

From equation (4.2), the conditional intensity of an Area Interaction process at a point $u_{kj} \in \boldsymbol{x_k}$ is

$$\lambda_{kj} = \lambda_{\beta,\eta}(u_{kj}; \boldsymbol{x_k}) = \beta\eta^{-D(u_{kj};\boldsymbol{x_k})} = \beta\eta^{-D_{kj}} \tag{4.6}$$

where $D_{kj} = D(u_{kj}; \boldsymbol{x_k})$. Hence the conditional intensity can be written in a loglinear form as: $\log(\lambda_{kj}) = \log(\beta) - D_{kj}\log(\eta) = \theta_1 - \theta_2 D_{kj}$. Thus, $\lambda_{kj} = \lambda_\theta(u_{kj}; \boldsymbol{x_k}) = \exp(\boldsymbol{\theta}^T \boldsymbol{S}(u_{kj}; \boldsymbol{x_k})) = \exp(\boldsymbol{\theta}^T \boldsymbol{S_{kj}})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2)^T = (\log(\beta), \log(\eta))^T$ and $\boldsymbol{S}(u_{kj}; \boldsymbol{x_k}) = \boldsymbol{S_{kj}} = (1, -D_{kj})^T = (1, -D(u_{kj}; \boldsymbol{x_k}))^T$. Using our algorithm we intend to compute valid parameter and standard error estimates of $\boldsymbol{\theta}$, which is a function of the regular parameters β and η. The parameter $\boldsymbol{\theta}$ quantifies the overall association between the clinical parameters CAL and BL. The irregular parameter r, the interaction radius, can be obtained by choosing the corresponding values which maximize the profile pseudolikelihood over a fixed grid of r values (Baddeley and Turner, 2000).

4.2.4 Inference for Multitype Point Patterns

Each point in a spatial point pattern may carry additional information called a 'mark'. Point pattern data with categorical marks are referred to as a multitype point pattern. For our purpose, we can classify the points in the point pattern shown in Figure 4.2 according to bicuspids and molars and treat these types of teeth as "marks" associated with the point pattern dataset.

Consider m subjects with the point pattern of the k th subject as $\mathbf{v_k} = \{(\mathbf{x_{k1}}, \mathbf{z_{k1}}), (\mathbf{x_{k2}}, \mathbf{z_{k2}})..., (\mathbf{x_{kn_k}}, \mathbf{z_{kn_k}})\}$; $x_{kj} \in W, z_{kj} \in M; j = 1, 2, ..n_k$. We consider the categorical or discrete mark for tooth j of subject k as

$$z_{kj} = \begin{cases} 1 & \text{if tooth } j \text{ is a molar} \\ 2 & \text{if tooth } j \text{ is a bicuspid} \end{cases}$$

Hence $M = \{1, 2\}$. Similar to equation (4.4), the log-pseudolikelihood of a multitype point process for subject k is given by

$$\begin{aligned} logPL(\theta; \boldsymbol{v_k}) &= \sum_{z_{ki} \in M} \sum_{x_{ki} \in W} log\lambda_\theta((x_{ki}, z_{ki}); \boldsymbol{v_k}) - \sum_{z_{ki} \in M} \int_W \lambda_\theta((u, z_{ki}); \boldsymbol{v_k}) du \\ &= \sum_{l=1}^{2} \sum_{i=1}^{n_{kl}} log\lambda_\theta((x_{ki}, l); \boldsymbol{v_k}) - \sum_{l=1}^{2} \int_W \lambda_\theta((u, l); \boldsymbol{v_k}) du \end{aligned}$$

Assuming m_{kl} quadrature points for person k and points of mark l, we can approximate the above integral by a summation as

$$\int_W \lambda_\theta((u, l); \boldsymbol{v_k}) du \approx \sum_{j=1}^{m_{kl}} \lambda_\theta((u_{kj}, l); \boldsymbol{v_k}) w_{kjl}$$

where w_{kjl} is the quadrature weight corresponding to quadrature point j for person k and points of mark l. Thus, we can write the log-pseudolikelihood of the point

process for subject k as

$$\begin{aligned} logPL(\theta; \boldsymbol{v_k}) &= \sum_{l=1}^{2}\sum_{i=1}^{n_k} log\lambda_\theta((x_{ki}, l); \boldsymbol{v_k}) - \sum_{l=1}^{2}\sum_{j=1}^{m_{kl}} \lambda_\theta((u_{kj}, l); \boldsymbol{v_k}) w_{kjl} \\ &= \sum_{l=1}^{2}\sum_{j=1}^{m_{kl}} (y_{kjl} log\lambda_{kjl} - \lambda_{kjl}) w_{kjl} \end{aligned}$$

$$z_{kjl} = \begin{cases} 1 & \text{if } u_{kjl} \text{ is a data point in } \boldsymbol{v_k} \text{ of type } l \ ; u_{kjl} \in \boldsymbol{v_k} \\ 0 & \text{if } u_{kjl} \text{ is a dummy point in } \boldsymbol{v_k} \text{ of type } l \ ; u_{kjl} \notin \boldsymbol{v_k} \end{cases}$$

For a fixed $\boldsymbol{v_k}$, the above is equivalent to a log-likelihood of weighted independent Poisson random variables Y_{kjl} with mean λ_{kjl} and weight w_{kjl}.

Denote the aggregation of point patterns for all the subjects as $\mathbf{V} = (\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_m})$. Thus, we can write the log-pseudolikelihood of the point process for all subjects is

$$logPL_W(\theta; \boldsymbol{V}) = \sum_{k=1}^{m} logPL_W(\theta; \boldsymbol{v_k}) = \sum_{k=1}^{m}\sum_{l=1}^{2}\sum_{j=1}^{m_{kl}} (y_{kjl} log\lambda_{kjl} - \lambda_{kjl}) w_{kjl} \qquad (4.7)$$

For categorical marks, the log-pseudolikelihood (4.7) can be expressed in the form of the log-pseudolikelihood (4.5) using categorical predictors in the corresponding design matrix. Thus the GEE algorithm described in Section 4.2.3 can be applied in a similar manner.

The conditional intensity of a two-type Area Interaction process for quadrature point j in the two-type marked point pattern v_k for person k is

$$\lambda_\theta((u_{kj}, l); v_k) = \beta_l \eta_{l1}^{-D_1((u_{kj},l);\boldsymbol{v_k})} \eta_{l2}^{-D_2((u_{kj},l);\boldsymbol{v_k})} \qquad (4.8)$$

where $D_1((u_{kj}, l); \boldsymbol{v_k}) = B_1((u_{kj}, l); \boldsymbol{v_k})/\pi r^2 - 1$, $D_2((u_{kj}, l); \boldsymbol{v_k}) = B_2((u_{kj}, l); \boldsymbol{v_k})/\pi r^2 - 1$ and $B_{l'}((u_{kj}, l), \boldsymbol{v_k})$ is the area of the disc of radius $r_{ll'}$ centered on u_{kj} (having mark

l) that is not covered by discs of radius $r_{ll'}$ centered at other points of the kth person $\mathbf{v_k} = \{(\mathbf{x_{k1}}, \mathbf{z_{k1}}), (\mathbf{x_{k2}}, \mathbf{z_{k2}})..., (\mathbf{x_{kn_k}}, \mathbf{z_{kn_k}})\}$ having mark l'. In our setup, $r_{12} = r_{21}$ and $\eta_{12} = \eta_{21}$.

The model may be cast in loglinear form (4.1) with the parameter vector $\boldsymbol{\theta} = (\log \beta_1, \log \beta_2, \log \eta_{11}, \log \eta_{12}, \log \eta_{22})$ and the corresponding predictors $I_1(l), I_2(l)$, $-D_{11}((u,l);\boldsymbol{v}), -D_{21}((u,l);\boldsymbol{v}) - D_{12}((u,l);\boldsymbol{v}), -D_{22}((u,l);\boldsymbol{v})$ respectively where $D_{11}((u,l);\boldsymbol{v}) = D_1((u,l);\boldsymbol{v})I_1(l), D_{21}((u,l);\boldsymbol{v}) = D_2((u,l);\boldsymbol{v})I_1(l), D_{12}((u,l);\boldsymbol{v}) = D_1((u,l);\boldsymbol{v})I_2(l), D_{22}((u,l);\boldsymbol{v}) = D_2((u,l);\boldsymbol{v})I_2(l)$ and $I_m(l) = \mathbf{1}\{m = l\}$. Using the parameter $\boldsymbol{\theta}$ we can quantify the difference in association between BL and CAL according to different types of teeth. The parameters β_1, β_2 controls the overall intensities for the two types of points while the interaction parameters $\eta_{11}, \eta_{12}, \eta_{22}$ govern the interaction or dependence between the points. η_{11} quantifies the association between any point of mark 1 and any other point having the same mark i.e. mark 1. $\eta_{11} < 1$ denote that any two points having mark 1 tend to repel each other while $\eta_{11} > 1$ denote that any two points having mark 1 tend to be close one other and thus there is a clustering between points. Similarly, $\eta_{12} = \eta_{21}$ quantifies the association between any point of type 1 and another point of type 2 and vice versa. $\eta_{12} < 1$ denote that two points having marks 1 and 2 tend to be far from one another while $\eta_{12} > 1$ denote that any two points having marks 1 and 2 tend to be close one other and thus there is a clustering between points of different marks.

4.3 Simulation Examples

4.3.1 Simulation of unmarked point pattern data

We assume that the observations arising from subject k follow an Area Interaction process with intensity parameter β and interaction parameter η. It can also be expressed in loglinear notation as: $\log(\lambda_{kj}) = \log(\beta) - D_{kj} \log(\eta)$. For each sub-

ject, we simulate data points having this conditional intensity and generate dummy points spanning the entire window. The data points are generated using a Metropolis-Hastings algorithm starting with 200 randomly generated points and is run for 100,000 iterations. These points, consisting of both data and dummy points, constitute the quadrature points that are used to compute the quadrature weights, responses and the covariate D as defined in equation (4.6). We simulate points using the algorithm for all subjects in the dataset. We consider different settings for varying values of the parameters β and η.

We use generalized estimating equations (GEE), assuming a "working" independence correlation structure, to get valid parameter and standard error estimates, for the model $\log(\lambda_{kj}) = \alpha_0 - \alpha_1 D_{kj}$ where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ are the parameters of interest. Also, note that we would expect $\exp(\hat{\alpha_0}) \approx \beta$ and $\exp(\hat{\alpha_1}) \approx \eta$.

We consider two simulation settings. In each setting, we simulated data for 30 subjects in each of 500 datasets. In each simulated dataset, we simulated an unmarked point pattern for every subject from an Area Interaction process with intensity parameter β, interaction parameter η and interaction radius $r = 0.5$ on a square window of dimension 13×13. We chose $\beta = 0.1, \eta = 1.6$ and $\beta = 0.2, \eta = 1.6$ for the two simulation settings, respectively. The results are displayed in Table 4.1.

The parameters for the simulation settings were chosen such that the simulated data might mimic actual periodontal data. However, in order to examine the performance of the algorithm for point pattern data, we need to have a sufficient number of points per subject, although realistically a subject can have no more than 28 points. Thus, we chose to use $\beta = 0.1, \eta = 1.6$ for one setting and $\beta = 0.2, \eta = 1.6$ for the other setting which on average contributed 17 and 36 observations per subject on a window of dimension 13×13 units in our simulated datasets. Since repulsion is unlikely in our setting, we chose to make $\eta > 1$ indicating interpoint attraction or clustering between the points. The value of the irregular parameter $r = 0.5$ implies

that points interact only if the distance between them is less than 2*0.5=1 unit, which is realistic as we are focusing on a window of dimension 13 $\times$ 13 units.

Table 4.1 Table showing the mean of estimated parameters and corresponding standard error estimates for 500 simulated datasets of 30 persons each where points from each subject follow an Area Interaction process with intensity parameter β, interaction parameter η and radius of interaction $r = 0.5$.

				Standard Error		
Setting	Parameter	Mean	Exp(mean)	Empirical	Model Based	Robust
$\beta = 0.1, \eta = 1.6$	$\hat{\alpha_0}$	-2.31	0.10	0.06	0.06	0.06
	$\hat{\alpha_1}$	0.38	1.46	0.42	0.28	0.41
$\beta = 0.2, \eta = 1.6$	$\hat{\alpha_0}$	-1.63	0.20	0.05	0.04	0.05
	$\hat{\alpha_1}$	0.45	1.57	0.22	0.15	0.22

From Table 4.1 note that $\exp(-2.31) = 0.10$, $\exp(0.38) = 1.46$ and $\exp(-1.63) = 0.20$, $\exp(0.45) = 1.56$ respectively which are close to the values of the parameters β, η which denote the intensity and interaction parameters for the Area Interaction process from which the data are simulated. Thus our algorithm yields valid estimates of the intensity and the interaction parameters. In both the scenarios, the model based standard errors calculating using pseudolikelihood are inconsistent as the correct likelihood function is not used. Furthermore, the robust standard errors are closer to the empirical standard errors for all the scenarios. The model based standard errors are negatively biased while the robust standard errors give valid estimates. In our setting, apart from bootstrapping and other resampling techniques there are no analytic formulas to calculate the standard errors of the parameter estimates. Thus our methods provide an alternative and relatively straightforward approach to obtain valid standard error estimates for modeling point pattern data.

4.3.2 Simulation of two type point pattern data

The simulation of multitype point pattern data having an Area Interaction conditional intensity is challenging as there is no straightforward algorithm for generating

such observations. Hence, to simulate observations for two types of points (pertaining to two types of teeth) in an individual we follow an approach similar to the one adopted in the previous section. Let us denote the two types of points as points having marks 1 and 2. For every subject, we simulate observations corresponding to an Area Interaction process with regular parameters (β_i, η_i) and interaction radius $r_i; i = 1, 2$. Based on this conditional intensity, we simulate data points corresponding to each mark for each subject. We also generate dummy points spanning the entire window for each of the two marks. These two types of quadrature points are used to calculate the corresponding quadrature weights and the responses. Using the location of the two types of points we calculate the covariates $D_{11}, D_{12}, D_{21}, D_{22}$ as defined in Section 4.2.4. Thus in our simulations though we know the intensity parameters: β_1, β_2 of the two types of points *apriori*, we do not know the true value of the interaction parameter η_{12} corresponding to the interaction between the two types of points.

We simulate points using the algorithm for different values of the intensity and interaction parameters. We fit a marginal model to the simulated data and use GEE to get valid parameter and standard error estimates. We fit the model:
$\log(\lambda_{kj}) = \log(\lambda(u_{kj}, l); \boldsymbol{v_k})) = \alpha_0^m + \alpha_1^m I_2(l) - \alpha_2^m D_{1kj} I_1(l) - \alpha_3^m (D_{1kj} I_2(l) + D_{2kj} I_1(l)) - \alpha_4^m D_{2kj} I_2(l)$ where $\boldsymbol{\alpha^m} = (\alpha_0^m, \alpha_1^m, \alpha_2^m, \alpha_3^m, \alpha_4^m)$ are the parameters of interest. We would expect $\exp(\hat{\alpha_0^m}) \approx \beta_1$, the overall intensity parameter for the type 1 points and $\exp(\hat{\alpha_0^m} + \hat{\alpha_1^m}) \approx \beta_2$, the overall intensity parameter for the type 2 points.

In each of the two settings, we consider 500 simulations each consisting of 30 persons with data on a rectangular window of dimension 10×10. In the first simulation example, we chose $\beta_1 = 0.06$, $\beta_2 = 0.08$, $\eta_1 = \eta_2 = 1.6$ and $r_1 = r_2 = 0.5$. In the second simulation example, we choose $\beta_1 = 0.08$, $\beta_2 = 0.10$, $\eta_1 = \eta_2 = 1.6$ and $r_1 = r_2 = 0.5$. The results are shown in Table 4.2.

As in the simulation settings for unmarked point pattern data, here also the parameter values for the simulations were chosen such that the simulated data might

be similar to what we would expect for periodontal data. On an average, intensity parameter values of 0.06, 0.08 and 0.1 contributed 6, 8 and 10 observations per subject. The values of the interaction parameters $\eta_1 = \eta_2 = 1.6$ and irregular parameters $r_1 = r_2 = 0.5$ imply that there is a clustering between the points and any two points interact only if the distance between them is less than 2*0.5=1 unit.

Table 4.2 Table showing the mean of estimated parameters and corresponding standard error estimates for 500 simulated datasets of 30 persons each where points from each subject follow a two-type Area Interaction process with parameters (i): $\beta_1 = 0.06, \beta_2 = 0.08, \eta_1 = \eta_2 = 1.6, r_1 = r_2 = 0.5$ and (ii): $\beta_1 = 0.08, \beta_2 = 0.10, \eta_1 = \eta_2 = 1.6, r_1 = r_2 = 0.5$.

			Standard Error		
Setting	Parameter	Mean	Empirical	Model Based	Robust
$\beta_1 = 0.06, \beta_2 = 0.08$	$\hat{\alpha_0^m}$	-2.83	0.11	0.10	0.16
	$\hat{\alpha_1^m}$	0.30	0.14	0.13	0.21
	$\hat{\alpha_2^m}$	0.29	1.12	0.68	0.94
	$\hat{\alpha_3^m}$	-0.09	0.63	0.43	0.60
	$\hat{\alpha_4^m}$	0.39	0.72	0.48	0.69
$\beta_1 = 0.08, \beta_2 = 0.10$	$\hat{\alpha_0^m}$	-2.54	0.10	0.09	0.14
	$\hat{\alpha_1^m}$	0.22	0.13	0.12	0.18
	$\hat{\alpha_2^m}$	0.42	0.68	0.48	0.68
	$\hat{\alpha_3^m}$	-0.01	0.42	0.33	0.46
	$\hat{\alpha_4^m}$	0.42	0.58	0.39	0.56

From Table 4.2, we note that $\exp(-2.83) = 0.06$, $\exp(-2.83 + 0.30) = 0.08$, $\exp(-2.54) = 0.08$ and $\exp(-2.54 + 0.22) = 0.10$, all of which are equal to the values of the parameters β_1, β_2 which denote the intensity parameters for the two Area interaction processes pertaining to the two types of points. Thus our algorithm yields valid estimates of the intensity parameters. However we are unable to compare the estimates of the interaction parameters since we are not simulating from a two-type Area Interaction process directly. Note that the model based standard errors are negatively biased while the robust standard errors are closer to the empirical standard errors in both the situations. The results are consistent with those obtained for an unmarked point pattern data considered in Tables 4.1. Hence, in both the

scenarios our proposed standard error estimator yields an estimate which is close to the empirical standard error estimate and thus our approach can be considered as an alternative to resampling methods currently used in practice for inference on point pattern data using Gibbs process models.

4.4 Data Analysis

We apply our methods to the baseline data described in Ramseier et al. (2009). We are interested in studying the overall association between BL and CAL and also how the association varies according to different types of teeth. Recall that we discarded the CAL values corresponding to the buccal and lingual sites of each tooth, and averaged the BL and CAL values on the distal and mesial sites of each tooth to compute tooth-averaged values. Overall, we have data on 99 subjects and 16 teeth (bicuspids and molars) which correspond to 1,364 observations. A scatter plot of the data is shown in Figure 4.2.

In order to compute the overall association between CAL and BL, we need to estimate the irregular parameter r as well as the intensity and interaction parameters. We used profile pseudolikelihood to estimate the interaction distance r. For each subject, we considered a range of possible values of r from 0 to 1 with increments of 0.1. For each candidate value of r, we fitted the point process model and computed the log-pseudolikelihood value. We chose the value of r with the largest value of the log-pseudolikelihood and used that value in determining the covariate D corresponding to the points for that subject.

We used the same model as that used in Section 4.3.1. Based on the fitted model, the estimated parameters are $\hat{\alpha_0} = -4.93$ (SE= 0.70) and $\hat{\alpha_1} = 5.96$ (SE= 0.83). Thus, the estimated values of the intensity and interaction parameters are $\exp(-4.93) = 0.01$ and $\exp(5.96) = 387.61$ respectively. We can infer that the overall intensity of the points is low but there is a very strong interpoint interaction between

the points and hence the points tend to be clustered together as supported in Figure 4.2. Therefore we expect one point to be closer to another point or in other words, given a tooth with a particular BL and CAL value, we would expect another tooth to have a similar BL and CAL value.

Furthermore, we are interested in studying the intensity and interaction between BL and CAL according to different types of teeth. Thus, we mimic the approach in Section 4.3.2, using the marginal model $\log(\lambda_{kj}) = \log(\lambda(u_{kj}, l); \boldsymbol{v_k})) = \alpha_0^m + \alpha_1^m I_2(l) - \alpha_2^m D_{1kj} I_1(l) - \alpha_3^m (D_{1kj} I_2(l) + D_{2kj} I_1(l)) - \alpha_4^m D_{2kj} I_2(l)$

where $\boldsymbol{\alpha^m} = (\alpha_0^m, \alpha_1^m, \alpha_2^m, \alpha_3^m, \alpha_4^m)$ are the parameters of interest. The results are shown in Table 4.3.

Table 4.3 Table showing the estimated parameters and corresponding standard error estimates assuming the points, obtained from tooth-level averages of BL and CAL measurements in the study of Ramseier et al. (2009), to be realizations of a two-type Area Interaction process corresponding to observations for two types of teeth.

	Estimate	Robust se
$\hat{\alpha_0^m}$	-4.81	0.61
$\hat{\alpha_1^m}$	-0.53	0.77
$\hat{\alpha_2^m}$	3.69	0.69
$\hat{\alpha_3^m}$	2.00	0.84
$\hat{\alpha_4^m}$	4.48	0.84

Thus the estimate of the intensity parameters for the points corresponding to molars (i.e. points of type=1) and bicuspids (i.e. points of type=2) are $\exp(-4.81) = 0.01$ and $\exp(-4.81 - 0.53) = 0.01$, respectively. Note that $\exp(3.69) = 40.04$ and $\exp(4.48) = 88.23$ quantifies the association between the clinical measures CAL and BL in the molars and bicuspids respectively. This signifies that there is a strong interpoint attraction and clustering within each type of tooth so that values of CAL and BL tend to be similar within the molars and within bicuspids separately. Also, the association between CAL and BL is slightly stronger in bicuspids than in molars. There is a high interpoint attraction between the molars and bicuspids, indicated by

$\exp(2.00) = 7.39$. Thus given the BL and CAL values of one molar we would expect another bicuspid to have similar BL and CAL values and vice versa. This might be expected as the molars and bicuspids, though different in terms of morphology (molars are multi-rooted while bicuspids usually have single roots), are similar in function and also adjacent to each other in the mouth. Since we found that the BL and CAL values of any two molars and bicuspids are similar to one another, this would suggest that we can take measurements from any molar or bicuspid and consider it to be representative of the measurements from any other tooth of the same type. Such findings would be appealing to periodontists, as fewer measures might be taken during a periodontal exam thereby reducing the time and manpower involved in conducting periodontal exams.

4.5 Discussion

Our motivating dataset, described in Ramseier et al. (2009), does not have measurements on cuspids and incisors, and we focused our inference to molars and bicuspids. However, our methods can be readily extended to accommodate data on other types of teeth and similar statistical models can be easily implemented. Furthermore, our methods can be applied to other research areas involving marked point pattern data like modeling the dependence between two species of trees in mixed forest stands considered in Eckel et al. (2009) or analyzing the dependence between tree locations and measures of tree sizes as studied in Schlather et al. (2004). The methods for obtaining parameter estimates using maximum pseudolikelihood approach are implemented in the R package "spatstat" (Baddeley and Turner, 2005) and thus are appealing to non-statisticians. However, the package does not contain any readily available method to obtain valid standard errors appropriate for our setting. Via our methods we propose a variance estimator which can be easily applied using the R package "gee" (Carey et al., 2011) or "geepack" (Hojsgaard et al., 2006) and thus

might be of interest for non-statisticians too. Our proposed variance estimator is computationally faster in comparison to bootstrapping and other resampling techniques proposed in existing literature.

For our research problem of understanding the association between BL and CAL, various other statistical methods can be implemented. The easiest approach would be to calculate the correlation coefficient between CAL and BL and calculate a confidence interval based on the bootstrap standard error estimate. However, such an approach fails to model the association between the CAL and BL for different types of teeth. An alternative approach would be to use various copulas (Nelsen, 2006) to model the association and to determine the parameters which govern the association between the clinical measures and this presents an interesting research problem. We intend to pursue this interesting area for future research and study the correspondence between point pattern data methods and copulas.

CHAPTER V

Conclusion

We have proposed novel statistical approaches to model periodontal data and our methods provide an improvement to the current state of art in periodontal literature. In Chapters II and III, we studied the distribution of affected teeth in the mouth and quantified the locations of the mouth that are most susceptible to periodontal disease. In contrast to previously published studies, we were able to base our inference on correlated observations from multiple teeth of subjects and our methods did not require any prior knowledge about which specific mouth locations to be included. In Chapter II, via a regression approach we were able to quantify the mean direction of affected teeth i.e. determine the location of the mouth which is most prone to periodontal disease. We modeled the directions of affected teeth as a function of various patient-level characteristics and obtained valid standard error estimates using a bias corrected variance estimator which is shown to perform well for moderate sample sizes common to periodontal studies. In Chapter III, we generalized the methods in Chapter II and studied the distribution and number of modes of affected teeth assuming Generalized von Mises (GvM) distribution. In our motivating dataset, we were able to detect three most periodontally susceptible regions: the upper right corner, the upper left corner, and the middle of the lower jaw. This supports a cursory exam of the data and corroborates the conjecture that there is a symmetry among the left and right sides of the mouth with respect to the distribution of periodontal disease.

To the best of our knowledge, ours is the first attempt to specifically model correlated periodontal outcomes using circular statistical methods and thus our methods also present an interesting application of circular statistics.

Modeling of marginal multimodal circular outcomes as a function of covariates presents an interesting future research problem. Such a methodology would not only help researchers quantify the number and location of modes but also help assess the importance of any patient characteristic eg. pathogen level, smoking status etc. in determining their effect on the distribution of affected teeth. However, this would necessitate developing a regression approach for GvM distributions. It should be noted here that though the von Mises family belongs to the class of dispersion models, the Generalized von Mises distribution does not belong to this family. Hence developing a circular regression model for marginal Generalized von Mises outcomes presents an interesting research problem to statisticians and can be applied to model various kinds of multimodal angular observations.

In Chapter IV we used spatial point pattern data analysis methods in order to model the association between CAL and BL. We extended existing methods for Gibbs point process models by proposing a standard error estimator which can be used as an alternative to bootstrapping and other resampling techniques. This would appreciably reduce computational time and is also relatively easy to implement using standard statistical softwares. Our methods show that the BL and CAL values are similar for molars and bicuspids, which imply that the clinical measures obtained on any tooth can be considered to be representative of the measures obtained from any molar or bicuspid. This would imply that periodontists could take measures on any specific tooth and consider them to be representative of measures obtained from that tooth type. Thus periodontists may not need to take measurements from all the molars and bicuspids, thereby reducing the time needed for a periodontal exam.

Our methods have relied on the assumption that missing observations are missing

completely at random (MCAR) or missing at random (MAR). Fortunately, in our motivating dataset, subjects had few missing teeth, and of the few teeth for which the measures were missing, there was no discernible pattern in the missingness at baseline or at six months. Also, in our data, there was no tooth that was diseased at baseline and missing at six months. However, it is likely that subjects with existing periodontal disease may have had some teeth removed due to chronic periodontal disease. This is referred to as informative missingness (Reich and Bandyopadhyay, 2010). Our methods need to be extended to be applicable in such a scenario. In such a situation, a complication is that missing teeth are not missing at random (MAR), which a necessary assumption for the validity of the robust standard errors produced by marginal approaches such as GEE (Kenward and Molenberghs, 1998). Thus, the extension and application of our methods to periodontal data with greater numbers of non-randomly missing teeth presents a future area of research and such an extension would make our methods more readily applicable to most periodontal studies.

BIBLIOGRAPHY

BIBLIOGRAPHY

M. Abramowitz and I. A. Stegun. Handbook of mathematical functions. *New York: Dover*, 1965.

H. Akaike. Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*, pages 267–281, 1973.

M. Arora, J. Weuve, J. Schwartz, and R. O. Wright. Association of Environmental Cadmium Exposure with Periodontal Disease in U.S. Adults. *Environ Health Perspect*, 117:739–744, 2009.

R. Artes, G. Paula, and R. Ranvaud. Analysis of circular longitudinal data based on generalized estimating equations. *Australian and New Zealand Journal of Statistics*, 42:347–358, 2000.

A. Baddeley and R. Turner. Practical Maximum Pseudolikelihood for Spatial Point Patterns. *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.

A. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005.

A. J. Baddeley and M. N. M. Van Lieshout. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47:601–619, 1995.

S. Banerjee, B. P. Carlin, and A. Gelfand. Hierarchical modeling and analysis for spatial data. *Boca Raton: Chapman and Hall / CRC*, pages 23–74, 2004.

J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1977.

D. D. Boos. Risk factors for periodontitis. *The American Statistician*, 46:327–333, 1992.

D. A. Braunholtz, S. J. L. Edwards, and R. J. Lilford. Are randomized clinical trials good for us (in the short term)? Evidence for a trial effect. *Journal of Clinical Epidemiology*, 54:217–224, 2001.

L. Brown and H. Loe. Prevalence, extent, severity and progression of periodontal disease. *Periodontology 2000*, 2:57–71, 1993.

V. J. Carey, T. Lumley, and B. Ripley. Gee: Generalized estimation equation solver. 2011. URL `http://CRAN.R-project.org/package=gee`. R package version 4.13-17.

I. B. Darby, A. Polster, G. J. S., Q. Guo, N. Henein, A. Heredia, H. Horina, D. Sanduja, and M. Radvar. Left-to-right distribution of periodontal disease. *International Journal of Dental Hygiene*, 10:74–79, 2012.

B. A. Dye, L. K. Barker, R. H. Selwitz, B. G. Lewis, T. Wu, C. D. Fryar, Y. Ostchega, E. D. Beltran, and E. Ley. Overview and quality assurance for the National Health and Nutrition Examination Survey (NHANES) oral health component, 1999-2002. *Community Dentistry and Oral Epidemiology*, 35:140–151, 2007a.

B. A. Dye, S. Tan, V. Smith, B. G. Lewis, L. K. Barker, G. Thornton-Evans, P. I. Eke, E. D. Beltrn-Aguilar, A. M. Horowitz, and C. H. Li. Trends in oral health status, United States, 1988-1994 and 1999-2004. *Vital and Health Statistics*, 11: 1–92, 2007b.

S. Eckel, F. Fleischer, P. Grabarnik, M. Kazda, A. Sarkka, and V. Schmidt. Modelling tree roots in mixed forest stands by inhomogeneous marked Gibbs point processes. *Biometrical Journal*, 51:522–39, 2009.

M. P. Fay and B. I. Graubard. Small-Sample Adjustments for Wald-type Tests using Sandwich Estimators. *Biometrics*, 57:1198–1206, 2001.

N. Fisher and A. Lee. Regression models for an angular response. *Biometrics*, 48: 665–677, 1992.

N. I. Fisher. Statistical analysis of circular data. *Cambridge University Press*, 1993.

R. Gatto. Some computational aspects of the generalized von Mises distribution. *Statistics and Computing*, 18:321–331, 2008.

R. Gatto and S. R. Jammalamadaka. The generalized von Mises distribution. *Statistical Methodology*, 4:341–353, 2007.

A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp. Handbook of Spatial Statistics. *Boca Raton: Chapman and Hall / CRC*, 2010.

R. J. Genco. Current View of Risk Factors for Periodontal Diseases. *Journal of Periodontology*, pages 1041–1049., 1996.

C. J. Geyer. Likelihood Inference for Spatial Point Processes. *London: Hall/ CRC*, pages 141–172, 1999.

A. T. Gillthorpe, M. S. v Zamzuri, G. S. Griffiths, I. H. Maddick, K. A. Eaton, and N. Johnson. Unification of the "Burst" and "linear" Theories of Periodontal Disease Progression: A Multilevel Manifestation of the Same Phenomenon. *Journal of Dental Research*, 82:200–205, 2003.

J. M. Goodson, A. D. Haffajee, and S. S. S. The relationship between attachment level loss and alveolar bone loss. *Journal of Clinical Periodontology*, 11:348–359, 1984.

E. Hausmann, K. Allen, J. Norderyd, W. Ren, O. Shibly, and E. Machtei. Studies on the relationship between changes in radiographic bone height and probing attachment. *Journal of Clinical Periodontology*, 21:128–132, 1994.

T. Hirotomi, A. Yoshihara, H. Ogawa, and H. Miyazaki. Tooth-related risk factors for periodontal disease in community-dwelling elderly people. *Journal of Clinical Periodontology*, 37:494–500, 2010.

E. B. Hoffman, P. K. Sen, and C. R. Weinberg. Within-cluster resampling. *Biometrika*, 88:1121–1134, 2001.

S. Hojsgaard, U. Halekoh, and J. Yan. The R Package geepack for Generalized Estimating Equations. *Journal of Statistical Software*, 15:1–11, 2006.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. Statistical Analysis and Modelling of Spatial Point Patterns. *John Wiley and Sons, Chichester*, 2008.

S. R. Jammalamadaka and A. SenGupta. Topics in Circular Statistics. *Singapore: World Scientific Press*, 2001.

M. K. Jeffcoat. Radiographic methods for the detection of progressive alveolar bone loss. *Journal of Periodontology*, 63:367–372, 1992.

B. Jorgensen. The theory of dispersion models. *London: Chapman and Hall*, 1997.

G. Kauermann and R. Carroll. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96:1387–1396, 2001.

M. Kenward and G. Molenberghs. Likelihood based frequentist inference when data are missing at random. *Statistical Science*, 13:236–47, 1998.

J. S. Kinney, T. Morelli, T. Braun, C. A. Ramseier, A. E. Herr, J. V. Sugai, C. A. Shelburne, L. A. Rayburn, A. K. Singh, and W. V. Giannobile. Saliva/Pathogen Biomarker Signatures and Periodontal Disease Progression. *Journal of Dental Research*, 90:752–758, 2011.

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

S. R. Lipsitz, G. M. Fitzmaurice, E. J. Orav, and N. M. Laird. Performance of generalized estimating equations in practical situations. *Biometrics*, 50:270–278, 1994.

H. Loe and L. Brown. Early onset periodontitis in the United States of America. *Journal of Periodontology*, 62:608–616, 1991.

H. Loe, A. Anerud, H. Boysen, and M. Smith. The Natural History of Periodontal Disease in Man. The Rate of Periodontal Destruction Before 40 Years of Age. *Journal of Periodontology*, 49:607–620, 1978.

B. Lu, J. S. Preisser, B. F. Qaqish, C. Suchindran, S. I. Bangdiwala, and M. Wolfson. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics*, 63:935–941, 2007.

V. M. Maksimov. Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups. *Theoria Verojatna*, 12:307–321, 1967.

L. Mancl and T. DeRouen. A covariance estimator for GEE with improved small-sample properties. *Biometrics*, 57:126–134, 2001.

K. V. Mardia. Statistics of directional data. *London: Academic Press*, 1972.

J. Moller and R. Waagepeterson. Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, 34:643–684, 2007.

A. Mombelli and C. Meier. On the symmetry of periodontal disease. *Journal of Clinical Periodontology*, 28:741–745, 2001.

R. B. Nelsen. An Introduction to Copulas. *Springer: Second Edition*, 2006.

M. C. Paik. Repeated measurement analysis for nonnormal data in small samples. *Communications in Statistics: Simulation and Computation*, 17:1155–1171, 1998.

W. Pan. Akaike's information criterion in generalized estimating equations. *Biometrics*, 57:120–125, 2001.

F. Papangelou. The conditional intensity of general point processes and an application to line processes. *Zeitschrift fur Wahscheinlichkeitstheorie und werwandte Gebiete*, 28:207–226, 1974.

C. A. Ramseier, J. S. Kinney, A. E. Herr, T. Braun, J. V. Sugai, C. A. Shelburne, L. A. Rayburn, H. M. Tran, A. K. Singh, and W. V. Giannobile. Identification of pathogen and host-response markers correlated with periodontal disease. *Journal of Periodontology*, 80:436–446, 2009.

B. J. Reich and D. Bandyopadhyay. A latent factor model for spatial data with informative missingness. *The Annals of Applied Statistics*, 4:439–459, 2010.

B. J. Reich and J. S. Hodges. Modeling longitudinal spatial periodontal data: A spatially adaptive model with tools for specifying priors and checking fit. *Biometrics*, 64:790–799, 2008.

B. J. Reich, J. S. Hodges, and B. P. Carlin. Spatial analyses of periodontal data using conditionally autoregressive priors having two classes of neighbor relations. *Journal of the American Statistical Association*, 102:44–55, 2007.

S. Renvert, A. Badersten, R. Niiveus, and J. Egelberg. Healing after treatment of periodontal intraosseous defects. I. Comparative study of clinical methods. *Journal of Clinieal Periodontology*, 8:387–399, 1981.

G. R. Riviere, K. S. Smith, N. Carranza Jr., E. Tzagaroulaki, K. S. Kay, and M. Dock. Subgingival distribution of Treponema denticola, Treponema socranskii, and pathogen-related oral spirochetes: prevalence and relationship to periodontal status of sampled sites. *Journal of Periodontology*, 66:829–837, 1995.

M. Schlather, P. J. Ribeiro Jr, and P. J. Diggle. Detecting Dependence between Marks and Locations of Marked Point Processes. *Collegium Antropologicum*, 66: 79–93, 2004.

S. S. Socransky, A. D. Haffajee, M. A. Cugini, C. Smith, and R. L. Kent. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, 25:134–144, 1998.

P. X. K. Song. Correlated Data Analysis: Modeling, Analytics and Applications. *Springer*, 2007.

S. Spalj and D. Plancak. The distribution of periodontal disease and loss of attachment in jaw sextants in different age groups-cross sectional study. *Collegium Antropologicum*, 27:183–190, 2003.

D. J. Strauss. A model for clustering. *Biometrika*, 63:467–475, 1975.

W. M. Thomson, R. Hashim, and A. R. C. Pack. The Prevalence and Intraoral Distribution of Periodontal Attachment Loss in a Birth Cohort of 26-Year-Olds. *Journal of Periodontology*, 71:1840–1845, 2000.

M. F. Timmerman and G. A. Van der Weijden. Risk factors for periodontitis. *International Journal of Dental Hygiene*, 4:2–7, 2006.

G. Tiwari. Biodegradable Microspheres For Controlled Delivery Of Metronidazole In The Treatment Of Periodontal Diseases: Formulation consideration. *International Journal of Pharma and Bio Sciences*, 1:1–13, 2010.

C. Tomasi, A. H. Leyland, and J. L. Wennstrom. Factors influencing the outcome of non-surgical periodontal treatment: a multi-level approach. *Journal of Clinical Periodontology*, 34:682–690, 2007.

K. Torrungruang, V. Gongsakdi, L. Laohaviraphab, K. Likittanasombat, and W. Ratanachaiwong. Association between cigarette smoking and the intraoral distribution of periodontal disease in Thai men over 50 years of age. *Journal of Investigative and Clinical Dentistry*, 3:1–7, 2011.

H. White. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50:1–25, 1982.

B. Widom and J. S. Rowlinson. New model for the study of liquid-vapor phase transitions. *The Journal of Chemical Physics*, 52:1670–1684, 1970.

C. B. Wiebe and E. E. Putnins. The periodontal disease classification system of the American Academy of Periodontology– An update. *Journal of the Canadian Dental Association*, 66:594–597, 2000.

E. A. Yfantis and L. E. Borgman. An extension of the von Mises distribution. *Communications in Statistics, Theory and Methods*, 11:1695–1706, 1982.

Y. Zhang, D. Todem, K. Kim, and E. Lesaffre. Bayesian latent variable models for spatially correlated tooth-level binary data in caries research. *Statistical Modeling*, 11:25–47, 2011.