STATISTICAL METHODS FOR REPEATED MEASURES IN EXPERIMENTAL GINGIVITIS WITH ADJUSTMENT FOR LEFT TRUNCATION DUE TO LOWER DETECTION LIMITS

Kelley D. Wekheye

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics, Gillings School of Global Public Health.

Chapel Hill
2014

Approved by:

John S. Preisser

Lloyd J. Edwards

Anastasia Ivanova

Steven Offenbacher

Pranab K. Sen

ABSTRACT

KELLEY D. WEKHEYE: Statistical Methods For Repeated Measures In Experimental Gingivitis With Adjustment For Left Truncation Due to Lower Detection Limits
(Under the direction of John S. Preisser)

In the characterization of biomarkers measured repeatedly over time, there is a need to summarize the information contained in the multivariate data. In experimental gingivitis (EG), for example, biomarker levels change when the benefits of toothbrushing are withheld during an induction phase, then restored during a resolution phase. The pattern of change over time of biomarker levels associated with gingivitis could reflect change in various directions; therefore, the statistical methodology utilized should consider this possibility. As such, area under the curve (AUC) can be implemented as a summary measure for estimating change in biomarker levels. Parametric statistical models for repeated measures analysis are useful for characterizing the nature of that change over time, particularly as they easily accommodate both truncated and missing data. In EG studies, left truncation results when a biomarker level falls below the lower limit of detection. We propose two parametric approaches to provide direct estimation of the trends in biomarkers over time while implementing adjustments for left truncation. The focus is on estimation and hypothesis testing for AUC.

The first paper derives a piecewise linear random-effects regression model fit to 3 biomarkers representing varying degrees of missingness due to lower detection limits using 2 *ad hoc* (naive) approaches for handling non-detect values and a likelihood approach accounting for left censoring (Lyles, Lyles and Taylor, 2000). These *naive* approaches replace non-detect biomarker values by the limit of detection and half that

limit, which may result in bias, while the maximum likelihood method gives valid results when dropouts are missing at random.

The second paper outlines AUC methodology for repeated measures biomarker data by using a nonlinear "Gamma Curve" mixed model with adjustment for left truncation based on a maximum likelihood approach in comparison to the *ad hoc* approaches outlined in the first paper.

The third paper presents a simulation study that includes methods from the first two papers as well as Wilcoxon Sign Rank test methods from Preisser, Sen, and Offenbacher (2011). The simulation design, motivated by EG studies, focuses on properties of hypothesis tests (size and power) in the presence of left truncation and/or missing data to evaluate whether the parametric methods are reliable for small sample sizes or whether larger samples are needed to reliably use the methods. Evidence for recommending certain sample sizes for EG studies and an evaluation of whether the nonparametric method is robust to left truncation and crude single imputation methods are also provided.

The proposed methodology is illustrated using longitudinal data from an EG study whereby the benefits of toothbrushing are temporarily withheld, then restored (Offenbacher et al, 2010).

To my dad, Nathaniel Peterson, for his love and dedication shown in raising two little girls on his own after the passing of his wife. To my mom, Geraldine Peterson, in loving memory. Her presence made me the woman I am today, motivated me to strive through the program, and continue to encourage me to reach my highest goals.

ACKNOWLEDGMENTS

First, all praise, honor, and glory to my Heavenly Father for making this degree even possible. Secondly, I am particularly thankful and grateful for my dear husband Chrisantos for his continuous encouraging words, love, support, and extreme patience during this process. We have gone through a lot together during this process and have come out victorious.

I would like to express my sincere gratitude to my academic and dissertation advisor, Dr. John S. Preisser. Without his guidance, it would have been challenging to complete the DrPH program and to write my dissertation. His support and feedback have been invaluable. Dr. Preisser has always been available to help and provided me with insightful references and ideas critical for this dissertation. He graciously reviewed several versions of the papers, and his suggestions have substantially contributed to the quality of my work.

I am very appreciative of the tremendous support of Dr. Lloyd Edwards. His superior knowledge of substantive issues concerning repeated measures analysis provided a unique perspective to the research problem and helped immensely in the interpretation of the results.

I am very grateful to Dr. Anastasia Ivanova for her graciousness and helpful comments during this dissertation process.

My heartfelt gratitude goes also to Dr. Steven Offenbacher. Dr. Offenbacher was kind enough to meet with me and Dr. Preisser early on in the research process to discuss the topic. He has provided me with insightful information critical for understanding the clinical and biological aspects of experimental gingivitis, the subject matter of this

dissertation. I would also like to thank Dr. Silvana Barros for providing insightful comments to my first paper.

I would also like to communicate my sincere thanks to Dr. P.K. Sen whose suggestions have helped strengthen the statistical methodology used in this dissertation. Dr. Sen was kind enough to recommend revisions that greatly enhance the quality of the manuscript and maintain scientific rigor.

This dissertation would also not have been possible without the prayers and support of numerous family members and friends. I would like to thank my sister Courtney Alexander, brother-in-law Cory Alexander, and my niece and nephews for their support, laughter, and sending gifts during this process. I would also like to thank my best friends Kyla Kurian (and family) and Rachael Redmond for their prayers, support, encouragement, and for helping select my final defense attire. You are truly like sisters to me. Thanks to my two other girlfriends/sisters Elexica McAlister (and family) and Julie Omishakin (and family) for their prayers and support. Thanks also go to Leon and Mashall Bullard (and family) for their endless support and encouragement. Special thanks to Isaiah Redmond and Christopher Paine for always making it a goal of theirs to keep me laughing whenever I see them.

Lastly, I would like to thank the countless other people who have supported me in various aspects during the completion of this dissertation, particularly the Kenyan community who have accepted me as part of the Kenyan family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA analysis of variance

AS asymmetrically sensitive

AUC area-under-the curve

CI confidence interval

EG experimental gingivitis

EM expectation maximization

FDR false discovery rate

FWER family-wise error rate

GCF gingival crevicular fluid

IS insensitive

LOD limit of detection

MAR missing at random

MMP matrixmetalloproteinase

NS negatively sensitive

OLS ordinary least squares

PROP proposed influence function

PS positively sensitive

RCM random coefficients model

SE standard error

CHAPTER 1: INTRODUCTION

In the progression of periodontal disease, molecular mediators of inflammation are often measured at multiple locations and taken repeatedly over time. The diagnosis of periodontal disease can be monitored through the concentration of microbial and host products in the gingival crevicular fluid (GCF). Specifically, in studies of experimental gingivitis experimental gingivitis (EG), samples of GCF are collected at multiple sites over time to assess changes in biofilm overgrowth and oral inflammation, potentially aiding in early recognition of periodontal disease susceptibility.

In research involving repeated measures of biomarker levels, there is a need to derive measurements that summarize the information contained in the multivariate data. Additionally, because the pattern of change over time could reflect change in many directions, the statistical methodology utilized should accommodate this possibility. As such, area-under-the curve (AUC) can be implemented as a summary measure for estimating change in biomarker levels. Parametric statistical methods for repeated measures analysis are often employed to determine whether outcomes exhibit change in their levels over time as well as to characterize the nature of that change. However, the performance characteristics of the statistical approaches have not been adequately assessed in these settings with respect to estimation of AUC and hypothesis testing, particularly in the presence of left truncation and missing data.

In the assessment of change over time, missing data can be a common occurrence when data are measured over multiple periods of time. For longitudinal studies involving repeated measures analysis, there can be many reasons for missing data, including

nonresponse, subject dropout, or as is often times the case in the measurement of biomarker data, missingness due to assay detection limits. Missing values, as well as outliers, can have a profound influence on statistical results, including estimation of summary measures of change and hypothesis testing. If the missingness mechanism is missing at random (MAR), i.e., the probability that a response is observed can only depend on the values of those other factors which have been observed, there are well developed computational methods for handling missing data under this assumption (Little and Rubin, 1987). Hughes (1999) described an EM algorithm for maximum likelihood estimation of a linear mixed effects model for estimating trends in CD4 counts over time in HIV-positive subjects, accounting for left and/or right censoring. Lyles et al. (2000) developed a likelihood method that addresses missing data due to left truncation as well as an extension to additionally accommodate informative dropout. Thiebaut and Jacqmin-Gadda (2004) applied a maximum likelihood approach for left-censored data based on a Marquardt algorithm (Marquardt, 1963) in HIV research. In reference to pharmacokinetic data with measurements below the quantification limit, Fang et al. (2011) developed a maximum likelihood method to estimate AUC and the ratio of two AUCs (i.e., relative exposure).

In dental research, because studies of inflammatory mediators tend to involve only a small to moderate number of subjects, parametric methods, which are particularly sensitive to outliers and deviations from Gaussian assumptions, may not be the preferred approach to analyze such data. Alternatively, nonparametric methods have been utilized in some settings due to the reliance on fewer assumptions and presumed greater robustness over their parametric counterparts. For the analysis of biomarker data in experimental gingivitis, most recently, a nonparametric multiple hypothesis testing approach was advocated for the analysis of repeated measures (Preisser et al., 2011) using AUC summary measures to assess the change over time in biomarker levels. Conse-

quently, a method to identify biomarkers using univariate and multivariate Wilcoxon signed rank tests for a set of four summary measures based upon AUC has been previously described (Preisser et al., 2011). A limitation of this approach is that it is a hypothesis testing approach; therefore, it does not provide estimation of the location (mean or median) response pattern over time. However, nonparametric procedures may improve many of the problems encountered with parametric methods and are more flexible in dealing with situations in which the number of biomarkers exceeds the number of subjects.

The purpose of this dissertation is to study statistical methods applicable to EG data. We propose to develop parametric mixed models accounting for left truncation under MAR using the estimation methods that are easy to implement. The parametric models will be fit to log-transformed data for 3 biomarkers representing varying degrees of truncation due to lower detection limits using 2 *ad hoc* (naive) approaches for handling non-detect values and a likelihood approach accounting for left censoring and outcomes missing at random ("ML1") from Lyles et al. (2000). The focus will be on providing direct estimation of the trends in biomarkers over time, calculations of AUC based on the associated parameters from Preisser et al. (2011), and hypothesis testing for AUC summaries in a study of EG. The proposed methodology will be illustrated using longitudinal data from an EG study previously conducted (Offenbacher et al., 2010).

CHAPTER 2: LITERATURE REVIEW

2.1 Overview of Gingivitis

Periodontitis is a chronic disease which results in destruction of the periodontal ligament and alveolar bone supporting a tooth, and which may eventually lead to tooth loss (DeRouen et al., 1995). Gingivitis, the mildest form of periodontal disease and a condition that can advance to periodontitis if left untreated, is often caused by inadequate oral hygiene, which leads to plaque buildup (American Academy of Periodontology 2010). The reversible process of gingivitis can be resolved with plaque removal from the tooth surface (Salvi et al., 2010). The inflammatory host response, a core component of periodontal disease, is thought to be the immediate cause of periodontal breakdown (Deinzer et al., 2007). Experimental Gingivitis (EG), first developed by Loe et al. (1965), is recognized as a well-controlled condition for the clinical investigation of gingivitis. In this study design, gingivitis is induced in healthy patients by stopping oral hygiene practices (Deinzer et al., 2007; Loe et al., 1965). Typically EG study designs involve a hygiene phase (ie, establishment of gingival health), an induction phase (i.e., neglect of gingival health), and a resolution phase (ie, re-establishment of gingival health). Many researchers have focused their attention on employing this study design to realize a better understanding of the host's immune response to periodontal pathogens (Deinzer et al., 2007). Therefore, the experimental gingivitis framework is a well-utilized analytical structure for elucidating the inflammatory response to undisturbed dental biofilm accumulation.

Characterization of chronic gingivitis is traditionally based on the presence of bleed-

ing, oedema, redness, and an increased flow of GCF (Salvi et al., 2010). The GCF is a serum transudate that is enriched with microbial and host products that arise as a result of the current inflammatory dynamics of the host-biofilm interaction (Offenbacher et al., 2010). The biochemical analysis of the fluid offers a non-invasive means of assessing the host response in periodontal disease. The active phase of the periodontal disease process can be assessed by the components of gingival fluid (Subrahmanyam and Sangeetha, 2003). Because it contains elevated levels of a vast array of biochemical factors, gingival crevicular fluid is a more attractive clinical marker of periodontal disease activity over more traditional methods. As such, some studies have shown that chronic gingivitis is associated with higher levels of inflammatory mediators as identified in GCF (Offenbacher et al., 2007). We will review the design and analysis issues in the experimental gingivitis studies using GCF as a means of assessing periodontal status.

2.2 Overview of the Experimental Gingivitis Study

Preisser et al. (2011) describe a study based on data previously published by Offenbacher et al. (2010) in which thirty-one inflammatory mediators within each GCF sample, including cytokines, matrixmetalloproteinases (MMPs) and adipokines were studied in 22 subjects to evaluate the changes in the GCF composition over time. The study recruited and enrolled subjects with naturally occurring gingivitis, defined as bleeding upon probing present, typically in at least 10% of dental sites, as these subjects were more likely to develop experimental gingivitis in the course of the study. The course of the experiment included a 1-week hygiene phase, a 3-week induction phase using two stents and a 4-week resolution phase. Gingivitis was induced by withholding tooth brushing by the use of intraoral acrylic stents that cover selected teeth in each arch during tooth brushing to induce local gingival inflammation. Mediator levels were

determined from the laboratory analysis of GCF. At the end of the induction phase, stents were removed and hygiene on all teeth was restored to resolve inflammation. Gingival crevicular fluid was collected from the same oral sites at the beginning of the hygiene phase (or Day -7, one week prior to baseline), weekly during the induction phase (Day 0 or baseline), Day 7, 14, 21 (end of the induction phase/baseline for the resolution phase), and biweekly during the resolution phase at Day 35 and 49. At the final time point, baseline levels were expected to be restored for all biomarkers.

At each time point, gingival crevicular fluid was collected from eight dental sites from the stent teeth and the volume of fluid collected from each sample was recorded. The average of the two concentration measurements from each site was considered the measurement for the particular site and time point. The goal of the experiment was to identify new candidate biomarkers that were sensitive to poor oral health care as identified by their patterns of change during induction and resolution of gingivitis. The data have been previously analyzed using parametric linear mixed modeling (Offenbacher et al., 2010). Although not specified, the analysis treated left truncated values as zeros. Furthermore, no adjustments were made for multiple hypothesis testing.

2.3 Review of Design and Analysis Methods in Experimental Gingivitis Studies

In review of statistical analysis methods used in the EG studies, fourteen studies measuring clinical parameters, microbiological parameters, and biomarker concentrations in GCF during experimental gingivitis were evaluated. As our focus is on the characterization of biomarkers within GCF, the clinical and microbiological profiles are not described. The sample size ranged from 10 to 50 subjects, with the majority of studies recruiting approximately 20 subjects to serve as internal controls in the analyses or to be divided into independent groups for analysis. One to 2 weeks prior to study

initiation, subjects received professional tooth cleaning and were given oral hygiene instructions in order to maintain perfect gingival health through the Baseline (Day 0) visit. Following this standard hygiene phase, at specified teeth, subjects abstained from oral hygiene ranging from a period of 4 to 28 days. The most frequent timeframes for this no-hygiene, gingivitis phase were a period of 21 or 28 days. The length of the resolution phase was not indicated in most study designs; however, when it was specified, the timeframe for oral hygiene restoration was 28 days. Gingival crevicular fluid was usually sampled from mesiobuccal, mesiopalatal, distopalatal, and distobuccal sites. For most studies, crevicular fluid samples were collected weekly only during the EG period and analyzed for measurement of 2-3 biomarkers, on average, within each sample. Although the number of biomarkers studied ranged from 1 to 33, the most consistently measured biomarkers were interleukin cytokines IL-1β, IL-1α, and IL-1ra.

Although gingivitis occurs over time as a steady-state inflammatory response, only approximately one-half of studies used a repeated measures analysis of variance (ANOVA) approach (Deinzer et al., 2004, 2007; Johnson et al., 1997; Waschul et al., 2003), with few studies examining the rate of increase by calculating AUC (Jepsen et al., 2003; Preisser et al., 2011; Salvi et al., 2010). Most studies that took a nonparametric approach to the analysis used a Wilcoxon-signed rank test for intra-subject comparisons and Mann-Whitney U-test or Wilcoxon rank-sum test to assess between-group differences (Giannopoulou et al., 2003; Konradsson et al., 2007; Konradsson and van Dijken, 2005; Staab et al., 2009; Tsalikis, 2010). Studies not using repeated measures analysis instead used paired t-tests to assess mean within group changes from baseline to each timepoint or 2-sample t-tests to assess between-group differences in mean levels at each timepoint (Konradsson et al., 2007; Salvi et al., 2010). Few studies mentioned using a log-transformation, or any other form of transformation, before analyzing the data. Additionally, most studies mentioned above did not mention or address the assay

lower limit of detection (LOD) and how values below the limit were handled in the analysis. Studies that did mention the LOD did not indicate how values left-censored due to being below the detection limit were addressed, if at all (Deinzer et al., 2007; Johnson et al., 1997; Konradsson et al., 2007). One study that examined the rate of development of gingivitis from Baseline to the end of induction phase dichotomized AUC at the mean and used a binary logistic regression analysis to identify a model to explain progression and severity of gingivitis. Another study used univariate and multivariate Wilcoxon-signed rank tests based on 4 AUC summary measures to assess the change over time in biomarker levels Preisser et al. (2011). To show increasing trend in biomarkers, one study used the large sample approximation Friedman test.

The studies that used a parametric repeated measures approach used an ANOVA model. Three studies used ANOVA to identify significant main effects and significant interactions with time. In these studies, to assess temporal stability of biomarkers, results were reported as Greenhouse-Geisser corrected values along with degrees of freedom, e and h^2 as indicators of effect size. These adjustments were made to correct for potential violations of sphericity. Studies not using repeated measures analysis instead used paired t-tests to assess mean within group changes from baseline to each timepoint or 2-sample t-tests to assess between-group differences in mean levels at each timepoint. Few studies mentioned using a log-transformation, or any other form of transformation, before analyzing the data.

Although majority of the studies assessed GCF mediator levels at multiple time-points, few studies took into account the multiple testing problems. Studies that did address this issue used the Bonferonni method. One study further indicated choice of control of family-wise error rate (FWER) or false discovery rate (FDR). Additionally, only 2 studies mentioned and addressed the assay lower limt of detection (LOD) and how values below the limit were handled in the analysis. Other studies that mentioned

the LOD did not indicate how values left-censored due to being below the detection limit were addressed, if at all. Overall, there still remain statistical issues that need to be considered when analyzing biomarker data in periodontal disease and uniformity in such analyses addressed.

2.4 Linear Mixed Models

Consider the matrix form of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.1}$$

where $\mathbf{y}$ is the response vector; $\mathbf{X}$ is the regression parameter design matrix, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{\epsilon} \sim N(0, \sigma^2)$ is the vector of errors. In this model, the relationship is described between a response variable and covariates that are measured at the same points in time as the response. Longitudinal studies are often employed to investigate the progression of a characteristic over time by taking repeated measurements within an observational unit. In some cases, the characteristic studied may exhibit differences at baseline as well as over time. If a characteristic varies linearly over time with intercept and/or slope varying between observational units, it may be more appropriate to model characteristic patterns from repeated measures using a random coefficients model (RCM). The RCM model is a two-stage model in which the mean response is modelled as a combination of fixed (population) and random (subject-specific) effects. The general form of the model can be expressed as

$$\begin{aligned} \mathbf{y_i} &= \mathbf{X_i}\boldsymbol{\beta} + \mathbf{Z_i}\mathbf{b_i} + \boldsymbol{\epsilon}_i \\ \mathbf{b_i} &\sim N(0, \mathbf{D}) \\ \boldsymbol{\epsilon}_i &\sim N(0, \sigma^2) \end{aligned} \tag{2.2}$$

where

- $\mathbf{y_i}$ is the n_i x 1 vector of observations with $E(\mathbf{y_i}) = \mathbf{X_i}\boldsymbol{\beta}$
- $\mathbf{X_i}$ is the n_i x p design matrix for the fixed effects
- $\boldsymbol{\beta}$ the p x 1 vector of regression coefficients
- $\mathbf{Z_i}$ the n_i x q design matrix for the random effects
- $\mathbf{b_i}$ is the q x 1 vector of $i.i.d$ random effect coefficients
- $\boldsymbol{\epsilon}_i$ is the n_i x 1 vector of $i.i.d$ random error terms

Random coefficients models are commonly used in analysis of data in which a subject-specific linear relationship is assumed between the response variable and time. In situations in which there are curvilinear effects in the response, polynomial models could be used to describe nonlinear relationships, for example, in the progression of disease. Visual review of EG biomarker levels indicate that the levels of GCF can vary widely within subjects. The nonlinear relationship of GCF levels with time indicates that a model using polynomials of times may be considered to fully characterize disease profiles.

2.5 Methods For Left Truncated Data

One of the most commonly used and easily implemented methods reported in environmental scientific literature to deal with values below detection limits is to substitute a fraction of the detection limit for each nondetect. Nondetects are values known only to be somewhere between zero and the laboratory assay's detection limits. In essence, this method replaces a single, unknown value with a single value for the specified data. Any substitution of a constant fraction of reporting limits is thought to distort estimates

of the standard deviation, thereby negatively affecting all (parametric) hypothesis tests using that statistic, as well as obscure patterns and trends in the data (Helsel, 2006). As such, substituted values using a fraction anywhere between 0 and 0.99 times the detection limit are considered to be incorrect, possibly leading to inaccurate interpretation of the study results (Helsel, 2006). Based on simulation studies, Lubin et al. (2004) found that imputing one-half the detection limit for nondetect values can be biased if the percentage of measurements below the detection limit is greater than 10%. Survival and reliability analysis methods used for analyzing data without substituting values has been suggested as a better approach. For environmental data, survival analysis has been shown as a better method over traditional substitution of values such as one-half the detection (Baccarelli et al., 2005; Helsel, 2006). Thompson and Nelson (2003) extended the work of Aitken (1981) by developing a maximum likelihood (ML) approach to Type I left- and interval-censored data. Through simulation studies, they confirmed the bias in the simple substitution approach, where censored observations are replaced by the midpoint of their censoring interval, and illustrated the effect of increased censoring level on power to detect significant relationships. Although slow in progression, the effect on power with increasing censoring was shown to be substantial (Thompson and Nelson, 2003).

Lubin et al. (2004) reviewed other strategies for handling data with detection limits. The small percentage of values below the detection limit requirement for using substitution methods led to a single-impute "fill-in" approach in which the form of the distribution is characterized, parameters are estimated, and randomly sampled values below the detection limit are assigned from the estimated distribution (Helsel (1990); Moschandreas et al. (001a,b)). Although the randomly-assigned values are from an appropriate distribution, this approach produces biased variance estimates when at least 30% of data are below the detection limit. Tobit regression and multiple imputation

were offered as two unbiased approaches. Although it requires a large amount of data, if values are needed for measurements below the detection limit, multiple imputation was determined to be the optimal approach in the presence of nonignorable data unless the proportion of missingness was substantial. When interest is in regression parameters, tobit regression is suitable (Little and Rubin, 1987).

Several estimation procedures for Type I censored data have been developed for normal and lognormal populations (Cohen, 1950, 1959; Gilliom and Helsel, 1986; Gleit, 1985; Persson and Rootzen, 1977; Schneider, 1986). In the presence of potential outliers, Singh and Nocerino (2002) evaluated classical and robust parameter estimating procedures (Cohen, 1959; Dempster et al., 1977; Persson and Rootzen, 1977) in terms of bias and mean square error (MSE). Maximum likelihood estimation (MLE) uses both the uncensored observations and the proportion of data below one or more detection limits to compute statistics for the entire dataset in which the distribution of the data is known. Cohen's method (Cohen's MLE) is an adaptation of the MLE method which uses a lookup table to calculate estimates of the mean and standard deviation by adjusting the statistics of the uncensored observations as a function of the amount of censoring in the data. Dempster et al. (1977) expectation maximization (EM) method is an iterative approach that uses a conditional expectation maximizing function to replace non-detect values by the conditional expected value. Gilliom and Helsel (1986); Hashimoto and Trussell (1983); Helsel (1990) used ordinary least squares (OLS) regression to extrapolate non-detects from the regression model.

The performance of these methods (including restricted MLE [RMLE] and unbiased MLE [UMLE] methods depend on sample size, percentage of censoring, and the detection limit value. Based on simulations, the UMLE method may be used for large sample sizes (i.e., at least 15 observations) if the percentage of censoring is less than 30%. The RMLE and Cohen's MLE methods provided similar estimates and bias and MSE results

that were stable and in close agreement for all sample sizes and censoring proportions. However, because of it's simplicity, for population parameter estimation, the RMLE was recommended for large censoring intensities (i.e., greater than 30% or larger sample sizes. Substitution methods, including the EM method, should be avoided when the sample size exceeds 10 observations. In the presence of outliers, the OLS regression method based on log-transformed data gives unreliable results due to the distortion of the outliers. In this case, Singh (1993) proposed influence function (PROP) function is recommended for the population parameter estimation.

For longitudinal studies involving repeated measures analysis, Lyles et al. (2000) developed a likelihood method that addresses missing data due to left censoring and informative dropout, simultaneously. Under this missingness mechanism, traditional longitudinal data analysis methods, such as generalized estimating equations (Liang and Zeger, 1986) and random effects linear models (Laird and Ware, 1982) can produce biased population average intercept and slope estimates (Lyles et al., 2000). The method proposed by Lyles et al. (2000) is a combination of a likelihood-based adaptation of the EM algorithm (Hughes, 1999) and a combined log-normal (dropout process) and linear random-effects (repeated measures response) model (Schluchter, 1992). The approach directly works with the likelihood functions to derive standard errors based on the observed information matrix.

Lyles et al. (2000) also reviewed a linear random effects model proposed by Schluchter (1992) for repeated measures with left-truncated data valid under MAR dropout that is easy to program with standard statistical packages (e.g., SAS PROC NLMIXED).

2.6 Area-Under-the-Curve Principle

For the re-analysis of the biomarker data of Offenbacher et al. (2010), all biomarker data are transformed to the log scale (base 10). Let L_{ijt} be the log GCF level of the jth

biomarker for the ith subject at the tth time point, for $t = 0, \ldots, 5$, corresponding to Days 0, 7, 14, 21, 35 and 49, respectively. Ideally, each biomarker should fit into one of the three categories: (a) positively sensitive positively sensitive (PS) biomarker - L_{ij0}, L_{ij1}, L_{ij2}, and L_{ij3} are expected to increase over time during the induction (stent) phase from L_{ij0} to L_{ij3} and decrease from L_{ij3} to L_{ij5} during the resolution (non-stent) phase; (b) negatively sensitive (NS) biomarker - decreasing trend during induction followed by a increasing trend during the resolution phase; (c) asymmetrically sensitive (AS) biomarker - levels not only return to baseline after Day 21, but temporarily elevate above baseline; and (d) insensitive (IS) biomarkers - the GCF levels remain constant over the six time periods. Change over time for some biomarkers may not be consistent with one of these patterns.

Because there are multiple directions for the change in biomarker levels, Preisser et al. (2011) determined that the statistical methodology employed in the analysis should allow for detecting change away from the null in any of the directions. The AUC approximates the average change between two observed time points of the biomarker levels. For the jth biomarker level from the ith subject, define the change from baseline to time t as $Y_{ijt} = L_{ijt} - L_{ij0}$. Using week as the unit of time, let the summaries of AUC be denoted as follows (Preisser et al., 2011):

$$\begin{aligned}
A_{ij} &= (Y_{ij1} + 2Y_{ij2} + Y_{ij3})/2 && = \text{area between week 1 and week 3} \\
B_{ij} &= Y_{ij3} + Y_{ij4} && = \text{area between week 3 and week 5} \\
C_{ij} &= Y_{ij1}/2 && = \text{area between week 0 and week 1} \\
D_{ij} &= Y_{ij4} + Y_{ij5} && = \text{area between week 5 and week 7}
\end{aligned}$$

Next, four variates of interest to be assessed in the statistical analysis are defined as

follows:

$$\begin{aligned} X_{ij1} = &\quad C_{ij} - \tfrac{1}{2}D_{ij} &&= \frac{1}{2}(Y_{ij1} - Y_{ij4} - Y_{ij5}) \\ X_{ij2} = &\quad A_{ij} - B_{ij} &&= \frac{1}{2}(Y_{ij1} + 2Y_{ij2} - Y_{ij3} - 2Y_{ij4}) \\ X_{ij3} = &\quad Y_{ij2} \\ X_{ij4} = &\quad Y_{ij4} - Y_{ij5} \end{aligned} \tag{2.3}$$

X_{ij1} and X_{ij2} are defined to examine whether the rate of induction is the same as the rate of resolution; rejection of the null hypothesis would point to asymmetry. The statistic X_{ij3} examines the rate of induction between Days 0 and 14. The statistic X_{ij4} examines the rate of resolution between Day 35 and Day 49. These four variates describe a biomarker's pattern of change over time and could result in increased statistical power for the alternative statistical hypotheses. Preisser et al. (2011) use univariate and multivariate Wilcoxon signed rank tests to assess whether the medians of the variates $X_{ij1}, \ldots X_{ij4}$ differ from zero.

With respect to experimental gingivitis, the interpretation of X_1, X_2, X_3, and X_4 as reflecting symmetry or asymmetry has potential implications relating to the biology of the system. They provide potential insight to discriminate whether there are differences in the homeostatic mechanisms which regulate the steady-state levels of different biomarkers (Preisser et al. (2011)).

2.7 Univariate and Multivariate Hypothesis Testing Based on Wilcoxon Signed Rank Statistics

A nonparametric statistical analysis was used to assess the pattern of response from Day 0 to Day 49 based on subject-level AUC summary measures. Exact p-values from the permutation distribution of univariate and multivariate Wilcoxon signed rank

tests were generated. Let $k = 1, 2, 3, 4$ index the variate. For each biomarker $j = 1, \ldots, J(J = 31)$, a four-variate Wilcoxon Signed Rank Test to $\mathbf{X_{ij}} = (X_{ij1}, X_{ij2}, X_{ij3}, X_{ij4})'$ was developed to examine the four variates simultaneously for departure from their null median values of 0. Alternatively, univariate tests are defined corresponding to the X_{ijk}, four such tests for each biomarker resulting in 124 total p-values. Using all available data, the univariate Wilcoxon Signed Rank tests individually examine the median of X_{jk} for departure from zero. Details of the procedure have been previously described (Preisser et al., 2011). The p-values were evaluated for statistical significance, taking into account multiplicity addressed by controlling FWER (Hochberg, 1988) and FDR (Benjamini and Hochberg, 1995). Given both methods, there were only small differences among the experimental gingivitis data for multivariate tests. The primary motivation for the analysis presented was that analysis of ranks provided tests less sensitive to outliers and Gaussian distribution assumptions than provided by parametric analysis.

Because not all subjects had complete data, three types of imputation were carried to increase the amount of usable information. For biomarker levels below the lower detection limited (recorded as zero in the data), the log biomarker response level L_{ijt} was imputed as the log base 10 applied to half the lower detection limit plus 1. Imputation by substitution was carried out by replacing missing Day 0 data with Day -7 (or Day 49 data since biomarker levels are expected to return to baseline at the end of the resolution phase if Day -7 was missing). Imputation by linear interpolation from previous and next visits were performed sequentially for biomarker levels missing at Days 7, 35, 21, and 14. The latter two methods used only within-subject information for imputations.

In the study of experimental gingivitis, an attractive feature of the nonparametric approach is its ability to assess a large number of biomarkers relative to the number of

subjects. Limitations of the approach include its inability to provide direction estimation of the patterns of change over time and the use of within-subject imputations for selected missing data.

2.8 Methods for Multiple Hypothesis Testing

Many different methods have been used to address issues in multiple testing. An independent test that can be applied to dependent tests is the simple sequentially rejective multiple test procedure (Holm, 1979). This procedure works by rejecting hypotheses one at a time. Secondary (intersection) hypotheses are rejected when any of the included basic (individual) hypotheses are rejected. Beginning with the smallest p-value, compare the p-value to α/(n-i+1) until $H_{(n-i+1)}$ cannot be rejected. Reject hypotheses one at a time until no further rejections can be made. The smallest p-value $P_{(1)}$ is examined and if $P_{(1)} \leq \alpha$/n then $H_{(1)}$ is rejected and the process continues with the next p-value $P_{(2)}$, compared with α/(n-1). If $H_{(2)}$ is not rejected, then the process is stopped and the remaining hypotheses $H_{(2)}$, $H_{(3)}$, ..., $H_{(n)}$ are accepted. The generalized sequentially rejective Bonferroni test process for rejecting the hypotheses is similar to the procedure previously described. The statistics are compared against $\alpha / \sum c_i$, i=1, 2, ..., n); the number of the ordered test, c_i positive constants. However, the most relevant hypotheses are chosen a suitable test statistic (for which the one-dimensional distribution is either exactly or approximately known) is assigned to each hypothesis. Direct the power towards the most important hypotheses by choosing proper positive constants (c_1, c_2, ..., c_n), greater for the more important hypotheses. This method has an advantage over the Bonferroni procedure in that it provides higher power (depending on the alternative hypothesis).

The Bonferroni procedure defines n p-values, P_1, ..., P_n, corresponding to n statistics for testing hypotheses H_1, ..., H_n. The null hypothesis, $H_0 = H_1$, ..., H_n is rejected

if any p-value $<$ α/n. If a specific H_i is rejected when $P_i \leq \alpha$/n, then the bonferroni inequality, $\text{pr}\{\bigcup_{i=1}^{n}(P_i \leq \alpha/\text{n})\} \leq \alpha$ ensures the probability of rejecting at least one hypothesis when all are true is no greater than α. The procedure is used when conducting multiple tests of significance to set an upper bound on the FWER. The advantages of this procedure are that it is simple to use, no distributional assumptions are needed, and it enables individual alternative hypotheses to be identified. However, it is conservative and less powerful if multiple highly correlated tests are undertaken. Simes (1986) improved the Bonferroni method by defining n ordered p-values for testing hypotheses $H_0 = \{ H_{(1)},..., H_{(n)} \}$. The null hypothesis is rejected if $P_{(i)} \leq \text{i}\alpha/\text{n}$ for any i=1,...,n. The test has level α under H_0 when p-values are independent. This method provides advantageous over the Bonferroni procedure due to the lower type II error rate for a given nominal significance level and higher power when test statistics are highly correlated and several alternative hypotheses are correct. Although it is mostly beneficial for independent tests for the null hypothesis, all procedures based upon the Simes inequality have the assumption that the result derived under independence is a conservative procedure for dependent tests (Kang, 2007).

Hochberg's procedure (Hochberg, 1988) is a method that provides strong control of the FWER. It begins with the largest p-value and compares the p-value to α/(n-i+1) until $H_{(n-i+1)}$ can be rejected. The hypothesis corresponding to the rejection along with all hypotheses with smaller or equal p-values are rejected. The largest p-value $P_{(n)}$ is examined and if $P_{(n)} \leq \alpha$ then all hypotheses are rejected. If not, then $H_{(n)}$ cannot be rejected and the process continues to compare $P_{(n-1)}$ with $\frac{1}{2}\alpha$. If smaller, then all $H_{(i)}$ (i=n-1, ..., 1) are rejected. If not, then $H_{(n-1)}$ cannot be rejected and the process continues to compare $P_{(n-2)}$ with $\frac{1}{3}\alpha$, etc. according to $P_{(i)} \leq \alpha$/(n-i+1). Again, this procedure only applies to independent tests; however, it is applicable to dependent tests through justification provided by Sen (2008).

For controlling the FDR, Benjamini and Hochberg (1995) procedure tests H_1, H_2, ..., H_n based on ordered p-values $P_{(1)} \leq P_{(2)} \leq ... \leq P_{(n)}$. $H_{(i)}$ is the hypothesis corresponding to $P_{(i)}$, the ith ordered p-value. Let k be the largest i for which $P_{(i)} \leq \frac{i}{n}$ q*, then reject all $H_{(i)}$, i=1,...,k and the FDR is controlled at q*. Start by comparing the largest p-value with $\frac{i}{n}$ q*, if smaller all hypotheses are rejected. If larger, proceed to the smaller p-values until one satisfies the condition. All hypotheses having p-values less than or equal to the condition are rejected. If all null are true, FDR is equivalent to FWER. Control of FDR implies control of FWER in the weak sense. When the number of true hypotheses (n_0) < number of null hypotheses (n), FDR < FWER. A procedure that controls the FWER also controls FDR. Although designed for independent tests, this procedure can be applied to dependent tests through justification provided by Sen (2008).

2.9 Summary and Proposed Research

In many periodontal research studies, there is often interest in identifying molecular mediators of inflammation via repeated measures analysis that can be induced to significant change over time as well as characterize the direction of the change. In the presence of potential outliers, nonparametric methods are preferred over parametric methods due to the reliance on fewer assumptions and presumed greater robustness over their parametric counterparts. For the analysis of biomarker data in experimental gingivitis, most recently, a nonparametric multiple hypothesis testing approach was advocated for the analysis of repeated measures Preisser et al. (2011) using AUC summary measures to assess the change over time in biomarker levels. Though this method has the advantage of being able to assess a large number of biomarkers relative to the number of subjects, it is unclear how to handle missing data in this context, particularly left truncation of observations due to a lower detection limit. For this setting,

it may be beneficial to use a parametric method that addresses limitations associated with parametric approaches, including the missingness process, as well as assumptions about the nature of left truncation of observations due to a lower detection limit. The most common strategy for measurement data with detection limits is substitution of the value below the limit with a fraction of the limit (i.e., 0.5). Though this strategy is simplistic and easy to implement, it often distorts results (Hughes, 1999; Lubin et al., 2004) and can provide biased estimation, particularly if a large percentage of the data are left-truncated (Fang et al., 2011).

In addition to the missingness issue due to assay detection limitations, in experimental gingivitis, the multiple hypothesis testing problem warrants further investigation. Current procedures for the multiplicity problem include controlling the FWER and FDR. Although they are commonly used techniques, their restrictions on the dependency of p-values, for example, a possibility in the multiple biomarkers studied in experimental gingivitis studies, could pose a problem for their use in this setting. Modifications of the Hochberg (1988) and Benjamini and Hochberg (1995) procedures based on the Chen-Stein Theorem (Chen, 1975) to the experimental gingivitis data and similar problems warrant investigation.

We propose to illustrate two parametric approaches to provide direct estimation of the trends in biomarkers over time based on AUC computations and describe how to estimate the associated parameters (Preisser et al., 2011). The outline of the remaining sections of this proposal are as follows. In Chapter 3, we derive a piecewise model based on a linear random-effects regression model as a parametric form of the model described in Preisser et al. (2011). The parametric model will be fit to log-transformed data for 3 biomarkers, MMP7, MMP3, and MIP-1β, representing varying degrees of missingness due to lower detection limits using 2 *ad hoc* (naive) approaches for handling non-detect values and a likelihood approach accounting for left censoring ("ML1") from Lyles et al.

(2000). These *naive* approaches, named "Naive1" and "Naive2", replace non-detect GCF values by the limit of detection and half that limit, respectively. In Section 3.4, we present the results comparing MLEs from "Naive1", "Naive2", and "ML1" approaches of the 3 biomarkers, separately.

In Chapter 4, we derive a gamma curve-like nonlinear mixed model with adjustment for left truncation based on the *ad hoc* approaches outlined in Chapter 3. In Chapter 5, we propose to compare approaches for hypothesis testing (size and power) in the presence of left truncation and/or missing data in a simulation study to evaluate whether the parametric methods are reliable for small sample sizes or whether larger samples are needed to reliably use the methods. Evidence for recommending certain sample sizes for EG studies and an evaluation of whether the nonparametric method is robust to left truncation and crude single imputation methods is also provided. In Chapter 6, we discuss future research for the investigation of nonparametric multiple hypothesis testing approaches in Preisser et al. (2011) and more methods for handling missing data. We plan to use the EG study (Offenbacher et al., 2010; Preisser et al., 2011) to illustrate our methods.

CHAPTER 3: PIECEWISE LINEAR MIXED MODEL

3.1 Background and Introduction

In the characterization of biomarkers measured repeatedly over time, there is a need to summarize the information contained in the multivariate data. In experimental gingivitis (EG), for example, biomarker levels change when the benefits of toothbrushing are withheld during an induction phase, then restored during a resolution phase. The pattern of change over time of biomarker levels associated with gingivitis could demonstrate changes in different directions and kinetics of responses mirroring the underlying dynamics of the biological response; therefore, the statistical methodology utilized should consider this possibility. As such, area under the curve (AUC) can be implemented as a summary measure for estimating change in biomarker levels. Parametric statistical models for repeated measures analysis are useful for characterizing the nature of that change over time, particularly as they easily accommodate both truncated and missing data. In EG studies, left truncation results when a biomarker level falls below the lower limit of detection. This article introduces a piecewise linear mixed model to provide direct estimation of AUC based on the trends in biomarkers over time while implementing adjustments for left truncation. Estimation and hypothesis testing results for area under the "line" curve are reported for three biomarkers.

In review of statistical analysis methods used in the EG studies, although gingivitis occurs over time as a steady-state inflammatory response, only approximately one-half of studies use a repeated measures ANOVA approach (Deinzer et al., 2004, 2007; Johnson et al., 1997; Waschul et al., 2003), with few studies examining the rate of increase by

calculating AUC (Jepsen et al., 2003; Preisser et al., 2011; Salvi et al., 2010). Most studies that took a nonparametric approach to the analysis used a Wilcoxon-signed rank test for intra-subject comparisons and Mann-Whitney U-test or Wilcoxon rank-sum test to assess between-group differences (Giannopoulou et al., 2003; Konradsson et al., 2007; Konradsson and van Dijken, 2005; Staab et al., 2009; Tsalikis, 2010). Studies not using repeated measures analysis instead used paired t-tests to assess mean within group changes from baseline to each timepoint or 2-sample t-tests to assess between-group differences in mean levels at each timepoint (Konradsson et al., 2007; Salvi et al., 2010). Few studies mentioned using a log-transformation, or any other form of transformation, before analyzing the data. Additionally, most studies mentioned above did not touch on or address the assay LOD and how values below the limit were handled in the analysis. Studies that did mention the LOD did not indicate how values left-censored due to being below the detection limit were addressed, if at all (Deinzer et al., 2007; Johnson et al., 1997; Konradsson et al., 2007).

For the analysis of biomarker data in experimental gingivitis, most recently, a nonparametric multiple hypothesis testing approach was advocated for the analysis of repeated measures (Preisser et al., 2011) using AUC summary measures to assess the change over time in biomarker levels. Though this method has the advantage of being able to assess a large number of biomarkers relative to the number of subjects, it is unclear how to handle missing data in this context, particularly left truncation of observations due to a lower detection limit. The most common strategy for measurement data with detection limits is substitution of the value below the limit with a fraction of the limit (e.g., 0.5). Though this strategy is simplistic and easy to implement, it often distorts results (Hughes, 1999; Lubin et al., 2004) and can provide biased estimation, particularly if a large percentage of the data are left-truncated (Fang et al., 2011).

In the assessment of change over time, missing data can be a common occurrence

when data are measured over multiple periods of time. For longitudinal studies involving repeated measures analysis, there can be many reasons for missing data, including nonresponse, subject dropout or, as is often times the case in the measurement of biomarker data, missingness due to assay detection limits. Missing values can have a profound influence on statistical results, including estimation of summary measures of change and hypothesis testing. If the missingness mechanism is MAR, i.e., the probability that a response is observed can only depend on the values of those other factors which have been observed, there are well developed computational methods for handling missing data under this assumption (Little and Rubin, 1987). Hughes (1999) described an EM algorithm for maximum likelihood estimation of a linear mixed effects model for estimating trends in CD4 counts over time in HIV-positive subjects, accounting for left and/or right censoring. Lyles et al. (2000) developed a likelihood method that addresses missing data due to left truncation as well as an extension to additionally accommodate informative dropout. Thiebaut and Jacqmin-Gadda (2004) applied a maximum likelihood approach for left-censored data based on a Marquardt algorithm (Marquardt, 1963) in HIV research. In reference to pharmacokinetic data with measurements below the quantification limit, Fang et al. (2011) developed a maximum likelihood method to estimate AUC and the ratio of two AUCs (i.e., relative exposure).

The challenge in using parametric procedures is that the validity of standard parametric models depends on certain underlying conditions being met, particularly for smaller sample sizes. These conditions include assumptions regarding the form of the mean model (including uncertainty pertaining to the transformation of the response) and the missingness process, as well as assumptions about the nature of left truncation of observations due to a lower detection limit (Preisser et al., 2011). Thus, we need new methods to fit a parametric model, while examining the form of the mean model

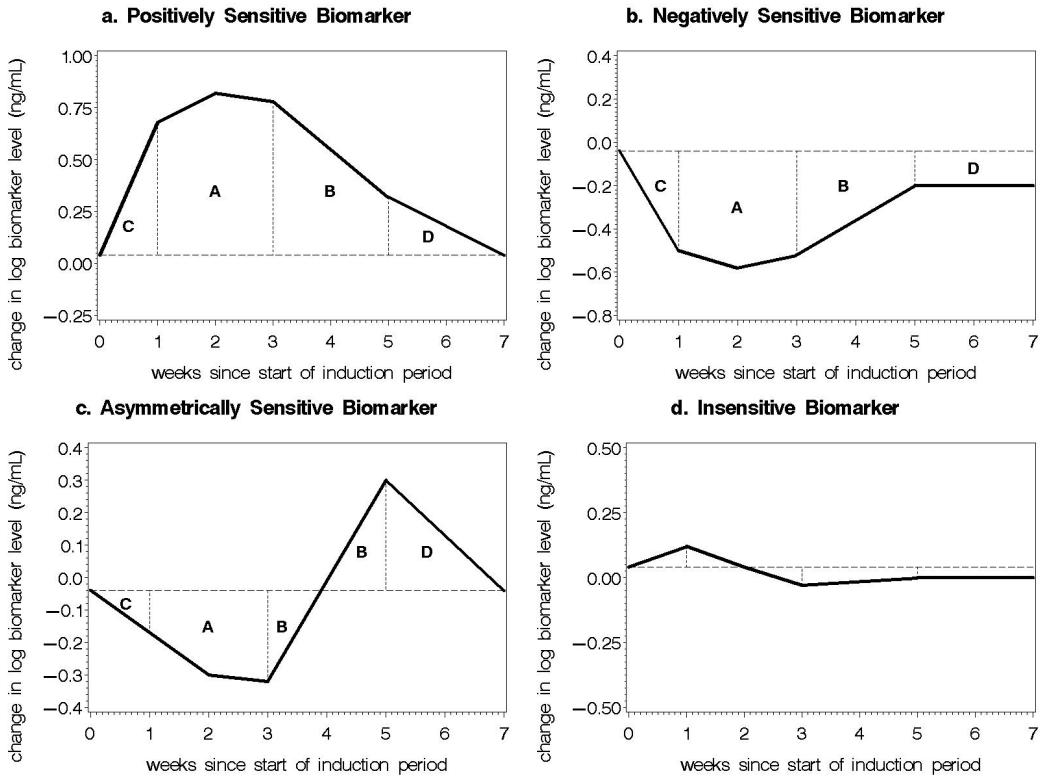
and missingness due to left truncation. For the parametric approach, we illustrate our methods under the framework of a random-effects model. Currently the most utilized *ad hoc* approaches for handling of missingness due to levels below the detectable threshold include substitution of all such values by some fraction of the limit. Parametric approaches using a survival analysis framework have been implemented as an ideal approach over simple substitution methods to address the missingness (Helsel, 2006). For repeated-measures problems, procedures for fitting linear mixed effects models to data include maximum likelihood methods. This paper proposes to develop a parametric model under framework (2.2) using the *ad hoc* approaches for handling left truncation due to lower detection limit, "Naive1" and "Naive2", from Lyles et al. (2000). Additionally, we apply Lyles et al. (2000) likelihood method accounting for left censoring ("ML1"). We will consider piecewise linear regression for fitting this model. The focus will be on providing direct estimation of the trends in biomarkers over time, calculations of AUC based on the associated parameters from Preisser et al. (2011), and hypothesis testing for AUC summaries in a study of experimental gingivitis.

3.2 Piecewise Linear Mixed Model

Previously, we mentioned that a limitation of using a nonparametric hypothesis testing approach is its inability to provide direct estimation of the patterns of change over time (Preisser et al., 2011). An important question in the identification of candidate biomarkers in experimental gingivitis is the nature of changes in trend patterns. Experimental gingivitis biomarker levels are typically expected to follow one of four patterns involving a directional change at a particular value (Figure 1). Experimental gingivitis biomarker levels have been described as occurring in phases, where there are distinctly different biomarker characteristics associated with GCF levels under different phases of change. For instance, for a positively sensitive biomarker, the levels are

expected to increase during the induction phase to a critical value and decrease during the resolution phase. The beginning of the resolution phase is thought to occur at or near the time when stents are discontinued and hygiene on all teeth is reinstituted to resolve inflammation.

Figure 1: Typical biomarker patterns of change over time for experimental gingivitis. Letters A, B, C, and D denote a partition of AUC for which summary measures of change can be estimated.



Piecewise linear regression can be applied to describe these trend data as it is a form of regression that allows multiple linear segments to be fit to the data for a set of pre-specified change points. To fix ideas, the change points are set to the measurement occasions assumed to be at 0, 1, 2, 3, 5, and 7 weeks as in, for example, Offenbacher et al. (2010). As an illustrative example of one commonly used piecewise regression model, suppose we have a two-segment linear regression model. Let (x_i, y_i),

$i = 1,\ldots,$n denote n pairs of observations. We assume that the x_i are ordered such that $x_1 \leq x_2 \leq \ldots \leq x_n$. Suppose $(x_i,\ y_i)$ is a sequence of independent observations satisfying the following model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 (x_{i1} - t) x_{i2} + \epsilon_i \tag{3.1}$$

where

- y_i is the log GCF level for subject i
- $x_{i1} = x_i$ is the time, in weeks, for subject i
- x_{i2} is a dummy variable (0, if $x_i \leq t$ and 1, if $x_i > t$)

and the independent error terms $\epsilon_i \sim \mathrm{N}(0,\ \sigma^2)$.

The corresponding linear regression functions are of the form

$$\begin{aligned} y_{i1} &= a_1 + b_1 x_i, \quad x_i \leq t \\ &= \beta_0 + \beta_1 x_i \\ y_{i2} &= a_2 + b_2 x_i, \quad x_i > t \\ &= \beta_0 + \beta_1 x_i + \beta_2 (x_i - t) \\ &= (\beta_0 - \beta_2 t) + (\beta_1 + \beta_2) x_i \end{aligned} \tag{3.2}$$

where $a_1 = \beta_0$ and $a_2 = \beta_0 - \beta_2 t$ and $b_1 = \beta_1$ and $b_2 = \beta_1 + \beta_2$ are the intercepts and slopes of the linear segments, respectively. For continuous models, the regression function is continuous at the change point t, satisfying the following:

$$a_1 + b_1 t = a_2 + b_2 t \tag{3.3}$$

Equation (3.3) is equivalent to $\beta_0 + \beta_1 t = (\beta_0 - \beta_2 t) + (\beta_1 + \beta_2)t$, which shows that (3.1) is continuous. Given what is understood about the nature of experimental gingivitis, we assume the function should be continuous. A piecewise linear regression model is defined to describe the biomarker pattern of change over the six timepoints to coincide with summary indices of change associated with experimental gingivitis described below. Let Y_{ik} be the GCF level of each biomarker (on the log base 10 scale for the application considered in this section) for the ith subject at the kth time point, for $k = 0, \ldots, 5$, which are $t_{i0} = 0, t_{i1} = 1, t_{i2} = 2, t_{i3} = 3, t_{i4} = 5, t_{i5} = 7$ weeks. Expanding the continuous two-phase model previously described to a continuous five-phase model with random intercept, the general form of the model (extending (2.2)) can be written as

$$Y_{ik} = \mathbf{Z}_\mathbf{i}'\boldsymbol{\beta} + b_{0i} + \boldsymbol{\epsilon}_{ik} \tag{3.4}$$

where $b_{0i} \sim N(0, \sigma_b^2)$ are subject-specific random intercepts and $\boldsymbol{\epsilon}_{ik} \sim N(0, \sigma^2)$ are random errors, with b_{0i}, $i = 1, \ldots, n$ and $\boldsymbol{\epsilon}_{ik}$: $i = 1, \ldots, n$; $k = 0, \ldots, 5$ mutually independent. We assume that $\mathbf{Z_i}$ is a vector of explanatory variables that are functions of time, including the intercept. An alternative covariance structure for $\mathbf{Y_i} = (Y_{i0}, Y_{i1}, \ldots, Y_{i5})'$ is given by a random slopes model

$$\mathbf{Y_{ik}} = \mathbf{Z_i'}\boldsymbol{\beta} + \mathbf{b_{1i}t_{ik}} + \boldsymbol{\epsilon_{ik}} \tag{3.5}$$

where $b_{1i} \sim N(0, \sigma_s^2)$ and $\boldsymbol{\epsilon}_{ik} \sim N(0, \sigma^2)$ are mutually independent. As in model (2.2), where both random intercept and slope coefficients are introduced, in practice, a model could contain both terms; however, because most EG studies are of small sample size, these studies would not likely allow estimation of more than 2 variance components. In model (3.4), $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$ and $\mathbf{Z_i} = (Z_{i0}, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5})'$ is de-

fined to give conjoined piecewise linear segments. To parameterize the model, define for $k = 0, 1, 2, 3, 4, 5$:

$$
\begin{aligned}
z_{i0} &= 1 \\
z_{i1} &= t_{ik} \\
z_{i2} &= (t_{ik} - 1)\mathrm{I}(t_{ik} > 1) \\
z_{i3} &= (t_{ik} - 2)\mathrm{I}(t_{ik} > 2) \\
z_{i4} &= (t_{ik} - 3)\mathrm{I}(t_{ik} > 3) \\
z_{i5} &= (t_{ik} - 5)\mathrm{I}(t_{ik} > 5)
\end{aligned}
$$

While covariates other than those that are functions of time are not considered in this dissertation, they could be easily added. The relationship between the z_i and y_i values can be described by the following six linear regression functions in terms of model βs:

$$
\begin{aligned}
Y_{i0} &= \beta_0 + b_{0i} + \epsilon_{i0} \\
Y_{i1} &= \beta_0 + \beta_1 + b_{0i} + \epsilon_{i1} \\
Y_{i2} &= \beta_0 + 2\beta_1 + \beta_2 + b_{0i} + \epsilon_{i2} \\
Y_{i3} &= \beta_0 + 3\beta_1 + 2\beta_2 + \beta_3 + b_{0i} + \epsilon_{i3} \\
Y_{i4} &= \beta_0 + 5\beta_1 + 4\beta_2 + 3\beta_3 + 2\beta_4 + b_{0i} + \epsilon_{i4} \\
Y_{i5} &= \beta_0 + 7\beta_1 + 6\beta_2 + 5\beta_3 + 4\beta_4 + 2\beta_5 + b_{0i} + \epsilon_{i5}
\end{aligned}
\tag{3.6}
$$

Next, using the definitions in Section 2.6, the summaries of AUC (with baseline adjust-

ment) can be estimated in terms of βs as follows:

$$
\begin{aligned}
E(A_i) &= E(Y_{i1} + 2Y_{i2} + Y_{i3})/2 - 2\beta_0 = \frac{1}{2}(8\beta_1 + 4\beta_2 + \beta_3) \\
E(B_i) &= E(Y_{i3} + Y_{i4}) - 2\beta_0 = 8\beta_1 + 6\beta_2 + 4\beta_3 + 2\beta_4 \\
E(C_i) &= E[(Y_{i0} + Y_{i1})/2] - \beta_0 = \frac{1}{2}\beta_1 \\
E(D_i) &= E(Y_{i4} + Y_{i5}) - 2\beta_0 = 12\beta_1 + 10\beta_2 + 8\beta_3 + 6\beta_4 + 2\beta_5
\end{aligned}
\tag{3.7}
$$

Four summary indices of change in equation (2.3) of Section 2.6 can now be defined in terms of βs as follows:

$$
\begin{aligned}
E(X_{i1}) &= E(C_i - \frac{1}{2}D_i) = \frac{1}{2}E[Y_{i1} - Y_{i4} - Y_{i5}] \\
&= -\frac{11}{2}\beta_1 - 5\beta_2 - 4\beta_3 - 3\beta_4 - \beta_5 \\
E(X_{i2}) &= E(A_i - B_i) = \frac{1}{2}E(Y_{i1} + 2Y_{i2} - Y_{i3} - 2Y_{i4}) \\
&= -4\beta_1 - 4\beta_2 - \frac{7}{2}\beta_3 - 2\beta_4 \\
E(X_{i3}) &= E(Y_{i2}) = 2\beta_1 + \beta_2 \\
E(X_{i4}) &= E(Y_{i4} - Y_{i5}) = -2\beta_1 - 2\beta_2 - 2\beta_3 - 2\beta_4 - 2\beta_5
\end{aligned}
\tag{3.8}
$$

where X_{i1} and X_{i2} examine whether the rate of induction is the same as the rate of resolution. X_{i3} examines the rate of induction between Week 0 and Week 2. X_{i4} examines the rate of resolution between Week 5 and Week 7. The statistical analysis of these variates addressing left truncation is performed based on the likelihood methods outlined in the next section.

3.3 Maximum Likelihood Estimation in the Presence of Left Truncation

In experimental gingivitis, GCF is collected at the beginning of the hygiene phase, the beginning of the induction phase, during and at the end of the induction phase,

and during the resolution phase. A limitation of using laboratory assays is that the biomarkers levels below the detection limit, as confirmed by internal controls, are not quantifiable. While many methods have been proposed to address the censored data, substitution methods are among the most popular. For "Naive1" and "Naive2" methods, GCF measurements are replaced by the limit of detection and half the limit of detection, respectively. For "ML1", for a biomarker, let the limit of detection be denoted by d. For $\mathbf{Y_i} = (Y_{i0}, Y_{i1}, \ldots, Y_{in_i})$, let n_{i1} represent the number of detectable GCF values and $n_i - n_{i1}$ represent the non-detectable GCF values ($n_{i1} \in \{0, \ldots, n_i\}$) (Lyles et al., 2000). The likelihood function for the parameter vector $\mathbf{\Omega} = (\beta,\ \sigma_b^2,\ \sigma^2)$, with an asterisk to denote that Y_i may include one or more non-detectable values, is

$$L(\Omega; \mathbf{Y}) = \prod_{i=1}^{l} f^*(\mathbf{Y}_{ik}; \Omega), \tag{3.9}$$

where $f^*(\mathbf{Y}_{ik}; \Omega) = \int_{-\infty}^{\infty} f^*(\mathbf{Y}|b_{0i})\, f(b_{0i})\ d\, b_{0i}$

As indicated in Lyles et al. (2000), a detectable value contributes $f(Y_{ik}|b_{0i})$ and a non-detectable value contributes the Bernoulli probability $F_y(d|b_{0i})$, F_y is the cumulative distribution function. The complete-data likelihood can be written as

$$L(\Omega; \mathbf{Y}) = \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{n_{i1}} f(\mathbf{Y}_{ik}|b_{0i}) \right\} \left\{ \prod_{k=n_{i1}+1}^{n_i} F_Y(d|b_{0i}) \right\} f(b_{0i}) db_{0i} \right] \tag{3.10}$$

The procedure treating all values as detectable ($n_{i1} = n_i$) is used for "Naive1" and "Naive2" methods, which is based on the standard log-likelihood function for a mixed model. The SAS NLMIXED procedure can be used to fit this maximum likelihood function (Thiebaut and Jacqmin-Gadda, 2004). The code for fitting this maximum likelihood function to the biomarker data is provided in Appendix A. Maximum likelihood estimation by adaptive Gauss-Hermite quadrature is used to estimate $\mathbf{\Omega}$.

3.4 Example

(Preisser et al., 2011) describe a study based on data previously published by Offenbacher et al. (2010) in which thirty-one inflammatory mediators within each gingival crevicular fluid (GCF) sample, including cytokines, matrix-metalloproteinases (MMPs) and adipokines were studied in 22 subjects to evaluate the changes in the GCF composition over time. The study recruited and enrolled subjects with naturally occurring gingivitis, defined as bleeding upon probing present, typically in at least 10% of dental sites, as these subjects were more likely to develop experimental gingivitis in the course of the study. The course of the experiment included a 1-week hygiene phase, a 3-week induction phase using two stents and a 4-week resolution phase. Gingival crevicular fluid was collected from the same oral sites at the beginning of the hygiene phase (or Day -7, one week prior to baseline), weekly during the induction phase (Day 0 or baseline), Day 7, 14, 21 (end of the induction phase/baseline for the resolution phase), and biweekly during the resolution phase at Day 35 and 49. At the final time point, baseline levels are often restored. However, some biomarkers may not have their expression levels restored to baseline levels.

At specified time points, gingival crevicular fluid from the stent teeth was collected from each sample. Gingival crevicular fluid is collected at multiple dental sites and fluid levels measured by use of different assays corresponding to different biomarkers. For each biomarker, the average of two concentration measurements was considered the measurement for the particular site and time point. The goal of the experiment was to identify new candidate biomarkers that were sensitive to poor oral health care as identified by their regulation patterns during induction and resolution of gingivitis as related to changes in clinical signs of disease. The data have been previously analyzed using parametric linear mixed modeling (Offenbacher et al., 2010) with zeros inserted for left-truncated values.

The extent of missingness or truncation for the three representative biomarkers chosen for this analysis, MMP7, MMP3, and MIP-1β, ranged from 16 (12.12%) to 36 (27.27%) observations. Approximately 1% to 12% of these missing values were due to the values being below the biomarker's detection limit. Table 1 outlines the level of missingness by biomarker.

Table 1: Amount of missingness across biomarkers

Biomarker	Left Truncated	Other	Total Missing	Observed	Total
MMP7	1 (0.76%)	15 (11.36%)	16 (12.12%)	116 (87.88%)	132
MIP-1β	16 (12.12%)	20 (15.15%)	36 (27.27%)	96 (72.73%)	132
MMP3	2 (1.52%)	15 (11.36%)	17 (12.88%)	115 (87.12%)	132
All Biomarkers	19 (4.80%)	50 (12.63%)	69 (17.42%)	327 (82.58%)	396

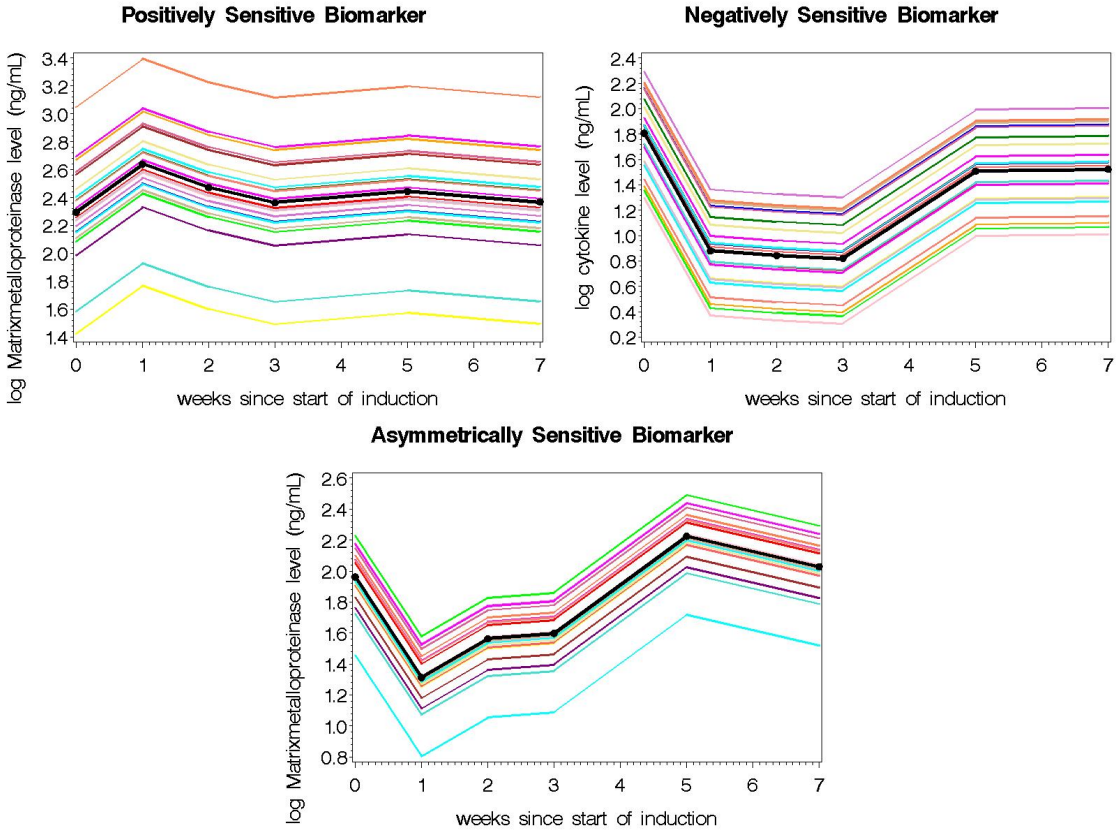
3.5 Results

3.5.1 Regression Estimates

Figure 2 presents the subject-level best linear unbiased predictions and population-averaged biomarker levels using the piecewise linear mixed model (and "ML1" to account for left censoring). The Model in equation (3.4) is fitted to the experimental gingivitis study data using the 2 *ad hoc* approaches ("Naive1" and "Naive2"). A standard linear mixed model procedure is used to obtain the MLEs and associated standard errors under the following assumption. The minimum detection limits for the 4 biomarkers in nanograms per millilitre on the log scale are as follows: MMP7: 1.470, MMP3: 0.415, and MIP-1β: -0.143. Next, the nonlinear mixed model procedure from Section 3.3 is used to obtain the MLEs and associated standard errors for the ML approach accounting for left censoring ("ML1"). Table 2 presents the ML estimates and associated standard errors from the piecewise linear regression model in equations (3.4)

and (3.6) for the 3 biomarkers based on the 3 missing data handling methods. Table 3 presents results from naive methods using the NLMIXED procedure. The NLMIXED results suggest that using maximum likelihood estimation by adaptive Gauss-Hermite quadrature for comparing the 3 methods would lead to marginally larger standard errors for the 2 naive approaches.

Figure 2: Subject-level best linear unbiased predictions and population-averaged biomarker levels over time for MMP7 (upper left), MIP-1β (upper right), and MMP3 (lower center) based on the piecewise linear mixed model.



For each biomarker, Figure 3 shows the population-averaged estimated curves for the three methods. For MMP7 and MMP3, the 3 methods show similar patterns over time, which is not surprising since these two biomarkers had the lowest rates of left truncated data. With MIP-1β, "ML1" population-averaged biomarker levels were

higher from the beginning of the induction phase through the end of the resolution phase. During this same timeframe, "Naive1" estimates were similar to the "ML1" estimates. However, the "Naive2" estimates were different during the induction phase through the end of the resolution phase, with an apparent greater decreasing trend in levels during the induction phase and a greater increasing trend in levels during the first 2 weeks of the resolution phase.

Figure 3: Population-averaged biomarker levels over time for MMP7 (upper left), MIP-1β (upper right), and MMP3 (lower center) for the three *ad hoc* methods based on the piecewise linear mixed model.

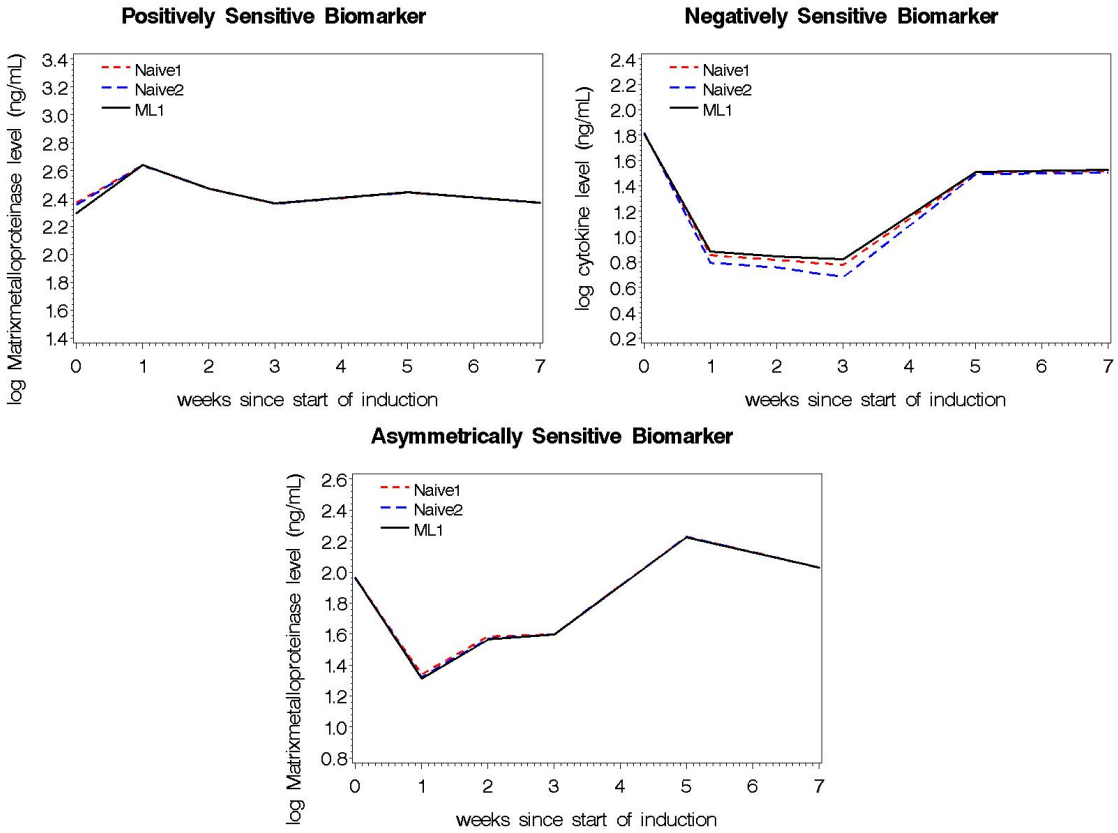


Table 2: Results from experimental gingivitis study using a piecewise linear mixed model: comparing ML estimates from *ad hoc* approaches with ML estimates from an approach accounting for non-detectable values

		Estimates (SEs)							
Biomarker	Method	β_0	β_1	β_2	β_3	β_4	β_5	σ_b^2	σ^2
MMP7	Naive1†	2.370	0.268	-0.434	0.057	0.149	-0.077	0.120	0.070
		(0.095)	(0.092)	(0.163)	(0.150)	(0.110)	(0.072)	(0.040)	(0.010)
	Naive2*	2.355	0.284	-0.450	0.058	0.149	-0.078	0.124	0.072
		(0.096)	(0.093)	(0.165)	(0.152)	(0.111)	(0.073)	(0.041)	(0.010)
	ML1$^\diamond$	2.362	0.276	-0.442	0.058	0.149	-0.077	0.122	0.071
		(0.096)	(0.092)	(0.164)	(0.151)	(0.110)	(0.073)	(0.041)	(0.010)
MIP-1β	Naive1†	1.812	-0.960	0.923	-0.0040	0.404	-0.357	0.125	0.190
		(0.135)	(0.161)	(0.275)	(0.246)	(0.179)	(0.126)	(0.050)	(0.028)
	Naive2*	1.819	-1.027	0.993	-0.044	0.483	-0.399	0.159	0.255
		(0.155)	(0.186)	(0.318)	(0.285)	(0.207)	(0.146)	(0.064)	(0.038)
	ML1$^\diamond$	1.818	-1.019	0.985	-0.048	0.490	-0.402	0.156	0.252
		(0.154)	(0.188)	(0.322)	(0.290)	(0.213)	(0.147)	(0.063)	(0.042)
MMP3	Naive1†	1.962	-0.622	0.865	-0.229	0.301	-0.414	0.043	0.179
		(0.105)	(0.146)	(0.259)	(0.239)	(0.175)	(0.116)	(0.024)	(0.026)
	Naive2*	1.962	-0.641	0.889	-0.219	0.284	-0.412	0.048	0.187
		(0.108)	(0.149)	(0.265)	(0.244)	(0.179)	(0.118)	(0.026)	(0.027)
	ML1$^\diamond$	1.962	-0.633	0.879	-0.222	0.291	-0.413	0.046	0.184
		(0.107)	(0.148)	(0.264)	(0.243)	(0.178)	(0.117)	(0.026)	(0.027)

† Substituting non-detectable values by the detection limit.
$*$ Substituting non-detectable values by half the detection limit.
$\diamond$ Accounting for non-detectable values via (3.10).
SAS PROC MIXED with ML option was used for the naive methods.

Table 3: Results from experimental gingivitis study using a piecewise linear mixed model fit in PROC NLMIXED: comparing ML estimates from *ad hoc* approaches with ML estimates from an approach accounting for non-detectable values

		Estimates (SEs)							
Biomarker	Method	β_0	β_1	β_2	β_3	β_4	β_5	σ_b^2	σ^2
MMP7	Naive1†	2.370	0.268	-0.434	0.057	0.149	-0.077	0.120	0.070
		(0.095)	(0.092)	(0.163)	(0.150)	(0.110)	(0.072)	(0.040)	(0.010)
	Naive2*	2.355	0.284	-0.450	0.058	0.149	-0.078	0.124	0.072
		(0.096)	(0.093)	(0.165)	(0.152)	(0.111)	(0.073)	(0.041)	(0.010)
	ML1$^\diamond$	2.362	0.276	-0.442	0.058	0.149	-0.077	0.122	0.071
		(0.096)	(0.092)	(0.164)	(0.151)	(0.110)	(0.073)	(0.041)	(0.010)
MIP-1β	Naive1†	1.812	-0.960	0.923	-0.0040	0.404	-0.357	0.125	0.190
		(0.135)	(0.161)	(0.275)	(0.246)	(0.179)	(0.127)	(0.050)	(0.028)
	Naive2*	1.819	-1.027	0.993	-0.044	0.483	-0.399	0.159	0.255
		(0.155)	(0.187)	(0.318)	(0.285)	(0.207)	(0.147)	(0.064)	(0.038)
	ML1$^\diamond$	1.818	-1.019	0.985	-0.048	0.490	-0.402	0.156	0.252
		(0.154)	(0.188)	(0.322)	(0.290)	(0.213)	(0.147)	(0.063)	(0.042)
MMP3	Naive1†	1.962	-0.622	0.865	-0.229	0.301	-0.414	0.043	0.179
		(0.105)	(0.146)	(0.259)	(0.239)	(0.175)	(0.116)	(0.024)	(0.026)
	Naive2*	1.962	-0.641	0.889	-0.219	0.284	-0.412	0.048	0.187
		(0.108)	(0.149)	(0.265)	(0.244)	(0.179)	(0.118)	(0.026)	(0.027)
	ML1$^\diamond$	1.962	-0.633	0.879	-0.222	0.291	-0.413	0.046	0.184
		(0.107)	(0.148)	(0.264)	(0.243)	(0.178)	(0.117)	(0.026)	(0.027)

† Substituting non-detectable values by the detection limit.
$*$ Substituting non-detectable values by half the detection limit.
$\diamond$ Accounting for non-detectable values via (3.10).

3.5.2 Estimates, SEs, and 95% Confidence Intervals of Area-Under-the-Curve (A, B, C, and D)

For the summary measures of AUC, let $\boldsymbol{\theta}_1 = \mathrm{E}(\mathbf{R_i})$ where $\mathbf{R_i} = (A_i, B_i, C_i, D_i)$, $i = 1, \ldots, \mathrm{n}$, are *i.i.d*. Then, based on equation (3.7), $\boldsymbol{\theta}_1 = \mathrm{C}\boldsymbol{\beta}$ where

$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$ and

$$C = \begin{pmatrix} 4 & 2 & \frac{1}{2} & 0 & 0 \\ 8 & 6 & 4 & 2 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 \\ 12 & 10 & 8 & 6 & 2 \end{pmatrix}$$

For the piecewise linear mixed regression model described in equations (3.4) and (3.6), the maximum likelihood estimation described in Section 3.4 is used to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{V_\beta} = \widehat{Var}(\widehat{\boldsymbol{\beta}})$ so that $\widehat{Var}(\widehat{\boldsymbol{\theta}}_1) = C\widehat{V_\beta}C'$. Estimation of $\widehat{Var}(\widehat{\boldsymbol{\theta}}_1)$ is carried out using the SAS Version 9.2 IML procedure (SAS, Inc., Cary, North Carolina). The procedure treating all values as detectable ($n_{i1} = n_i$) is used for "Naive1" and "Naive2" methods, reducing to the standard estimation approach (Laird and Ware, 1982). Table 4 presents the estimates, standard errors, and 95% confidence intervals (CIs) of AUC from the piecewise linear mixed model for the 3 biomarkers based on *ad hoc* and "ML1" data handling methods. For MMP7, the rates of induction and resolution from the "ML1" method fell within the range of the estimated rates for both naive methods, with lower estimated rates for "Naive1" and higher estimated rates for "Naive2". Similar results were seen for both MMP3 and MIP-1β.

Table 4: Estimates (SEs), and 95% CIs of AUC from experimental gingivitis study using piecewise linear mixed model: comparisons of *ad hoc* approaches with an approach accounting for non-detectable values

		Estimates (SEs) and 95% CIs			
Biomarker	Method	*A*	*B*	*C*	*D*
MMP7	Naive1†	0.233	0.068	0.134	0.075
		(0.141)	(0.146)	(0.046)	(0.146)
		(-0.047, 0.512)	(-0.222, 0.357)	(0.043, 0.225)	(-0.215, 0.364)
	Naive2∗	0.264	0.098	0.142	0.105
		(0.143)	(0.148)	(0.046)	(0.148)
		(-0.020, 0.547)	(-0.195, 0.392)	(0.050, 0.234)	(-0.188, 0.398)
	ML1◇	0.249	0.084	0.138	0.090
		(0.142)	(0.147)	(0.046)	(0.147)
		(-0.047, 0.545)	(-0.222, 0.390)	(0.042, 0.234)	(-0.216, 0.340)
MIP-1β	Naive1†	-1.993	-1.344	-0.480	-0.603
		(0.256)	(0.263)	(0.080)	(0.263)
		(-2.502, -1.485)	(-1.866, -0.822)	(-0.640, -0.320)	(-1.125, -0.081)
	Naive2∗	-2.143	-1.466	-0.513	-0.642
		(0.296)	(0.304)	(0.093)	(0.304)
		(-2.732, -1.554)	(-2.070, -0.861)	(-0.699, -0.328)	(-1.247, -0.038)
	ML1◇	-2.131	-1.457	-0.510	-0.630
		(0.297)	(0.304)	(0.094)	(0.303)
		(-2.418, -1.427)	(-1.795, -0.779)	(-0.620, -0.308)	(-1.092, -0.076)
MMP3	Naive1†	-0.871	-0.098	-0.311	0.333
		(0.224)	(0.232)	(0.073)	(0.232)
		(-1.316, -0.425)	(-0.559, 0.364)	(-0.456, -0.166)	(-0.129, 0.794)
	Naive2∗	-0.896	-0.100	-0.321	0.330
		(0.229)	(0.237)	(0.074)	(0.237)
		(-1.351, -0.440)	(-0.572, 0.371)	(-0.469, -0.173)	(-0.141, 0.801)
	ML1◇	-0.886	-0.099	-0.317	0.331
		(0.228)	(0.236)	(0.074)	(0.235)
		(-1.359, -0.412)	(-0.589, 0.391)	(-0.470, -0.163)	(-0.159, 0.821)

† Substituting all non-detectable values by the detection limit.
∗ Substituting all non-detectable values by half the detection limit.
◇ Accounting for non-detectable values via (3.10).

3.5.3 Estimates and Standard Errors of Summary Indices of Change (X1, X2, X3, and X4) and P-values

For the summary indices of change, let $\boldsymbol{\theta}_2 = \mathrm{E}(\mathbf{X_i})$ where $\mathbf{X_i} = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, $i = 1, \ldots, \mathrm{n}$, are $i.i.d.$ Then $\boldsymbol{\theta}_2 = \mathrm{D}\boldsymbol{\beta}$ and

$$D = \begin{pmatrix} -\frac{11}{2} & -5 & -4 & -3 & -1 \\ -4 & -4 & -\frac{7}{2} & -2 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ -2 & -2 & -2 & -2 & -2 \end{pmatrix}$$

For each biomarker, univariate tests are defined corresponding to the 4 variates, X_{i1}, X_{i2}, X_{i3}, X_{i4}, for each method of handling missing data. The univariate testing approach applies the univariate χ^2 test to each of the measures for the 3 biomarkers studied under the 4 variates for addressing missingness due to left truncation resulting in 12 p-values. A univariate test for the jth element of $\boldsymbol{\theta}_1$, $H_0 : \boldsymbol{\theta}_{1j} = 0$ vs. $H_1 : \boldsymbol{\theta}_{1j} \neq 0$ is

$$T_j = \frac{(\boldsymbol{C}_j\widehat{\boldsymbol{\beta}})^2}{\boldsymbol{C}_j\widehat{V_\beta}\boldsymbol{C}_j'} \tag{3.11}$$

where $\boldsymbol{C}_j$ is the jth row of $\boldsymbol{C}$. T has an asymptotic χ^2_1 distribution. A multivariate test of $\boldsymbol{\theta}_1$, $H_0 : \boldsymbol{\theta}_1 = 0$ vs. $H_1 : \boldsymbol{\theta}_1 \neq 0$ is

$$T = (\boldsymbol{C}\widehat{\boldsymbol{\beta}})'[\boldsymbol{C}\widehat{V_\beta}\boldsymbol{C}']^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}}) \tag{3.12}$$

has an asymptotic χ^2_4 distribution. Testing for $\boldsymbol{\theta}_2$ is performed analogously.

Table 5 reports the p-values for univariate and multivariate tests. For MMP3, tests for X_1, X_2, and X_3 using "ML1" were statistically significant at $\alpha = 0.05$. For this biomarker, this suggests suppression of mediator levels during the first two weeks of the

induction phase and asymmetry around the end of the induction phase/beginning of the resolution phase (Week 3). Similar results were seen for the *ad hoc* approaches. X_4 was not significant suggesting similarity of mediator levels during the last two weeks of the resolution phase, perhaps, resolution by Week 5. All multivariate test p-values were significant at the 0.05 level. For MMP7, X_2 achieved statistical significance for "ML1" at $\alpha = 0.10$ suggesting asymmetry between the rate of induction and the rate of resolution during the first two weeks of the resolution phase. None of the multivariate test p-values were significant at the 0.05 level. For MIP-1β, X_2 and X_3 were significant for "ML1" suggesting asymmetry about Week 3, and suppression of biomarker levels within the first two weeks of induction. Similar results were seen for the *ad hoc* approaches. All of the multivariate test p-values were significant at the 0.05 level.

Table 5: Estimates (SEs) of summary indices of change and p-values from experimental gingivitis study using piecewise regression: comparing significance from *ad hoc* approaches with significance from ML approach accounting for non-detectable values

		Estimates (SEs) and P-values				
Biomarker	Method	X_1	X_2	X_3	X_4	MVT
MMP7	Naive1†	0.097	0.165	0.102	0.074	
		(0.063)	(0.096)	(0.085)	(0.083)	
		0.123	0.087	0.228	0.373	0.147
	Naive2*	0.089	0.165	0.117	0.074	
		(0.063)	(0.097)	(0.086)	(0.084)	
		0.159	0.090	0.170	0.378	0.147
	ML1◇	0.093	0.165	0.110	0.074	
		(0.063)	(0.097)	(0.085)	(0.084)	
		0.141	0.088	0.197	0.376	0.147
MIP-1β	Naive1†	-0.178	-0.649	-0.996	-0.014	
		(0.107)	(0.165)	(0.149)	(0.143)	
		0.096	0.0001	<0.0001	0.923	<0.0001
	Naive2*	-0.192	-0.678	-1.061	-0.012	
		(0.124)	(0.191)	(0.173)	(0.165)	
		0.121	0.0004	<0.0001	0.940	<0.0001
	ML1◇	-0.195	-0.675	-1.054	-0.012	
		(0.124)	(0.193)	(0.174)	(0.165)	
		0.117	0.0004	<0.0001	0.943	<0.0001
MMP3	Naive1†	-0.477	-0.773	-0.378	0.199	
		(0.099)	(0.153)	(0.135)	(0.133)	
		<0.0001	<0.0001	0.005	0.133	<0.0001
	Naive2*	-0.486	-0.796	-0.393	0.197	
		(0.101)	(0.156)	(0.138)	(0.136)	
		<0.0001	<0.0001	0.004	0.146	<0.0001
	ML1◇	-0.482	-0.787	-0.387	0.198	
		(0.101)	(0.156)	(0.137)	(0.135)	
		<0.0001	<0.0001	0.005	0.142	<0.0001

† Substituting all non-detectable values by the detection limit.
∗ Substituting all non-detectable values by half the detection limit.
◇ Accounting for non-detectable values via (3.10).

3.6 Discussion and Conclusion

This paper has presented a piecewise linear mixed model to provide direct estimation of AUC based on the trends in biomarkers over time. The method was illustrated using data from an experimentatl gingivitis study. Our objective was two-fold: (1)

to outline current methods in the experimental gingivitis literature for analyzing repeated measures biomarker data and motivate an easy to implement parametric model that provides estimation of change in biomarker levels and (2) to present a maximum likelihood methodology for addressing left truncation while illustrating the potential negative effect on population level estimates and hypothesis testing when simple imputations methods are used to correct left censoring. Lee and Kong (2011) have proposed median regression for longitudinal left-censored biomarker data subject to a detection limit. However, their approach based on weighted estimating equations may not be broadly accessible to analysts. For this setting, it may be beneficial to use a well-established parametric method that addresses the missingness process, as well as assumptions about the nature of left truncation of observations due to a lower detection limit. The general approach in this article could be extended to modeling piecewise polynomial (e.g., quadratic and cubic) curves (Edwards et al., 2006)

Although the proposed methodology was motivated by trends in repeated measures experimental gingivitis data, the approach can be applied to other areas with a similar longitudinal setup and multiple outcomes. Additionally, the likelihood framework presented in this article can be extended to more fixed and random effects; however, special consideration should be given for studies of small sample size like the experimental gingivitis study when introducing more than two random effects. For the naive methods, SAS PROC MIXED with ML option was used. For small sample sizes, the restricted ML option with finite-sample correction (i.e., Kenward-Roger method) for computing denominator degrees of freedom may be a better option; however, it would lead to larger variance component estimates and larger standard errors for the parameter estimates and variance components.

In conclusion, the examples presented here illustrate the potential for using a parametric statistical model for analyzing AUC as a summary measure for estimating change

in biomarker levels. However, emphasis is placed on the importance of using an appropriate means of taking into account left truncation in order to avoid inaccurate results for population parameter estimates and hypothesis tests based on linear mixed models. Mild distortion of results was illustrated (Figure 3) for one biomarker (MIP-1β) for which 12% of the data were left truncated. Consistent with earlier studies, a larger percentage of left truncated values are required for naive data imputation methods to fundamentally alter conclusions with respect to the theoretically superior "ML1" method. Specifically, Lubin et al. (2004) report very small bias when the percentage of measurements below detection limits is small (5-10%); however, they reported distortion in inferences when 30% of more of the data are below detection limits. In conclusion, using the maximum likelihood procedure discussed in this dissertation can provide a uniform means of analyzing biomarker data in periodontal disease to identify changes as a function of time and for handling left truncation.

CHAPTER 4: NONLINEAR GAMMA-LIKE MIXED MODEL

4.1 Background and Introduction

In the characterization of biomarkers measured repeatedly over time, there is a need to summarize the information contained in the multivariate data. In experimental gingivitis, for example, biomarker levels change when the benefits of toothbrushing are withheld during an induction phase, then restored during a resolution phase. The pattern of change over time of biomarker levels associated with gingivitis could reflect change in various directions; therefore, the statistical methodology utilized should consider this possibility. As such, AUC can be implemented as a summary measure for estimating change in biomarker levels. Parametric statistical models for repeated measures analysis are useful for characterizing the nature of that change over time, particularly as they easily accommodate both truncated and missing data. In EG studies, left truncation results when a biomarker level falls below the lower limit of detection. This article introduces a gamma curve-like non-linear mixed model to provide direct estimation of AUC based on the trends in biomarkers over time while implementing adjustments for left truncation. Estimation and hypothesis testing results for AUC are reported for three biomarkers.

In periodontal research, there is often interest in identifying molecular mediators of inflammation that can be induced to significant change over time as well as quantify the magnitude of the change within an experimental repeated measures study. Dental researchers often employ experimental gingivitis (EG) to realize a better understanding of the hosts immune response to periodontal pathogens (Deinzer et al., 2007). As such,

EG is widely used for elucidating the inflammatory response to undisturbed dental biofilm accumulation in a well-controlled pretest-posttest experimental design framework. The course of EG includes a hygiene phase, an induction phase, and resolution phase. Gingivitis is induced by withholding tooth brushing by the use of intraoral acrylic stents that cover selected teeth in each arch during tooth brushing to induce local gingival inflammation. Mediator levels are determined from the laboratory analysis of GCF as a means of assessing periodontal status. At the end of the induction phase, stents are removed and hygiene on all teeth is restored to resolve inflammation.

In review of statistical analysis methods used in the EG studies, although gingivitis occurs over time as a steady-state inflammatory response, only approximately one-half of studies use a repeated measures ANOVA approach (Deinzer et al., 2004, 2007; Johnson et al., 1997; Waschul et al., 2003), with few studies examining the rate of increase by calculating AUC (Jepsen et al., 2003; Preisser et al., 2011; Salvi et al., 2010). Most studies that took a nonparametric approach to the analysis used a Wilcoxon-signed rank test for intra-subject comparisons and Mann-Whitney U-test or Wilcoxon rank-sum test to assess between-group differences (Giannopoulou et al., 2003; Konradsson et al., 2007; Konradsson and van Dijken, 2005; Staab et al., 2009; Tsalikis, 2010). Studies not using repeated measures analysis instead used paired t-tests to assess mean within group changes from baseline to each timepoint or 2-sample t-tests to assess between-group differences in mean levels at each timepoint (Konradsson et al., 2007; Salvi et al., 2010). Few studies mentioned using a log-transformation, or any other form of transformation, before analyzing the data. Additionally, most studies mentioned above did not touch on or address the assay LOD and how values below the limit were handled in the analysis. Studies that did mention the LOD did not indicate how values left-censored due to being below the detection limit were addressed, if at all (Deinzer et al., 2007; Johnson et al., 1997; Konradsson et al., 2007).

For the analysis of biomarker data in experimental gingivitis, most recently, a non-parametric multiple hypothesis testing approach was advocated for the analysis of repeated measures (Preisser et al., 2011) using AUC summary measures to assess the change over time in biomarker levels. Though this method has the advantage of being able to assess a large number of biomarkers relative to the number of subjects, it is unclear how to handle missing data in this context, particularly left truncation of observations due to a lower detection limit. The most common strategy for measurement data with detection limits is substitution of the value below the limit with a fraction of the limit (e.g., 0.5). Though this strategy is simplistic and easy to implement, it often distorts results (Hughes, 1999; Lubin et al., 2004) and can provide biased estimation, particularly if a large percentage of the data are left-truncated (Fang et al., 2011).

In the assessment of change over time, missing data can be a common occurrence when data are measured over multiple periods of time. For longitudinal studies involving repeated measures analysis, there can be many reasons for missing data, including nonresponse, subject dropout or, as is often times the case in the measurement of biomarker data, missingness due to assay detection limits. Missing values can have a profound influence on statistical results, including estimation of summary measures of change and hypothesis testing. If the missingness mechanism is MAR, i.e., the probability that a response is observed can only depend on the values of those other factors which have been observed, there are well developed computational methods for handling missing data under this assumption (Little and Rubin, 1987). Hughes (1999) described an EM algorithm for maximum likelihood estimation of a linear mixed effects model for estimating trends in CD4 counts over time in HIV-positive subjects, accounting for left and/or right censoring. (Lyles et al., 2000) developed a likelihood method that addresses missing data due to left truncation as well as an extension to additionally accommodate informative dropout. Thiebaut and Jacqmin-Gadda (2004)

applied a maximum likelihood approach for left-censored data based on a Marquardt algorithm (Marquardt, 1963) in HIV research. In reference to pharmacokinetic data with measurements below the quantification limit, Fang et al. (2011) developed a maximum likelihood method to estimate AUC and the ratio of two AUCs (i.e., relative exposure).

In Chapter 3, we illustrate use of a piecewise linear mixed model adjusting for left truncation as a parametric approach in the analysis of experimental gingivitis biomarker data. The disadvantage of using a parametric model is the assumptions placed on the form of the mean model and the missingness mechanism, including missing data due to the lower detection limit (Preisser et al., 2011). For this parametric approach, we illustrate our methods under the framework of a random-effects model. Currently the most utilized *ad hoc* approaches for handling of missingness due to levels below the detectable threshold include substituting the values by some fraction of the detection limit. Parametric approaches using a survival analysis framework have been implemented as an ideal approach over simple substitution methods to address the missingness (Helsel, 2006). For repeated-measures problems, procedures for fitting linear mixed effects models to data include maximum likelihood methods.

This paper presents a smooth nonlinear gamma curve-like (i.e., expected response takes the form of a gamma density function) mixed model accounting for left truncation under MAR using estimation methods that are easy to implement. A parametric model is fit to log-transformed data for 3 biomarkers representing varying degrees of truncation due to lower detection limits using 2 *ad hoc* (naive) approaches for handling non-detect values and a likelihood approach accounting for left censoring and outcomes missing at random ("ML1") from Lyles et al. (2000). The focus will be on providing direct estimation of the trends in biomarkers over time based on AUC computations and hypothesis testing for AUC to assess biomarker changes over time in a study of

experimental gingivitis.

4.2 Gamma Curve-Like Nonlinear Mixed Model

Experimental gingivitis biomarker levels are typically expected to follow one of two general patterns involving directional temporal changes (Figure 4). Experimental gingivitis biomarker levels have been described as occurring in phases, where there are distinctly different biomarker characteristics associated with GCF levels under different phases of change. For instance, for a positively sensitive (PS) biomarker, the levels are expected to increase during the induction phase to a critical value and decrease during the resolution phase (Figure 5). The beginning of the resolution phase is thought to occur at or near the time when stents are discontinued and hygiene on all teeth is reinstituted to resolve inflammation. A negatively sensitive (NS) biomarker has the opposite pattern. A gamma density function can be applied to describe these data trend as it provides a smooth fit to a set of pre-specified change points. To fix ideas, the change points are set to the measurement occasions assumed to be at 0, 1, 2, 3, 5, and 7 weeks as in, for example, Offenbacher et al. (2010).

Figure 4: Typical biomarker patterns of change over time for experimental gingivitis. Letters A, B, C, and D denote a partition of AUC for which summary measures of change can be estimated.

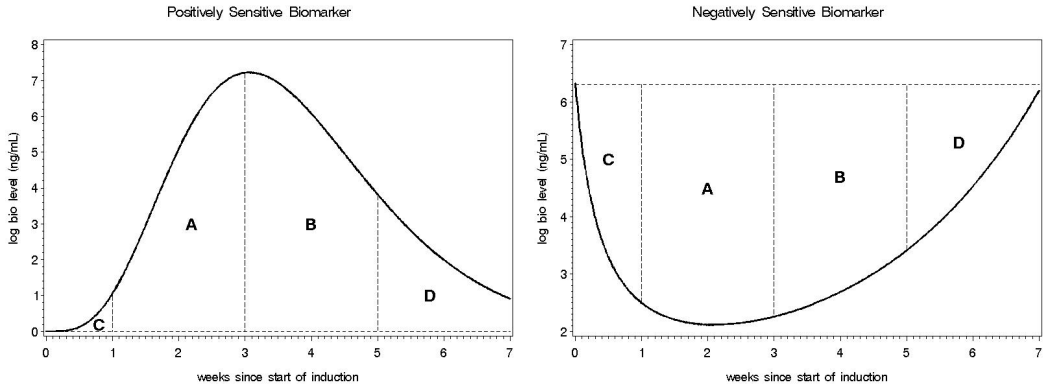
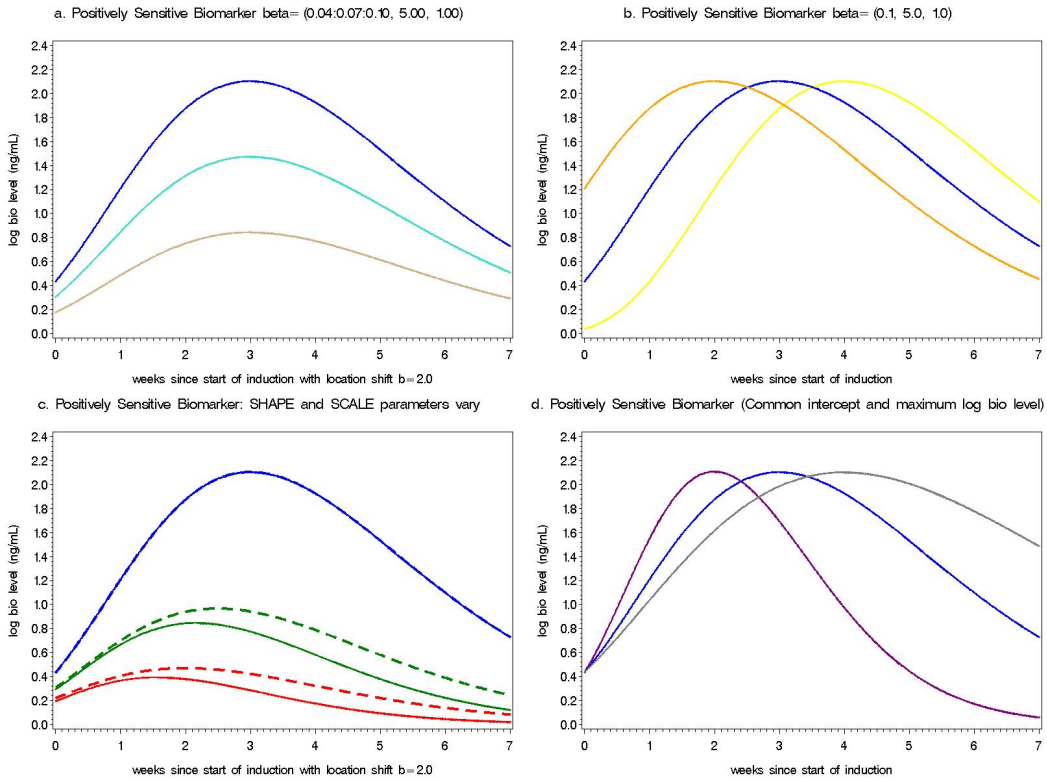


Figure 5: Patterns of change over time for positively sensitive biomarker with variation in intercept (β_0), shape (β_1), scale (β_2), and location shift (θ_1) parameters. Panel a: β_0 =0.04, 0.07, 0.10; β_1 = 5.00; β_2 = 1.00; θ_1 = 2.0. Panel b: β_0 =0.10; β_1 = 5.00; β_2 = 1.00; θ_1 = 1.0, 2.0, 3.0. Panel c (solid line): β_0 =0.10; β_1 = 4.00, 4.50, 5.00; β_2 = 1.00; θ_1 = 2.0. Panel c (dashed line): β_0 =0.10; β_1 = 5.00; β_2 = 1.00, 1.20, 1.40; θ_1 = 2.0. Panel d: β_0 =0.089, 0.10, 0.116; β_1 = 8.19, 5.00, 3.66; β_2 = 2.047, 1.00, 0.61; θ_1 = 2.0.



Given what is understood about the nature of experimental gingivitis, we assume the function should be contiguous. A gamma curve-like mixed model is defined to describe the biomarker pattern of change over the six timepoints to coincide with summary indices of change associated with experimental gingivitis described below. Let Y_{it}, t = 0, ..., T_i - 1, be repeated measures for subjects i = 1, ..., n for a fixed set of $T \geq$ max(T_i) measurement times (with possibly missing visits that are missing at random). Y_{it} is the GCF level of each biomarker (on the log base 10 scale for the application considered in this chapter). A nonlinear repeated measures model for Y_{it}, the i-th

individual's response at time t, follows a gamma-type curve of the form

$$Y_{it} = \beta_0(t+\theta_1)^{\beta_1}e^{-\beta_2(t+\theta_1)} + b_i + \epsilon_{it}, t = 0, \ldots T-1 \tag{4.1}$$

where $\beta_0 > 0$, $\theta_1 > 0$, β_1 and β_2 are unrestricted, and where $b_i \sim \mathrm{N}(0, \sigma_b^2)$ for $i = 1, \ldots, n$ subjects and $\epsilon_{it} \sim \mathrm{N}(0, \sigma_e^2)$ for $t = 0, \ldots, T_i$ are mutually independent random variables. The expected response at time t is

$$\begin{aligned} E(Y_t) &= \beta_0(t+\theta_1)^{\beta_1}e^{-\beta_2(t+\theta_1)} \\ &= \theta_0(t+\theta_1)^{\beta_1}e^{-\beta_2 t}, t = 0, \ldots T-1 \end{aligned} \tag{4.2}$$

where $\theta_0 = \beta_0 e^{-\beta_2\theta_1}$. In the absence of fixed effects covariates, the subscript i is dropped. For $t = 0$, $\mathrm{E}(Y_0) = \theta_0\theta_1^{\beta_1} > 0$, showing that $\theta_1 > 0$ with $\beta_0 > 0$ implies a positive-valued intercept. For a PS biomarker, the shape parameter $\beta_1 > 0$ and scale parameter $\beta_2 > 0$. For a NS biomarker, $\beta_1 < 0$ and $\beta_2 < 0$. Regardless of the pattern of change, the location-shift parameter θ_1 alters the values of the intercept, while not affecting the shape nor scale of the curve. Figure 5 gives examples from the family of curves given by equation (4.2), illustrating the role of each parameter in the model, and the effects of combinations of parameters. For example, Figure 5c illustrates that the extrema, in this case the maximum point on the curve, occurs at $(\beta_1 / \beta_2) - \theta_1$.

Consider the $Gamma(\beta_1, \beta_2)$ distribution function,

$$f(u) = \frac{\beta_2^{\beta_1+1}u^{\beta_1}e^{-\beta_2 u}}{\Gamma(\beta_1+1)}, u > 0, \beta_1 > 0, \beta_2 > 0 \tag{4.3}$$

which takes the form of equation (4.2) assuming θ_1 is a known constant, such as a, and with $\alpha_0 = \beta_2^{\beta_1+1}/\Gamma(\beta_1+1)$ in place of β_0. We are interested in the area under the curve given in equation (4.2) between two points t_1 and t_2 (for $t_1 < t_2$) above

the horizontal line extending out from the intercept given by

$$AUC(t_1, t_2) = \left\{ \int_{t_1}^{t_2} \beta_0 (t+a)^{\beta_1} e^{-\beta_2 (t+a)} dt \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \tag{4.4}$$

for a PS marker, and

$$AUC(t_1, t_2) = \theta_0 a^{\beta_1} [t_2 - t_1] - \left\{ \int_{t_1}^{t_2} \beta_0 (t+a)^{\beta_1} e^{-\beta_2 (t+a)} dt \right\} \tag{4.5}$$

for a NS marker. Letting $u = t + a$ and rearranging terms to express the area as a scaled difference between two *Gamma* cdfs gives

$$\begin{aligned} AUC(t_1, t_2) &= \left\{ \int_{t_1+a}^{t_2+a} \beta_0 u^{\beta_1} e^{-\beta_2 u} du \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \\ &= \left\{ \frac{\beta_0}{\alpha_0} \int_{t_1+a}^{t_2+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \\ &= \frac{\beta_0}{\alpha_0} \left[\int_0^{t_2+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du - \int_0^{t_1+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du \right] - \theta_0 a^{\beta_1} [t_2 - t_1] \end{aligned} \tag{4.6}$$

Next, define

$$\begin{aligned} E(A_i) &= AUC(1, 3) \\ E(B_i) &= AUC(3, 5) \\ E(C_i) &= AUC(0, 1) \\ E(D_i) &= AUC(5, 7) \end{aligned} \tag{4.7}$$

where A_i and C_i correspond to areas under the curve during the induction phase and B_i and D_i correspond to the resolution phase.

Four summary indices of change can now be defined as follows:

$$
\begin{aligned}
E(X_{i1}) &= E(C_i - \frac{1}{2}D_i) = AUC(0,1) - \frac{1}{2}AUC(5,7) \\
E(X_{i2}) &= E(A_i - B_i) = AUC(1,3) - AUC(3,5) \\
E(X_{i3}) &= E(Y_{i2}) = \beta_0(2+\theta_1)^{\beta_1}e^{-\beta_2(2+\theta_1)} \\
E(X_{i4}) &= E(Y_{i4} - Y_{i5}) = \beta_0(4+\theta_1)^{\beta_1}e^{-\beta_2(4+\theta_1)} - \beta_0(5+\theta_1)^{\beta_1}e^{-\beta_2(5+\theta_1)}
\end{aligned}
\tag{4.8}
$$

where X_{i1} and X_{i2} examine whether the rate of induction is the same as the rate of resolution, X_{i3} examines the rate of induction between Week 0 and Week 2, and X_{i4} examines the rate of resolution between Week 5 and Week 7 ((Preisser et al., 2011)). The statistical analysis of these variates addressing left truncation is performed based on the likelihood methods outlined in the next section.

4.3 Maximum Likelihood Estimation in the Presence of Left Truncation

In experimental gingivitis, GCF is collected at the beginning of the hygiene phase, the beginning of the induction phase, during and at the end of the induction phase, and during the resolution phase. Let $\mathbf{Y_i} = (Y_{i0}, Y_{i1}, \ldots, Y_{in_i})$. For model (4.1), the likelihood function for the parameter vector $\mathbf{\Omega} = (\beta,\ \sigma_b^2,\ \sigma_e^2)$ is

$$
L(\Omega; \mathbf{Y}) = \prod_{i=1}^{l} f(\mathbf{Y}_{ik}; \Omega), \tag{4.9}
$$

where $f(\mathbf{Y}_{ik}; \Omega) = \int_{-\infty}^{\infty} f(\mathbf{Y}|b_{0i})\, f(b_{0i})\ d\, b_{0i}$, $f(b_{0i})$ and $f(\mathbf{Y}|b_{0i})$ are the probability density functions for the random effects and responses given the random effects, respectively.

A limitation of using laboratory assays is that the biomarkers levels below the detection limit are not quantifiable. While many methods have been proposed to ad-

dress the censored data, substitution methods are among the most popular. For a biomarker, let the limit of detection be denoted by d. For "Naive1" and "Naive2" methods, GCF measurements are replaced by d and $d/2$, respectively. As indicated in Lyles et al. (2000), a detectable value contributes $f(Y_{ik}|b_{0i})$ and a non-detectable value contributes the Bernoulli probability $F_y(d|b_{0i})$, F_y is the cumulative distribution function. Let n_{i1} represent the number of detectable GCF values and $n_i - n_{i1}$ represent the non-detectable GCF values ($n_{i1} = 0, \ldots, n_i$) (Lyles et al., 2000). The complete-data likelihood can be written as

$$L(\Omega; \mathbf{Y}) = \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{n_{i1}} f(\mathbf{Y}_{ik}|b_{0i}) \right\} \left\{ \prod_{k=n_{i1}+1}^{n_i} F_Y(d|b_{0i}) \right\} f(b_{0i}) db_{0i} \right] \tag{4.10}$$

The SAS NLMIXED procedure can be used to fit this maximum likelihood function (Thiebaut and Jacqmin-Gadda, 2004). The code for fitting this maximum likelihood function to the biomarker data is provided in Appendix B. Maximum likelihood estimation by adaptive Gauss-Hermite quadrature is used to estimate $\mathbf{\Omega}$.

4.4 Example

Preisser et al. (2011) describe a study based on data previously published by Offenbacher et al. (2010) in which thirty-one inflammatory mediators within each gingival crevicular fluid (GCF) sample, including cytokines, MMPs and adipokines were studied in 22 subjects to evaluate the changes in the GCF composition over time. The study recruited and enrolled subjects with naturally occurring gingivitis, defined as bleeding upon probing present, typically in at least 10% of dental sites, as these subjects were more likely to develop experimental gingivitis in the course of the study. The course of the experiment included a 1-week hygiene phase, a 3-week induction phase using two stents and a 4-week resolution phase. Gingival crevicular fluid was collected from the

same oral sites at the beginning of the hygiene phase (or Day -7, one week prior to baseline), weekly during the induction phase (Day 0 or baseline), Day 7, 14, 21 (end of the induction phase/baseline for the resolution phase), and biweekly during the resolution phase at Day 35 and 49. At the final time point, baseline levels were expected to be restored for all biomarkers.

Gingival crevicular fluid is collected at multiple dental sites and fluid levels measured by use of different assays corresponding to different biomarkers. At specified time points, gingival crevicular fluid from the stent teeth was collected from each sample. For each biomarker, the average of two concentration measurements was considered the measurement for the particular site and time point. The goal of the experiment was to identify new candidate biomarkers that were sensitive to poor oral health care as identified by their patterns of change during induction and resolution of gingivitis. The data have been previously analyzed using parametric linear mixed modeling (Offenbacher et al., 2010) with zeros inserted for left-truncated values.

The extent of missingness or truncation for the three representative biomarkers chosen for this analysis, MMP7, MMP3, and MIP-1β, ranged from 16 (12.12%) to 36 (27.27%) observations. Approximately 1% to 12% of these missing values were due to the values being below the biomarker's detection limit. Table 6 outlines the level of missingness by biomarker.

Table 6: Amount of missingness across biomarkers

Biomarker	Left Truncated	Other	Total Missing	Observed	Total
MMP7	1 (0.76%)	15 (11.36%)	16 (12.12%)	116 (87.88%)	132
MMP3	2 (1.52%)	15 (11.36%)	17 (12.88%)	115 (87.12%)	132
MIP-1β	16 (12.12%)	20 (15.15%)	36 (27.27%)	96 (72.73%)	132
All Biomarkers	25 (4.73%)	65 (12.31%)	90 (17.05%)	438 (82.95%)	528

4.5 Results

4.5.1 Regression Estimates

Figure 6 presents the subject-level best linear unbiased predictions and population-averaged biomarker levels using the gamma curve-like mixed model (and "ML1" to account for left censoring). The MMP-7 population-averaged biomarker levels show increasing trends from the beginning of the induction phase until a peak at Week 1. After the peak, the levels decrease to a near constant rate through the end of the resolution phase. In contrast, the average biomarker levels for MMP-3 decrease from Week 0 through Week 1, then steadily increase through the end of the resolution phase. A similar trend was seen with the MIP-1β biomarker levels.

The model in equation (4.1) is fitted to the experimental gingivitis study data using the 2 *ad hoc* approaches ("Naive1" and "Naive2"). A nonlinear mixed model procedure is used to obtain the MLEs and associated standard errors under the following assumption. The minimum detection limits for the 3 biomarkers in nanograms per millilitre on the log scale are as follows: MMP7: 1.470, MMP3: 0.415, and MIP-1β: -0.143. Next, the nonlinear mixed model procedure from Section 2.2 is used to obtain the MLEs and associated standard errors for the ML approach accounting for left censoring ("ML1"). Table 7 presents the ML estimates and associated standard errors from the gamma curve-like mixed model in equation (4.1) for the 3 biomarkers based on the 3 missing data handling methods.

For each biomarker, Figure 7 shows the population-averaged estimated curves and 95% pointwise confidence bands about the mean curve constructed based on the delta method for variance estimation for the three methods. For MMP7, the 3 methods show similar mean patterns and confidence bands over time. A similar result was seen with the MMP3 population-averaged biomarker levels over time. For MIP-1β, "ML1"

population-averaged biomarker levels were lower from the beginning of the induction phase through the end of the resolution phase. The "ML1" population-averaged confidence band width fell within the band widths of both *ad hoc* approaches from the beginning of the induction phase through the end of the resolution phase. During this same timeframe, "Naive2" biomarker levels and confidence band were similar to "ML1". However, corresponding estimates for "Naive1" were different during the induction phase through the end of the resolution phase, with a higher biomarker level and more narrow confidence band width during the induction phase through the end of the resolution phase. Of all the biomarkers, MIP-1β had the more varied band widths between the different methods, likely due to having the largest percentage of missingness due to the LOD leading to higher variability.

Figure 6: Subject-level best linear unbiased predictions and population-averaged biomarker levels over time for MMP7 (upper left), MIP-1β (upper right), and MMP3 (lower center) based on the gamma curve-like mixed model.

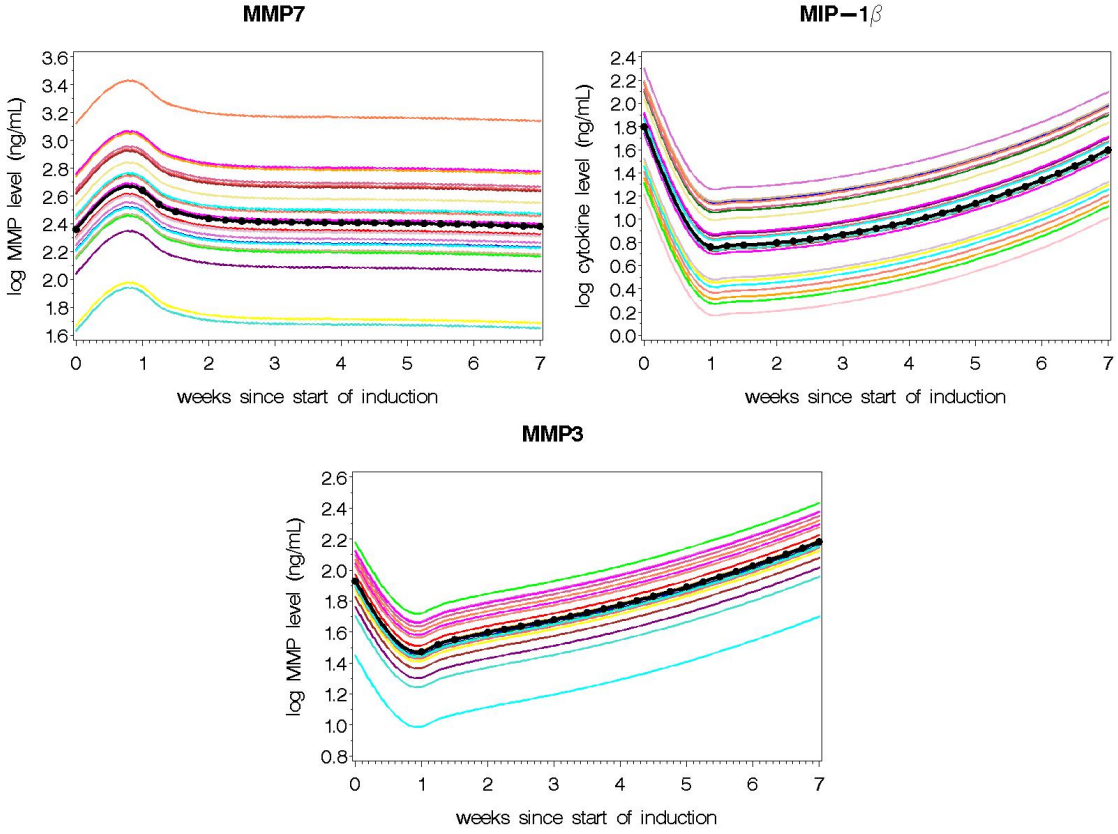
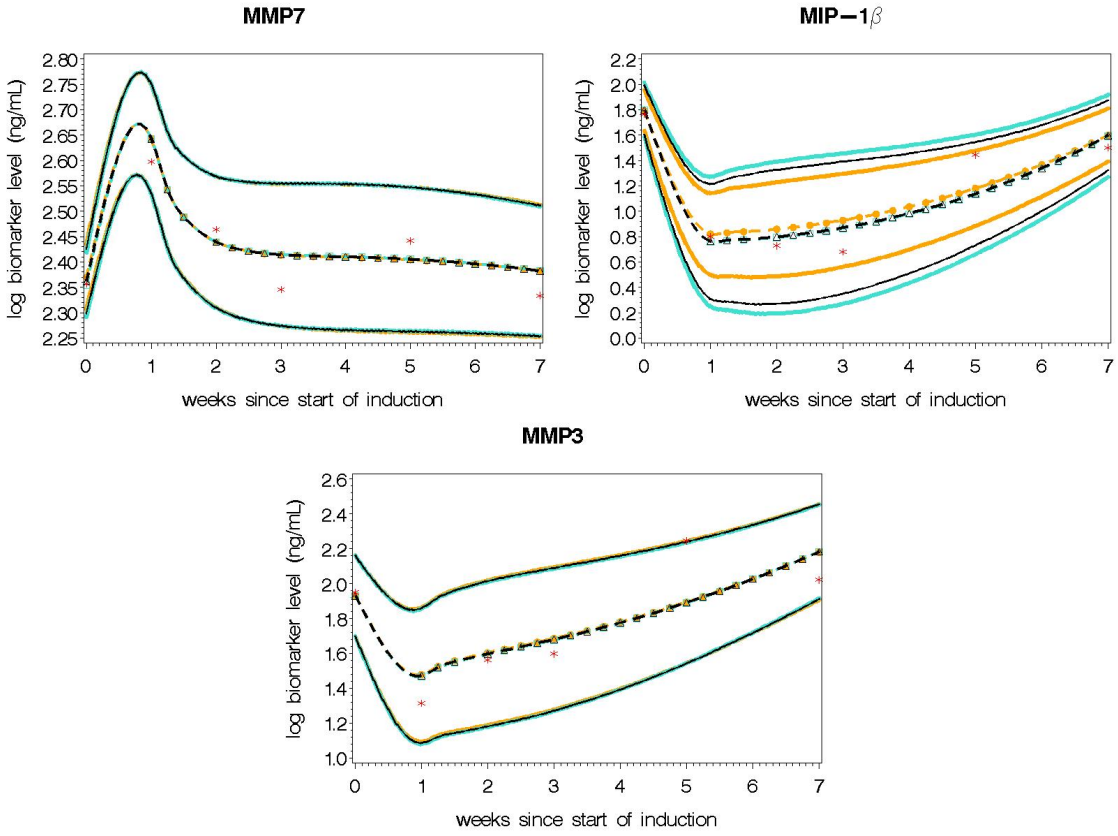


Table 7: Results from experimental gingivitis study using a nonlinear gamma curve-like mixed model: comparing ML estimates from *ad hoc* approaches with ML estimates from an approach accounting for non-detectable values

		Estimates (SEs)					
Biomarker	Method	β_0	β_1	β_2	θ_1	σ_b^2	σ_e^2
MMP7	Naive1†	1.618	0.195	0.200	0.174	0.121	0.071
		(0.401)	(0.136)	(0.109)	(0.093)	(0.040)	(0.010)
	Naive2*	1.598	0.205	0.206	0.178	0.124	0.073
		(0.408)	(0.143)	(0.113)	(0.096)	(0.041)	(0.011)
	ML1$^\diamond$	1.607	0.201	0.203	0.176	0.122	0.072
		(0.405)	(0.140)	(0.111)	(0.095)	(0.041)	(0.010)
MIP-1β	Naive1†	0.972	-0.544	-0.364	0.365	0.115	0.219
		(0.495)	(0.122)	(0.232)	(0.484)	(0.048)	(0.033)
	Naive2*	0.872	-0.580	-0.384	0.333	0.145	0.292
		(0.823)	(0.240)	(0.429)	(0.864)	(0.062)	(0.043)
	ML1$^\diamond$	0.902	-0.597	-0.397	0.352	0.143	0.291
		(0.746)	(0.214)	(0.381)	(0.721)	(0.061)	(0.048)
MMP3	Naive1†	2.108	-0.291	-0.263	0.154	0.043	0.209
		(1.044)	(0.311)	(0.225)	(0.185)	(0.026)	(0.031)
	Naive2*	2.104	-0.299	-0.268	0.156	0.048	0.218
		(1.062)	(0.320)	(0.230)	(0.188)	(0.028)	(0.032)
	ML1$^\diamond$	2.100	-0.294	-0.265	0.154	0.046	0.215
		(1.054)	(0.317)	(0.228)	(0.187)	(0.028)	(0.032)

† Substituting non-detectable values by the detection limit.
$*$ Substituting non-detectable values by half the detection limit.
$\diamond$ Accounting for non-detectable values via (4.10).

Figure 7: Population-averaged biomarker levels with 95% pointwise confidence bands over time for MMP7 (upper left), MIP-1β (upper right), and MMP3 (lower center) for the three *ad hoc* methods ("Naive1" [orange], "Naive2" [turquoise], and "ML1" [black]) based on the nonlinear gamma curve-like mixed model.



4.5.2 Estimates, Standard Errors, and 95% Confidence Intervals of Area-Under-the-Curve (A, B, C, and D)

For the summary measures of AUC, let $\boldsymbol{\theta}_1 = \mathrm{E}(\mathbf{R_i})$ where $\mathbf{R_i} = (A_i, B_i, C_i, D_i)$, $i = 1, \ldots,$ n, are $i.i.d.$ Then, based on equation (4.6), $\boldsymbol{\theta}_1$ is estimated by two incomplete intervals that can be computed by SAS with the function "CDF('GAMMA', u, b,lambda)" where $b = \beta_1 + 1$, lambda $= 1/\beta_2$, and $u = t_2 + a$ and $u = t_1 + a$, respectively.

For the nonlinear gamma mixed model described in equation (4.1), the maximum

likelihood estimation described in Section 4.4 is used to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{V_\beta} = \widehat{Var}(\widehat{\boldsymbol{\beta}})$. Estimation of $\widehat{Var}(\widehat{\boldsymbol{\theta}}_1)$ is carried out using bootstrapping resampling methods.

Table 8 presents the estimates, standard errors, and 95% CIs of AUC from the non-linear gamma curve-like mixed model for the 3 biomarkers based on *ad hoc* and "ML1" data handling methods. The confidence intervals were estimated using bootstrap sampling methods. The 25th and the 975th items from 1000 resampled AUCs compose the 95% CI for the AUC. For MMP7, the rates of induction and resolution from the "ML1" method fell within the range of the estimated rates for both naive methods, with lower estimated rates for "Naive1" and higher estimated rates for "Naive2". Similar results were seen for both MMP3. For MIP-1 β, the rates were lower than both naive methods.

Table 8: Estimates (SEs), and 95% CIs of AUC from experimental gingivitis study using a nonlinear gamma curve-like mixed model: comparisons of *ad hoc* approaches with an approach accounting for non-detectable values

		Estimates (SEs) and 95% CIs			
Biomarker	Method	*A*	*B*	*C*	*D*
MMP7	Naive1†	0.204	0.039	0.171	0.027
		(0.108)	(0.081)	(0.057)	(0.059)
		(-0.001, 0.410)	(-0.123, 0.187)	(0.065, 0.287)	(-0.090, 0.136)
	Naive2*	0.216	0.051	0.174	0.041
		(0.109)	(0.081)	(0.057)	(0.058)
		(0.008, 0.422)	(-0.111, 0.198)	(0.067, 0.289)	(-0.074, 0.149)
	ML1◇	0.210	0.045	0.173	0.034
		(0.108)	(0.081)	(0.057)	(0.059)
		(0.003, 0.416)	(-0.116, 0.193)	(0.065, 0.288)	(-0.081, 0.143)
MIP-1β	Naive1†	-1.847	-1.472	-0.479	-0.813
		(0.148)	(0.119)	(0.070)	(0.097)
		(-2.016, -1.686)	(-1.573, -1.355)	(-0.555, -0.399)	(-0.874, -0.747)
	Naive2*	-1.957	-1.589	-0.527	-0.864
		(0.157)	(0.125)	(0.074)	(0.100)
		(-2.131, -1.767)	(-1.702, -1.460)	(-0.608, -0.440)	(-0.929, -0.795)
	ML1◇	-2.004	-1.617	-0.536	-0.864
		(0.153)	(0.121)	(0.073)	(0.096)
		(-2.173, -1.821)	(-1.725, -1.495)	(-0.616, -0.451)	(-0.924, -0.799)
MMP3	Naive1†	-0.616	-0.298	-0.344	0.222
		(0.044)	(0.032)	(0.023)	(0.022)
		(-0.690, -0.520)	(-0.355, -0.230)	(-0.386, -0.295)	(0.179, 0.268)
	Naive2*	-0.625	-0.302	-0.347	0.218
		(0.047)	(0.034)	(0.024)	(0.023)
		(-0.704, -0.523)	(-0.363, -0.229)	(-0.392, -0.294)	(0.174, 0.267)
	ML1◇	-0.621	-0.303	-0.345	0.215
		(0.046)	(0.034)	(0.024)	(0.023)
		(-0.698, -0.521)	(-0.363, -0.232)	(-0.389, -0.293)	(0.171, 0.264)

† Substituting all non-detectable values by the detection limit.
∗ Substituting all non-detectable values by half the detection limit.
◇ Accounting for non-detectable values via (4.10).

4.5.3 Estimates and Standard Errors of Summary Indices of Change (X1, X2, X3, and X4) and P-values

For the summary indices of change, let $\boldsymbol{\theta}_2 = \mathrm{E}(\mathbf{X_i})$ where $\mathbf{X_i} = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, $i = 1, \ldots, \mathrm{n}$, are $i.i.d$. For each biomarker and for a method of handling missing data due to left truncation, univariate tests are defined corresponding to the 4 variates, X_{i1}, X_{i2}, X_{i3}, X_{i4}.

A univariate test for the jth element of $\boldsymbol{\theta}_1$, $H_0\colon\ \boldsymbol{\theta}_{1j} = 0$ vs. $H_1\colon\ \boldsymbol{\theta}_{1j} \neq 0$ is

$$T_j = \left(\frac{\widehat{\boldsymbol{\theta}}_{1j}}{se(\widehat{\boldsymbol{\theta}}_{1j})} \right)^2 \tag{4.11}$$

where $se(\widehat{\boldsymbol{\theta}}_{1j})$ is the estimated bootstrapped standard error. T has an asymptotic χ^2_1 distribution. Also, the overall test of any change is $H_0\colon\ \beta_1 = \beta_2 = 0$. For the overall test, the NLMIXED procedure constructs an approximate F test using the delta method (Cox, 1998).

Table 9 reports the p-values for univariate tests and overall test. For MMP3, tests for X_1, X_2, X_3, and X_4 using "ML1" were statistically significant at $\alpha = 0.05$. For this biomarker, this suggests suppression of mediator levels during the first two weeks of the induction phase and asymmetry around the end of the induction phase/beginning of the resolution phase (Week 3), and during the last two weeks of the resolution phase. Similar results were seen for the *ad hoc* approaches. For MMP7, X_1 and X_3 were statistically significant at $\alpha = 0.05$. Similar results were seen for the *ad hoc* approaches. For MIP-1β, X_2 and X_3 were significant for "ML" suggesting asymmetry about Week 3, and suppression of biomarker levels within the first two weeks of induction. Similar results were seen for the *ad hoc* approaches.

Table 9: Estimates (SEs) of summary indices of change and p-values from experimental gingivitis study using a nonlinear gamma curve-like model: comparing significance from *ad hoc* approaches with significance from ML approach accounting for non-detectable values

		Estimates (SEs) and P-values				
Biomarker	Method	X_1	X_2	X_3	X_4	Overall $P*$
MMP7	Naive1†	0.158	0.165	1.218	0.136	<0.0001
		(0.064)	(0.137)	(0.140)	(0.105)	
		0.014	0.227	<0.0001	0.194	
	Naive2*	0.154	0.165	1.196	0.136	<0.0001
		(0.064)	(0.137)	(0.140)	(0.103)	
		0.017	0.229	<0.0001	0.189	
	ML1◇	0.156	0.165	1.207	0.136	<0.0001
		(0.064)	(0.137)	(0.140)	(0.104)	
		0.015	0.228	<0.0001	0.192	
MIP-1β	Naive1†	-0.072	-0.376	0.656	0.098	<0.0001
		(0.043)	(0.098)	(0.120)	(0.082)	
		0.092	0.0001	<0.0001	0.229	
	Naive2*	-0.095	-0.368	0.582	0.090	<0.0001
		(0.046)	(0.108)	(0.121)	(0.080)	
		0.040	0.0006	<0.0001	0.262	
	ML1◇	-0.104	-0.387	0.591	0.092	<0.0001
		(0.045)	(0.104)	(0.118)	(0.077)	
		0.020	0.0002	<0.0001	0.231	
MMP3	Naive1†	-0.456	-0.318	1.496	0.194	0.002
		(0.025)	(0.055)	(0.133)	(0.093)	
		<0.0001	<0.0001	<0.0001	0.038	
	Naive2*	-0.456	-0.323	1.484	0.195	0.002
		(0.027)	(0.058)	(0.133)	(0.093)	
		<0.0001	<0.0001	<0.0001	0.035	
	ML1◇	-0.453	-0.318	1.487	0.194	0.002
		(0.026)	(0.057)	(0.133)	(0.093)	
		<0.0001	<0.0001	<0.0001	0.037	

† Substituting all non-detectable values by the detection limit.
∗ Substituting all non-detectable values by half the detection limit.
◇ Accounting for non-detectable values via (4.10).
∗ $\beta_1 = \beta_2 = 0$.

4.6 Discussion and Conclusion

We have presented a nonlinear gamma curve-like mixed model to provide direct estimation of AUC based on the trends in biomarkers over time. The method was illustrated using data from an experimental gingivitis study. Our objective was two-fold: (1) to outline current methods in the experimental gingivitis literature for analyzing repeated measures biomarker data and motivate a biologically plausible parametric model that provides estimation of change in biomarker levels via a smooth function and (2) to present a maximum likelihood methodology for addressing left truncation while illustrating the potential negative effect on population level estimates and hypothesis testing when simple imputations methods are used to correct left censoring. The maximum likelihood approach suggested here provides a straightforward way for analysts to evaluate longitudinal data with missing observations due to left truncation.

Although the proposed methodology was motivated by trends in repeated measures experimental gingivitis data, the approach can be applied to other areas with a similar longitudinal setup and multiple outcomes. Additionally, the likelihood framework presented in this article can be extended to more fixed and random effects; however, special consideration should be given for studies of small sample size like the experimental gingivitis study when introducing more than two random effects.

In conclusion, the examples presented here illustrate the potential for using a smooth gamma curve-like parametric statistical model as a potential biologically plausible model for analyzing AUC as a summary measure for estimating change in biomarker levels. However, emphasis is placed on the importance of using an appropriate means of taking into account left truncation in order to avoid inaccurate interpretation of results from population parameters, random effects, and hypothesis tests. Overall, using the maximum likelihood procedure can provide a uniform means of analyzing biomarker data in periodontal disease and other conditions.

CHAPTER 5: COMPARISON OF STATISTICAL METHODS

5.1 Background and Introduction

In periodontal research, repeated measures analysis is sometimes used to identify molecular mediators of inflammation that can be induced to significant change over time. In Experimental Gingivitis (EG), for example, biomarker levels change when the benefits of toothbrushing are withheld during an induction phase, then restored during a resolution phase. The pattern of change over time of biomarker levels associated with gingivitis can be varied. The statistical methodology utilized to detect that change should address left-truncation, which occurs in EG studies when a biomarker level falls below its lower limit of detection. Parametric repeated measures models are useful for characterizing the nature of change over time, particularly as they easily accommodate both truncated and missing biomarker data. In the presence of outliers and/or highly skewed data, however, nonparametric methods may be preferred due to their robustness and reliance on fewer assumptions. In this article, a simulation study was performed to compare parametric and nonparametric hypothesis testing methods for detection of any overall biomarker level change in hypothetical EG studies with six repeated measures and n=20, 30 or 50 subjects. For mild left-truncation (10%), the correctly specified mixed effects model that explicitly accounts for left-truncation maintained the nominal Type I error for all sample sizes; for severe left-truncation (25%), it maintained the nominal Type I error only for n=50. In comparison, a multivariate Wilxocon signed-rank (WSR) permutation test, directed at a set of area-under-the-curve indices summarizing change, maintained the nominal Type I error for n=30 and n=50 under

mild or severe left-truncation. It also suffered a maximum power loss of only 7% relative to the parametric test for n=20 and had similar power for n=30 and n=50. Nonparametric methods are generally recommended for EG studies.

In periodontal research, there is often interest in identifying molecular mediators of inflammation that can be induced to significant change over time as well as quantify the magnitude of the change within an experimental repeated measures study. Dental researchers often employ experimental gingivitis (EG) to realize a better understanding of the hosts immune response to periodontal pathogens (Deinzer et al., 2007). As such, EG is widely used for elucidating the inflammatory response to undisturbed dental biofilm accumulation in a well-controlled pretest-posttest experimental design framework. The course of EG includes a hygiene phase, an induction phase, and resolution phase. Gingivitis is induced by withholding tooth brushing by the use of intraoral acrylic stents that cover selected teeth in each arch during tooth brushing to induce local gingival inflammation. Mediator levels are determined from the laboratory analysis of gingival crevicular fluid (GCF) as a means of assessing periodontal status. At the end of the induction phase, stents are removed and hygiene on all teeth is restored to resolve inflammation.

In review of statistical analysis methods used in the EG studies, although gingivitis occurs over time as a steady-state inflammatory response, only approximately one-half of studies use a repeated measures ANOVA approach (Deinzer et al., 2004, 2007; Johnson et al., 1997; Waschul et al., 2003), with few studies examining the rate of increase by calculating AUC (Jepsen et al., 2003; Preisser et al., 2011; Salvi et al., 2010). Most studies that took a nonparametric approach to the analysis used a Wilcoxon-signed rank test for intra-subject comparisons and Mann-Whitney U-test or Wilcoxon rank-sum test to assess between-group differences (Giannopoulou et al., 2003; Konradsson et al., 2007; Konradsson and van Dijken, 2005; Staab et al., 2009; Tsalikis, 2010). Studies

not using repeated measures analysis instead used paired t-tests to assess mean within group changes from baseline to each timepoint or 2-sample t-tests to assess between-group differences in mean levels at each timepoint (Konradsson et al., 2007; Salvi et al., 2010). Few studies mentioned using a log-transformation, or any other form of transformation, before analyzing the data. Additionally, most studies mentioned above did not touch on or address the assay lower limit of detection (LOD) and how values below the limit were handled in the analysis. Studies that did mention the LOD did not indicate how values left-censored due to being below the detection limit were addressed, if at all (Deinzer et al., 2007; Johnson et al., 1997; Konradsson et al., 2007).

In the assessment of change over time, missing data can be a common occurrence when data are measured over multiple periods of time. For longitudinal studies involving repeated measures analysis, there can be many reasons for missing data, including nonresponse, subject dropout or, as is often times the case in the measurement of biomarker data, missingness due to assay detection limits. Missing values can have a profound influence on statistical results, including estimation of summary measures of change and hypothesis testing. If the missingness mechanism is missing at random (MAR), i.e., the probability that a response is observed can only depend on the values of those other factors which have been observed, there are well developed computational methods for handling missing data under this assumption (Little and Rubin, 1987). Hughes (1999) described an EM algorithm for maximum likelihood estimation of a linear mixed effects model for estimating trends in CD4 counts over time in HIV-positive subjects, accounting for left and/or right censoring. Lyles et al. (2000) developed a likelihood method that addresses missing data due to left truncation as well as an extension to accommodate informative dropout. Thiebaut and Jacqmin-Gadda (2004) applied a maximum likelihood approach for left-censored data based on a Marquardt algorithm (Marquardt, 1963) in HIV research. In reference to pharmacokinetic data

with measurements below the quantification limit, Fang et al. (2011) developed a maximum likelihood method to estimate AUC and the ratio of two AUCs (i.e., relative exposure).

In papers 1 and 2, we illustrate use of a piecewise linear mixed model and a nonlinear gamma-curve-like mixed model, respectively, adjusting for left truncation as parametric approaches in the analysis of experimental gingivitis biomarker data. An advantage of the parametric approaches is that they can be easily used to provide estimation of trends in biomarker levels over time while accommodating missing or truncated data. However, because nonparametric methods have weaker assumptions and are less sensitive to outliers, they may be preferred over their parametric counterparts. For the analysis of biomarker data in experimental gingivitis, recently, a nonparametric multiple hypothesis testing approach was advocated for the analysis of repeated measures (Preisser et al., 2011) using AUC summary measures to assess the change over time in biomarker levels. Though this method has the advantage of being able to assess a large number of biomarkers relative to the number of subjects, it is unclear how to handle missing data in this context, particularly left truncation of observations due to a lower detection limit. It has been suggested by Preisser et al. (2011) that permutation tests alleviate many problems encountered when deterministic imputation methods are utilized; however, further evaluation is needed. The primary motivation for nonparametric analyses is that analysis of ranks provide tests less sensitive to outliers and Gaussian distribution assumptions than provided by parametric analysis.

This paper compares parametric mixed models accounting for left truncation under MAR with nonparametric multivariate analysis methods. The methods are applied to log-transformed data simulated from the "true" model with 2 different levels of truncation due to lower detection limits. The focus will be on evaluating size of test based on the "true" model and evaluating power based on testing for any change in small

or large-sized experimental gingivitis studies. An evaluation of whether the nonparametric method is robust to left truncation and crude single imputation methods is also provided.

5.2 Statistical Methods

5.2.1 Piecewise Linear Mixed Model

Experimental gingivitis biomarker levels generally follow one of four patterns involving a directional change at a particular point. Experimental gingivitis biomarker levels have been described as occurring in phases, where there are distinctly different biomarker characteristics associated with GCF levels under different phases of change. As an example, for a positively sensitive biomarker, the levels are expected to increase during the induction phase to a critical value and decrease during the resolution phase. The beginning of the resolution phase is thought to occur at or near the time when stents are discontinued and hygiene on all teeth is reinstituted to resolve inflammation. Piecewise linear regression is applied to describe these trend data as it is a form of regression that allows multiple linear segments to be fit to the data for a set of prespecified change points. The change points are set to the measurement occasions at 0, 1, 2, 3, 5, and 7 weeks as in, for example, Offenbacher et al. (2010).

Given what is understood about the nature of experimental gingivitis, we assume the function should be contiguous. A piecewise linear regression model is defined to describe the biomarker pattern of change over the six timepoints to coincide with summary indices of change associated with experimental gingivitis described below. Let Y_{ik} be the GCF level of each biomarker (on the log base 10 scale for the application considered in this article) for the ith subject at the kth time point, for $k = 0, \ldots, 5$, which are $t_{i0} = 0, t_{i1} = 1, t_{i2} = 2, t_{i3} = 3, t_{i4} = 5, t_{i5} = 7$ weeks. The general form of the model

can be written as

$$Y_{ik} = \mathbf{Z}_\mathbf{i}'\boldsymbol{\beta} + b_{0i} + \boldsymbol{\epsilon}_{ik} \tag{5.1}$$

where $b_{0i} \sim N(0, \sigma_b^2)$ are subject-specific random intercepts and $\boldsymbol{\epsilon}_{ik} \sim N(0, \sigma^2)$ are random errors, with b_{0i}, $i = 1, \ldots, n$ and ϵ_{ik}: $i = 1, \ldots, n$; $k = 0, \ldots, 5$ mutually independent. We assume that $\mathbf{Z_i}$ is a vector of explanatory variables that are functions of time, including an intercept. In principle, we could introduce a model with multiple variance components; however, because most EG studies are of small sample size, these studies would not likely allow estimation of more than 2 variance components.

In model (3.4), $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$ and $\mathbf{Z_i} = (Z_{i0}, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5})'$ is defined to give conjoined piecewise linear segments. To parameterize the model, define for $k = 0, 1, 2, 3, 4, 5$:

$$z_{i0} = 1$$

$$z_{i1} = t_{ik}$$

$$z_{i2} = (t_{ik} - 1)\mathrm{I}(t_{ik} > 1)$$

$$z_{i3} = (t_{ik} - 2)\mathrm{I}(t_{ik} > 2)$$

$$z_{i4} = (t_{ik} - 3)\mathrm{I}(t_{ik} > 3)$$

$$z_{i5} = (t_{ik} - 5)\mathrm{I}(t_{ik} > 5)$$

The summaries of AUC (with baseline adjustment) can be estimated in terms of βs in

the following manner:

$$
\begin{aligned}
E(A_i) &= E(Y_{i1} + 2Y_{i2} + Y_{i3})/2 - 2\beta_0 = \frac{1}{2}(8\beta_1 + 4\beta_2 + \beta_3) \\
E(B_i) &= E(Y_{i3} + Y_{i4}) - 2\beta_0 = 8\beta_1 + 6\beta_2 + 4\beta_3 + 2\beta_4 \\
E(C_i) &= E[(Y_{i0} + Y_{i1})/2] - \beta_0 = \frac{1}{2}\beta_1 \\
E(D_i) &= E(Y_{i4} + Y_{i5}) - 2\beta_0 = 12\beta_1 + 10\beta_2 + 8\beta_3 + 6\beta_4 + 2\beta_5
\end{aligned} \tag{5.2}
$$

Four summary indices of change can now be defined as follows:

$$
\begin{aligned}
E(X_{i1}) &= E(C_i - \frac{1}{2}D_i) = \frac{1}{2}E[Y_{i1} - Y_{i4} - Y_{i5}] \\
&= -\frac{11}{2}\beta_1 - 5\beta_2 - 4\beta_3 - 3\beta_4 - \beta_5 \\
E(X_{i2}) &= E(A_i - B_i) = \frac{1}{2}E(Y_{i1} + 2Y_{i2} - Y_{i3} - 2Y_{i4}) \\
&= -4\beta_1 - 4\beta_2 - \frac{7}{2}\beta_3 - 2\beta_4 \\
E(X_{i3}) &= E(Y_{i2}) = 2\beta_1 + \beta_2 \\
E(X_{i4}) &= E(Y_{i4} - Y_{i5}) = -2\beta_1 - 2\beta_2 - 2\beta_3 - 2\beta_4 - 2\beta_5
\end{aligned} \tag{5.3}
$$

The statistical analysis of these variates addressing left truncation is performed based on the likelihood methods outlined in Section 5.3.3.

5.2.2 Gamma Curve-Like Mixed Model

Typical biomarker levels follow one of several trends, where 2 of the general patterns involve a directional temporal change. A gamma density function is applied to describe these data trend as it provides a smooth fit to a set of pre-specified change points.
A gamma curve-like mixed model is defined to describe the biomarker pattern of change over the six timepoints to coincide with summary indices of change associated with experimental gingivitis described below. Let Y_{it}, $t = 0, \ldots, T_i$ - 1, be repeated

measures for subjects $i = 1, \ldots, n$ for a fixed set of $T \geq \max(T_i)$ measurement times (with possibly missing visits that are missing at random). Y_{it} is the GCF level of each biomarker (on the log base 10 scale for the application considered in this article). A nonlinear repeated measures model for positive-valued Y_{it}, the i-th individual's response at time t, follows a gamma-type curve of the form

$$Y_{it} = \beta_0(t+\theta_1)^{\beta_1} e^{-\beta_2(t+\theta_1)} + b_i + \epsilon_{it}, t = 0, \ldots T-1 \tag{5.4}$$

where $\beta_0 > 0$, $\theta_1 > 0$, β_1 and β_2 are unrestricted, and where $b_i \sim \mathrm{N}(0, \sigma_b^2)$ for $i = 1, \ldots, n$ subjects and $\epsilon_{it} \sim \mathrm{N}(0, \sigma_e^2)$ for $t = 0, \ldots, T_i$ are mutually independent random variables. The expected response at time t is

$$\begin{aligned} E(Y_t) &= \beta_0(t+\theta_1)^{\beta_1} e^{-\beta_2(t+\theta_1)} \\ &= \theta_0(t+\theta_1)^{\beta_1} e^{-\beta_2 t}, t = 0, \ldots T-1 \end{aligned} \tag{5.5}$$

where $\theta_0 = \beta_0 e^{-\beta_2\theta_1}$. In the absence of fixed effects covariates, the subscript i is dropped. For a PS biomarker, the shape parameter $\beta_1 > 0$ and scale parameter $\beta_2 > 0$. For a NS biomarker, $\beta_1 < 0$ and $\beta_2 < 0$.

Consider the $Gamma(\beta_1, \beta_2)$ distribution function,

$$f(u) = \frac{\beta_2^{\beta_1+1} u^{\beta_1} e^{-\beta_2 u}}{\Gamma(\beta_1+1)}, u > 0, \beta_1 > 0, \beta_2 > 0 \tag{5.6}$$

which takes the form of equation (5.5) assuming θ_1 is a known constant, such as a, and with $\alpha_0 = \beta_2^{\beta_1+1} / \Gamma(\beta_1 + 1)$ in place of β_0. We are interested in the area under the curve given in equation (5.5) between two points t_1 and t_2 (for $t_1 < t_2$) above

the horizontal line extending out from the intercept given by

$$AUC(t_1, t_2) = \left\{ \int_{t_1}^{t_2} \beta_0 (t+a)^{\beta_1} e^{-\beta_2 (t+a)} dt \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \tag{5.7}$$

for a PS biomarker, and

$$AUC(t_1, t_2) = \theta_0 a^{\beta_1} [t_2 - t_1] - \left\{ \int_{t_1}^{t_2} \beta_0 (t+a)^{\beta_1} e^{-\beta_2 (t+a)} dt \right\} \tag{5.8}$$

for a NS biomarker. Letting $u = t + a$ and rearranging terms to express the area as a scaled difference between two $Gamma$ cdfs gives

$$\begin{aligned} AUC(t_1, t_2) &= \left\{ \int_{t_1+a}^{t_2+a} \beta_0 u^{\beta_1} e^{-\beta_2 u} du \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \\ &= \left\{ \frac{\beta_0}{\alpha_0} \int_{t_1+a}^{t_2+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du \right\} - \theta_0 a^{\beta_1} [t_2 - t_1] \\ &= \frac{\beta_0}{\alpha_0} \left[\int_0^{t_2+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du - \int_0^{t_1+a} \alpha_0 u^{\beta_1} e^{-\beta_2 u} du \right] - \theta_0 a^{\beta_1} [t_2 - t_1] \end{aligned} \tag{5.9}$$

Next, define $E(A_i) = AUC(1,3)$, $E(B_i) = AUC(3,5)$, $E(C_i) = AUC(0,1)$, and $E(D_i) = AUC(5,7)$ where A_i and C_i correspond to areas under the curve during the induction phase and B_i and D_i correspond to the resolution phase.

The rationale for the choice of a gamma-like model was two-fold. Firstly, for the biomarker analysis, a smooth model is more biologically plausible for analyzing AUC as a summary measure for estimating change in biomarker levels. Secondly, this model would allow us to assess the maximum likelihood approach for evaluating longitudinal data with missing observations due to left truncation.

5.2.3 Maximum Likelihood Estimation in the Presence of Left Truncation

In experimental gingivitis, GCF is collected at the beginning of the hygiene phase, the beginning of the induction phase, during and at the end of the induction phase, and during the resolution phase. Let $\mathbf{Y_i} = (Y_{i0}, Y_{i1}, \ldots, Y_{in_i})$. For model (5.4), let $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \theta_1)$. For models (5.1, 5.4), the likelihood function for the parameter vector $\mathbf{\Omega} = (\beta, \sigma_b^2, \sigma_e^2)$ is

$$L(\Omega; \mathbf{Y}) = \prod_{i=1}^{l} f(\mathbf{Y}_{ik}; \Omega), \tag{5.10}$$

where $f(\mathbf{Y}_{ik}; \Omega) = \int_{-\infty}^{\infty} f(\mathbf{Y}|b_{0i}) f(b_{0i}) \, d\, b_{0i}$, $f(b_{0i})$ and $f(\mathbf{Y}|b_{0i})$ are the probability density functions for the random effects and responses given the random effects, respectively.

A limitation of using laboratory assays is that the biomarkers levels below the detection limit are not quantifiable. While many methods have been proposed to address the censored data, substitution methods are among the most popular. For a biomarker, let the limit of detection be denoted by d. For 2 ad hoc (naive) approaches for handling non-detect values "Naive1" and "Naive2", GCF measurements are replaced by d and $d/2$, respectively. Alternatively, as indicated in Lyles et al. (2000), a detectable value contributes $f(Y_{ik}|b_{0i})$ and a non-detectable value contributes the Bernoulli probability $F_y(d|b_{0i})$, F_y being the cumulative distribution function. Let n_{i1} represent the number of detectable GCF values and $n_i - n_{i1}$ represent the non-detectable GCF values $(n_{i1} = 0, \ldots, n_i)$ (Lyles et al., 2000). The complete-data likelihood can be written as

$$L(\Omega; \mathbf{Y}) = \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} \left\{ \prod_{k=1}^{n_{i1}} f(\mathbf{Y}_{ik}|b_{0i}) \right\} \left\{ \prod_{k=n_{i1}+1}^{n_i} F_Y(d|b_{0i}) \right\} f(b_{0i}) db_{0i} \right] \tag{5.11}$$

Maximum likelihood estimation by adaptive Gauss-Hermite quadrature is used to es-

timate $\boldsymbol{\Omega}$.

5.2.4 Hypothesis Testing

For the piecewise model, the maximum likelihood estimation previously described is used to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{V_\beta} = \widehat{Var}(\widehat{\boldsymbol{\beta}})$. Estimation is carried out using the SAS Version 9.2 IML procedure (SAS, Inc., Cary, North Carolina). For the summary indices of change, let $\boldsymbol{\theta}_2 = \mathrm{E}(\mathbf{X_i})$ where $\mathbf{X_i} = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, $i = 1, \ldots,$ n, are $i.i.d$. Then $\boldsymbol{\theta}_2 = \mathrm{D}\boldsymbol{\beta}$ and

$$D = \begin{pmatrix} -\frac{11}{2} & -5 & -4 & -3 & -1 \\ -4 & -4 & -\frac{7}{2} & -2 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ -2 & -2 & -2 & -2 & -2 \end{pmatrix}$$

A multivariate test of $\boldsymbol{\theta}_2$, $H_0 : \boldsymbol{\theta}_2 = 0$ vs. $H_1 : \boldsymbol{\theta}_2 \neq 0$ is

$$T = (\boldsymbol{D}\widehat{\boldsymbol{\beta}})'[\boldsymbol{D}\widehat{V_\beta}\boldsymbol{D}']^{-1}(\boldsymbol{D}\widehat{\boldsymbol{\beta}}) \tag{5.12}$$

has an asymptotic χ^2_4 distribution.

For the nonlinear gamma mixed model, define $\beta^{(G)} = (\beta_1, \beta_2)$. The maximum likelihood estimation previously described is used to obtain $\widehat{\boldsymbol{\beta}}^{(G)}$ and Var($\widehat{\boldsymbol{\beta}}^{(G)}$). The overall test of any change is H_0: $\beta_1 = \beta_2 = 0$. For the overall test, the NLMIXED procedure constructs an approximate F test using the delta method (Cox, 1998).

For the nonparametric analysis, a four-variate Wilcoxon Signed Rank Test was developed to examine the four variates $\mathbf{X_{ij}} = (X_{ij1}, X_{ij2}, X_{ij3}, X_{ij4})'$ simultaneously for departure from their null median values of 0. Details of the asymptotic and permutation hypothesis testing procedures have been previously described (Preisser et al., 2011).

5.3 Simulation Studies

We conducted simulations to assess performance of six statistical methods. The primary goal was to test the ability of each method to achieve advertised size of test and power based on the "true" model and various scenarios for the number of data points. Ultimately we are interested in evaluating how the six methods perform in the presence of left truncated values; therefore, we created simulated datasets based on two different values for the lower limits of detection. Therefore, the scope of the simulation was limited in testing and evaluating missingness due to left truncation.

The continuous outcomes Y_{it}, representing the gingival crevicular fluid level of a biomarker (on the log base 10 scale) were generated based on specifying model parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \theta_1)'$, randomly generating subject-specific random intercepts (b_{0i}) and random errors (ϵ_{it}) at time t, and forcing simulated outcomes below the detection limit to be left censored. We assumed the nonlinear gamma curve-like mixed model in equation (5.4) to describe a biomarker's pattern of change over time with $b_{0i} \sim^{iid} N(0, 0.1)$ and $\epsilon_{it} \sim^{iid} N(0, 0.3)$. The following values were assigned as parameter coefficients: $\beta_0 = 1.6; \beta_1 = 0.4; \beta_2 = 0.2$; and $\theta_1 = 0.2$. Data were considered left-truncated if $Y_{it} < d$, where d (in nanograms per millilitre on the log scale) was assigned as 0.272 and 0.687, respectively such that (1) 10% and (2) 25% of the data were left-truncated. To assess power, data were generated from the "true" model under alternative $\boldsymbol{\beta} \neq 0$. For size of test, the data were generated under H_0 model $Y_{ik} = \beta_0 + b_{0i} + \epsilon_{ik}$.

The following models and procedures for addressing left-truncation were compared:

a. Piecewise linear regression with random intercept and left-truncated data replaced by half the limit of detection ("Naive2")[PW Naive2], $Y_{ik} = \mathbf{Z}_\mathbf{i}'\beta + b_{0i} + \epsilon_{ik}$

b. Piecewise linear regression with random intercept and maximum likelihood for

left-truncated data [PW ML1]

c. Gamma-curve-like nonlinear model with random intercept and left-truncated data replaced by $d/2$ [Gamma Naive2]

d. Gamma-curve-like nonlinear model with random intercept and maximum likelihood for left-truncated data [Gamma ML1]

e. Wilcoxon signed-rank test (Preisser et al., 2011), asymptotic [WSR Asy]

f. Wilcoxon signed-rank test (Preisser et al., 2011), permutation [WSR Perm].
For any record where the result of the test was rejected at $\alpha = 0.05$, a value of 1 was be assigned; else 0 was assigned. Size or power was the proportion of results with value of 1.

Size and power of tests were evaluated for the null hypothesis of no change in mean (or median) response versus the alternative hypothesis of any change. For the piecewise models, the multivariate test of any change tests H_0: $E(X_1) = E(X_2) = E(X_3) = E(X_4) = 0$, χ^2_4. For the gamma models, the overall test of any change tests H_0: $\beta_1 = \beta_2 = 0$, χ^2_2. For the Wilcoxon signed-rank tests, the test of any change tests H_0: $\text{med}\,(X_1) = \text{med}\,(X_2) = \text{med}\,(X_3) = \text{med}\,(X_4)$ (all medians simultaneously equal 0), χ^2_4. For each scenario, the simulations were run 1000 times on increasing sample sizes of N=20, 30, and 50.

5.4 Simulation Results

The simulation results (Table 10, 11, and 12 showed that all six methods performed better with increasing sample sizes. At N = 20, the Gamma-curve-like nonlinear model using ML methods ("ML1") to address left-truncated, size of test achieved the expected α when no more than 10% of the data were below the detection limit. Moreover, this

method did outperform both the piecewise and WSR approaches, which failed to achieve the advertised size of test. The piecewise methods had greater power than the other methods; however, these results must be interpreted with caution given the fact that the piewcewise models are misspecified models. At N $\geq$ 30, when no more than 10% of the data were missing due to the detection limit, the nominal test was achieved and power was high for the Gamma-curve-like nonlinear model using ML methods. The results from the percentage of measurements below the detection limit being 25% showed deflated test size in the simple substitution approach for Gamma Naive2, where left-truncated observations are replaced by the half the detection limit, and illustrated the effect of increased missingness due to values being below the detection limit on power. The effect on the power is consistent with previous simulation studies that showed the significant effect of increased censoring level on decreasing power to detect significant relationships (Thompson and Nelson, 2003). It is noteworthy that both WSR approaches performed well at N $\geq$ 30, with the permutation test method displaying slightly decreased power relative to the asymptotic approach. Based on these results, it can be inferred that when the sample size is large (N $\geq$ 30), permutation tests do alleviate many problems encountered when deterministic imputation methods are utilized as suggested by Preisser et al. (2011).

Table 10: Results from 6 approaches for analyzing experimental gingivitis data based on 1000 datasets simulated from a gamma-curve model with random intercept and 10% left-truncation

	N=20		N=30		N=50	
	Size	Power	Size	Power	Size	Power
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
PW Naive2$*$	0.011 (0.003)	0.950 (0.007)	0.016 (0.004)	0.997 (0.002)	0.018 (0.004)	1.00
PW ML1$\diamond$	0.010 (0.003)	0.950 (0.007)	0.016 (0.004)	0.997 (0.002)	0.018 (0.004)	1.00
Gamma Naive2$*$	0.053 (0.007)	0.906 (0.009)	0.060 (0.008)	0.982 (0.004)	0.065 (0.008)	1.00
Gamma ML1$\diamond$	0.048 (0.007)	0.899 (0.010)	0.051 (0.007)	0.979 (0.005)	0.056 (0.007)	1.00
WSR Asy	0.018 (0.004)	0.834 (0.012)	0.045 (0.007)	0.983 (0.004)	0.046 (0.007)	1.00
WSR Perm$\triangle$	0.038 (0.006)	0.883 (0.010)	0.045 (0.006)	0.952 (0.007)	0.045 (0.006)	0.977 (0.005)

Asy = Asymptotic, Perm = Permutation, PW = Piecewise, WSR = Wilcoxon Signed-Rank.
$*$ Substituting non-detectable values by half the detection limit.
$\diamond$ Accounting for non-detectable values via (3.10).
$\triangle$ Based on monte carlo sampling 150,000 of 2^N possible permutations.

Table 11: Results from 6 approaches for analyzing experimental gingivitis data based on 1000 datasets simulated from a gamma-curve model with random intercept and 25% left-truncation

	N=20		N=30		N=50	
	Size	Power	Size	Power	Size	Power
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
PW Naive2$*$	0.018 (0.004)	0.933 (0.008)	0.015 (0.004)	0.997 (0.002)	0.014 (0.004)	1.00
PW ML1$\diamond$	0.014 (0.004)	0.929 (0.008)	0.016 (0.004)	0.995 (0.002)	0.013 (0.004)	1.00
Gamma Naive2$*$	0.048 (0.007)	0.901 (0.009)	0.059 (0.007)	0.980 (0.004)	0.062 (0.004)	1.00
Gamma ML1$\diamond$	0.036 (0.006)	0.875 (0.010)	0.040 (0.006)	0.973 (0.005)	0.051 (0.007)	1.00
WSR Asy	0.022 (0.005)	0.796 (0.013)	0.047 (0.007)	0.970 (0.005)	0.047 (0.007)	1.00
WSR Perm$\triangle$	0.041 (0.006)	0.856 (0.011)	0.047 (0.007)	0.930 (0.008)	0.045 (0.006)	0.974 (0.005)

Asy = Asymptotic, Perm = Permutation, PW = Piecewise, WSR = Wilcoxon Signed-Rank.
$*$ Substituting non-detectable values by half the detection limit.
$\diamond$ Accounting for non-detectable values via (3.10).
$\triangle$ Based on monte carlo sampling 150,000 of 2^N possible permutations.

Table 12: Results from 4 approaches for analyzing experimental gingivitis data based on 1000 datasets simulated from a gamma-curve model with random intercept and without left truncation

	N=20		N=30		N=50	
	Size	Power	Size	Power	Size	Power
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Piecewise	0.012 (0.003)	0.957 (0.006)	0.016 (0.004)	0.997 (0.002)	0.018 (0.004)	1.00
Gamma	0.052 (0.004)	0.904 (0.009)	0.059 (0.007)	0.980 (0.004)	0.063 (0.008)	1.00
WSR Asy	0.020 (0.004)	0.824 (0.012)	0.042 (0.006)	0.983 (0.004)	0.046 (0.007)	1.00
WSR Perm△	0.038 (0.006)	0.885 (0.010)	0.043 (0.006)	0.951 (0.007)	0.040 (0.006)	0.941 (0.007)

Asy = Asymptotic, Perm = Permutation, PW = Piecewise, WSR = Wilcoxon Signed-Rank.
△ Based on monte carlo sampling 150,000 of 2^N possible permutations.

5.5 Discussion

We compared six methods for hypothesis testing while accounting for left truncation. Our simulation results suggest that the Gamma-curve-like nonlinear model with random intercept and maximum likelihood for left-truncated data works well in small sample sizes ($N \geq 20$) if the percentage of measurements below the detection limit is greater than 10%. These results further indicate that Gamma method performs reasonably well in moderate to large sample size (say $N \geq 30$) if the censoring is increased. However, the results for the Gamma ML method are based on a correctly assumed model; the results may not be as favorable under a misspecified model includsing deviations from the fixed effects or random effects structure. Finally, the WSR (particularly the permutation test) approaches worked satisfactorily well with sample sizes exceeding 30 regardless of the level of censoring.

This research is relevant for the following reasons. Periodontal research studies often have interest in identifying molecular mediators of inflammation via repeated measures analysis that can be induced to significant change over time. Although nonparametric methods are preferred over parametric methods, it is unclear how to handle missing data in this context, particularly left truncation of observations due to a lower detection limit. For this setting, it may be beneficial to use a parametric method that addresses limitations associated with parametric approaches, including the missingness process, as well as assumptions about the nature of left truncation of observations due to a lower detection limit. However, in order to use parametric methods for research, it is important to examine whether the properties of the model underlying hypothesis testing are supported.

The results from the gamma-curve-like nonlinear model using maximum likelihood suggest that this method may be a good parametric approach for analyzing experimental gingivitis data when interest is in mean (or median) response over time in biomarker

levels and characterizing the direction of the change. The simulations suggest that this method is reliable for small sample sizes (N=20) when the level of truncation is no more than 10%. Otherwise, larger samples are needed to reliably use the method to achieve adequate power to detect the changes in biomarker levels. In the case of the nonparametric methods, the permutation test performed well under some level of truncation with larger samples. The implication for applications is that, in the presence of truncation, with small sample sizes, this method may have some power loss relative to parametric methods to detect the change in a given biomarker. Thus, when designing experimental gingivitis studies using these analytic methods, consideration should be given to sample size and power issues if left censoring in the data is probable.

In summary, this study shows that both parametric and nonparametric methods can be reliably used to analyze biomarker levels in experimental gingivitis data. However, as demonstrated by our simulations, the appropriateness of the approach does vary depending on the sample size and the appropriateness of model and distributional assumptions.

CHAPTER 6: FUTURE RESEARCH

6.1 Alternative Models

In this research, the gamma-curve-like mixed effects model was considered the correctly specified model. One future area of research could be to evaluate size of test and power based on other models for data generation. One possible model could be Fang et al. (2011) PK model. An alternative could be a piecewise polynomial regression model (Edwards et al., 2006). Further, a gamma-curve-like model with multiplicative error structure could be considered as well.

6.2 Multiple Hypothesis Testing

Regardless of the statistical methodology used in analysis of periodontal research studies, these studies often require a number of outcomes to be examined and many hypotheses to be tested. Such testing involving multiple outcome measures may increase the risk of Type I errors when multiple simultaneous hypotheses are tested at individual-level (uncorrected) α-levels. One popular approach to this multiplicity for confirmatory analysis is control of the family-wise error rate (FWER) Hochberg (1988). Defined as the probability that at least one true null hypothesis is rejected when any of the null hypotheses hold, the FWER can be conservative when there are many hypotheses tested. Alternatively, the false discovery rate (FDR), defined by Benjamini and Hochberg (1995) as the expected proportion of Type I error among the number of rejections, is the preferred approach when the aims of a study are ex-

ploratory. Although the FDR procedure is better suited than the FWER procedures for an experimental gingivitis study with a large number of biomarkers and tests (e.g., Offenbacher et al. (2010)), an improved procedure which gives attention to the patterns of dependency among tests is needed to compare the multiple hypothesis testing procedures in experimental gingivitis and to determine whether adjustments to Hochbergs or Benjamini-Hochbergs multiple hypothesis testing procedures lead to improved inference for problems in similar settings.

APPENDIX A: SAS CODE FOR PIECEWISE MODEL

PROC NLMIXED is used to fit maximum likelihood to the biomarker data in the presence of left truncation.

proc nlmixed data=*data* cov;
parms b0=*b0* b1=*b1* b2=*b2* b3=*b3* b4=*b4* b5=*b5* sigsqerr=*sigsqerr* sigsqb=*sigsqb*;
eta = b0i + b0 + b1*time + b2*x1 + b3*x2 + b4*x3 + b5*x4;
if (*d* ne 0) then f = (1 / sqrt(2*constant('PI')*sigsqerr)) * exp(-0.5*((*y*-eta)**2)/sigsqerr);
else if (*d* = 0) then f = CDF('NORMAL', *d*, eta, sqrt(sigsqerr));
ll = log(f);
model *y* ~ general(ll);
random b0i ~ normal(0, sigsqb) subject=*subject*;
run;

The initial parameters (parms) are estimated from a repeated measures model using PROC MIXED.

APPENDIX B: SAS CODE FOR GAMMA-CURVE-LIKE MODEL

PROC NLMIXED is used to fit maximum likelihood to the biomarker data in the presence of left truncation.

```
proc nlmixed data=data cov;
parms b0=b0 b1=b1 b2=b2 theta1=theta1 sigsqerr=sigsqerr sigsqb=sigsqb;
f1 = (time+theta1)**b1;
f2 = exp(-b2*(time+theta1));
eta = b0*f1*f2;
if (d ne 0) then f = (1 / sqrt(2*constant('PI')*sigsqerr)) * exp(-0.5*((y-eta)**2)/sigsqerr);
else if (d = 0) then f = CDF('NORMAL', d, eta, sqrt(sigsqerr));
ll = log(f);
model y ~ general(ll);
random b0i ~ normal(0, sigsqb) subject=subject;
run;
```

BIBLIOGRAPHY

Aitken, M. (1981). A note on regression analysis of censored data. *Technometrics*, 23:161–163.

Baccarelli, A., Pfeiffer, R., Consonni, D., Pesatori, A. C., Bonzini, M., Jr, D. G. P., Bertazzi, P. A., and Landi, M. T. (2005). Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the seveso chloracne study. *Chemosphere*, 60:898–906.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Annals of Probability*, 3:534–545.

Cohen, A. C. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *The Annals of Mathematical Statistics*, 21(4):557–569.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics*, 1(3):217–237.

Deinzer, R., Waschul, B., and Herforth, A. (2004). Effects of experimental gingivitis on crevicular pge_2 in a split mouth trial. *Journal of Clinical Periodontology*, 31:501–505.

Deinzer, R., Weik, U., Kolb-Bachofen, V., and Herforth, A. (2007). Comparison of experimental gingivitis with persistent gingivitis: differences in clinical parameters and cytokine concentrations. *Journal of Periodontal Research*, 42:318–324.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39:1–38.

DeRouen, T. A., Hujoel, P. P., and Mancl, L. A. (1995). Concise review: statistical issues in periodontal research. *Journal of Dental Research*, 74:1731–1737.

Edwards, L., Stewart, P., MacDougall, J., and Helms, R. (2006). A method for fitting regression splines with varying polynomial order in the linear mixed model. *Statistics in Medicine*, 25:513–527.

Fang, L., Chen, P., Chunlei, K., and Lee, E. (2011). Estimating area under the curve and relative exposure in a pharmacokinetic study with data below quantification limit. *Journal of the Society for Industrial and Applied Mathematics*, 21:66–76.

Giannopoulou, C., Cappuyns, I., and Mombelli, A. (2003). Effect of smoking on gingival crevicular fluid cytokine profile during experimental gingivitis. *Journal of Clinical Periodontology*, 30:996–1002.

Gilliom, R. J. and Helsel, D. R. (1986). Estimation of distributional parameters for censored trace level water quality data, 1. estimation techniques. *Water Resources Research*, 22(2):135–146.

Gleit, A. (1985). Estimation for small normal data sets with detection limits. *Environmental Science and Technology*, 19(12):1201–1206.

Hashimoto, L. and Trussell, R. (1983). Evaluating water quality data near the detection limit. *Advanced Technology Conference, American Water Works Association, Las Vegas, NV*, 39(1).

Helsel, D. R. (1990). Less than obvious-statistical treatment of data below the detection limit. *Environmental Science and Technology*, 24(12):1766–1774.

Helsel, D. R. (2006). Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere*, 65:2434–2439.

Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75:800–802.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Hughes, J. P. (1999). Mixed-effects model with censored data with application to hiv rna levels. *Biometrics*, 55:625–629.

Jepsen, S., Eberhard, J., Fricke, D., Hedderich, J., Siebert, R., and Acil, Y. (2003). Interleukin-1 gene polymorphisms and experimental gingivitis. *Journal of Clinical Periodontology*, 30:102–106.

Johnson, T. C., Reinhardt, R. A., Payrte, J., Dyer, J., and Patil, K. (1997). Experimental gingivitis in periodontitis-susceptible subjects. *Journal of Clinical Periodontology*, 24:618–625.

Kang, M. (2007). Multiple testing in genome-wide studies.

Konradsson, K., Claesson, R., and van Dijken, J. W. V. (2007). Dental biofilm, gingivitis and interleukin-1 adjacent to approximal sites of a bonded ceramic. *Journal of Clinical Periodontology*, 34:1062–1067.

Konradsson, K. and van Dijken, J. W. V. (2005). Interleukin-1 levels in gingival crevicular fluid adjacent to restorations of calcium aluminate cement and resin composite. *Journal of Clinical Periodontology*, 32:462–466.

Laird, N. M. and Ware, J. H. (1982). Random-effects model for longitudinal data. *Biometrics*, 38:963–974.

Lee, M. and Kong, L. (2011). Median regression for longitudinal left-censored biomarker data subject to detection limit. *Statistics in Biopharmaceutical Research*, 3(2):363–371.

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

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Loe, H., Theilade, E., and Jensen, S. B. (1965). Experimental gingivitis in man. *Journal of Periodontology*, 36:177–187.

Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., Bernstein, L., and Hartge, P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112(17):1691–1696.

Lyles, R. H., Lyles, C. M., and Taylor, D. (2000). Random regression models for human immunodeficiency virus ribonucleic acid data subject to left censoring and informative drop-outs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(4):485–497.

Marquardt, D. W. (1963). An algorithm for least squares estimation on nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Moschandreas, D. J., Karuchit, S., Kim, Y., Ari, H., Lebowitz, M. D., ORourke, M. K., Gordon, S., and Robertson, G. (2001a). On predicting multi-route and multimedia residential exposure to chlorpyrifos and diazinon. *Journal of Exposure Analysis and Environonmental Epidemiology*, 11(1):56–65.

Moschandreas, D. J., Karuchit, S., Kim, Y., Ari, H., Lebowitz, M. D., ORourke, M. K., Gordon, S., and Robertson, G. (2001b). In-residence, multiple route exposures to chlorpyrifos and diazinon estimated by indirect method models. *Atmospheric Environment*, 35(12):2201–2213.

Offenbacher, S., Barros, S., Mendoza, L., Mauriello, S., Preisser, J., Moss, K., de Jager, M., and Aspiras, M. (2010). Changes in gingival crevicular fluid inflammatory mediator levels during the induction and resolution of experimental gingivitis in humans. *Journal of Clinical Periodontology*, 37:324–333.

Offenbacher, S., Barros, S. P., Singer, R. E., Moss, K., Williams, R. C., and Beck, J. D. (2007). Periodontal disease at the biofilm-gingival interface. *Journal of Periodontology*, 78:1911–1925.

Persson, T. and Rootzen, H. (1977). Simple and highly efficient estimators for a type 1 censored normal sample. *Biometrika*, 64(1):123–128.

Preisser, J. S., Sen, P. K., and Offenbacher, S. (2011). Multiple hypothesis testing for experimental gingivitis based on wilcoxon signed rank statistics. *Statistics in Biopharmaceutical Research*, 3(2):372–384.

Salvi, G. E., Franco, L. M., Braun, T. M., Lee, A., Persson, G. R., Lang, N. P., and Giannobile, W. V. (2010). Pro-inflammatory biomarkers during experimental gingivitis in patients with type 1 diabetes mellitus: a proof-of-concept study. *Journal of Clinical Periodontology*, 37:9–16.

Schluchter, M. D. (1992). Methods for the analysis if informatively censored longitudinal data. *Statistics in Medicine*, 11:1861–1870.

Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations*, volume 70. Marcel Dekker.

Sen, P. K. (2008). Kendalls tau in high-dimensional genomic parsimony. *IMS Collections*, 3:251–266.

Singh, A. (1993). Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers. In Patil, G. and Rao, C., editors, *Multivariate Environmental Statistics*, pages 445–488. Elsevier Science.

Singh, A. and Nocerino, J. (2002). Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*, 60:69–86.

Staab, B., Eick, S., Knofler, G., and Jentsch, H. (2009). The influence of a probiotic milk drink on the development of gingivitis: a pilot study. *Journal of Clinical Periodontology*, 36:850–856.

Subrahmanyam, M. V. and Sangeetha, M. (2003). Gingival crevicular fluid a marker of the periodontal disease activity. *Indian Journal of Clinical Biochemistry*, 18(1):5–7.

Thiebaut, R. and Jacqmin-Gadda, H. (2004). Mixed models for longitudinal left-censored repeated measures. *Computer Methods and Programs in Biomedicine*, 74:255–260.

Thompson, M. L. and Nelson, K. P. (2003). Linear regression with type i interval- and left-censored response data. *Environmental and Ecological Statistics*, 10:221–230.

Tsalikis, L. (2010). The effect of age on the gingival crevicular fluid composition during experimental gingivitis. a pilot study. *The Open Dentistry Journal*, 4:13–26.

Waschul, B., Herforth, A., Stiller-Winkler, R., Idel, H., Granrath, N., and Deinzer, R. (2003). Effects of plaque, psychological stress and gender on crevicular il-1β and il-1ra secretion. *Journal of Clinical Periodontology*, 30:238–248.