Testing for Serial Homogeneity and Pooled Correlation for Longitudinally Measured Biomarkers

by

Su Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2013

Doctoral Committee:

Professor Thomas M. Braun, Chair
Professor William V. Giannobile
Professor Timothy D. Johnson
Assistant Professor Lu Wang

For my family and friends

ACKNOWLEDGEMENTS

At this time I would like to express my deep appreciation and gratitude to the following people for helping me complete this dissertation.

First and foremost, I would like to thank Dr. Thomas Braun, my academic advisor and committee chair, for giving me the opportunity to work on my research projects and guiding me with great patience and outstanding knowledge throughout my years at the Department of Biostatistics. Dr. Braun has always been supportive and considerate. From him I learned not only statistical knowledge and techniques, but also a lot of academic writing and presentation skills. Dr. Braun has also been supporting me by making great suggestions on my career development. Without Dr. Braun's generous advice and encouragement, this work would not be possible. His considerable expertise and passion for his students will always be an inspiration to me.

I would also like to express my appreciation to my committee members. I thank Dr. William Giannobile for being my cognate member and giving me suggestions in the midst of his busy schedule. I would also like to thank Dr. Timothy Johnson, for his valuable advice on my dissertation and his patience with my research needs. From him I have learned a great deal in the Bayesian field. Dr. Johnson's Bayesian class has been one of my most enjoyable courses. I also thank Dr. Lu Wang for her thoughtful comments on my research and continuous support during my Ph.D life.

I am full of gratitude to my colleagues from the Department of Biostatistics at the University of Michigan for their full support over the years and the congenial environment they created. I enjoyed learning statistical knowledge from all of the knowledgeable professors, sharing research ideas and other academic experiences with all of the talented fellow students. I thank them for many wonderful moments in such an outstanding research community.

Special appreciation goes to my brothers and sisters in Christ at the Chinese Christian Fellowship Church in Ann Arbor for persistently showing me God's great love and always remembering me in their prayers. I thank them all for their caring and encouragement.

Finally, I am grateful to have supportive family members. I would like to express my profound thanks to my parents for their endless love. During my every decision making, they have stood firmly behind me and provided me with unconditional support. Last but not least, I would like to thank my dear husband, Min, who is also my brother in Christ, for being with me and showing me his continuous love and full support.

TABLE OF CONTENTS

LIST OF TABLES

Table

ABSTRACT

Testing for serial homogeneity and pooled correlation for longitudinally measured biomarkers

by

Su Chen

Chair: Thomas M. Braun

Salivary biomarkers play an important role in predicting oral disease status along with oral bacterial pathogens. Thus, our work is motivated by a study that longitudinally measured periodontal biomarkers and levels of bacterial pathogens in the oral cavity with the intent of testing whether the correlation between each biomarker and each pathogen is homogeneous over time.

We first developed both frequentist and Bayesian approaches for testing for serial homogeneity of correlation coefficients. We proposed two Wald tests and an F-test based on the asymptotic distributions of sample correlation coefficients. We found that the Wald test based on Fisher's Z-transformation and the F-test have nominal sizes when the data fit our assumed model, while the other Wald test has a more inflated size in small samples. The Wald test based on Fisher's Z-transformation is generally robust to mis-specified models and heavier tailed data.

We then applied the concepts of Bayesian credible intervals and Bayesian posterior predictive p-values. We decomposed the variance/covariance matrix of the data to

standard deviation elements and correlation elements and ran a Metropolis-Hastings algorithm within Gibbs with a set of parameters being updated at one time. Our simulation results showed that Bayesian tests provide an alternative way of testing homogeneity of serial correlations.

Under an assumption of homogeneity, we then developed a Mantel-Haenszel-type estimator of the pooled correlation coefficient and its asymptotic variance estimate as the sample size goes to infinity. Through simulations, we found that our proposed Mantel-Haenszel estimator is very close to the true common correlation, and that the variance estimator also performs well even with a small sample size. In addition, the variance estimator remains robust to model mis-specification.

When applied to actual data, we found some significant, time-invariant correlation did exist between MMP-8 and MMP-9 and some red complex pathogens. These results are supported by published clinical research and demonstrate the utility of our methods for providing guidance to investigators as to which biomarker/pathogen pairs might best describe disease severity over time.

CHAPTER I

Introduction

Our methods are motivated by data collected from a small longitudinal study of gingivitis, or inflammation of the gums (gingivae) (Salvi et al., 2010). Investigators enrolled nine subjects with Type I diabetes and nine control subjects without Type I diabetes. The eighteen subjects in this study were instructed to refrain from all oral health practices for 21 days so that the natural progression of gingivitis could occur. After 21 days, the subjects were instructed to return to usual oral health practices for two weeks. Each patient was examined at baseline (Day 0), 21 days after enrollment (Day 21), when progression of gingivitis had occurred, and 35 days after enrollment (Day 35), when gingivitis would be resolved. At each of these three time points, investigators collected samples of plaque and gingival crevicular fluid (GCF), the fluid between gums and teeth, from multiple sites of the mouths of the eighteen subjects.

The GCF was analyzed for levels of several biomarkers, including tumor necrosis factor (TNF)-α, calprotectin, matrix metalloproteinase-8 (MMP-8), and MMP-9, all of which are known inflammatory markers and have been shown to exist in higher levels in periodontally diseased subjects than in periodontally healthy subjects (Rai et al., 2008; Yucel et al., 2008). The plaque was analyzed for levels of numerous bac-

terial pathogens, three of which, *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*, are known collectively as the "red complex" (Socransky et al., 1998) and have been associated with higher levels of gingival inflammation (Socransky et al., 2002; Zhang et al., 2002). Given that both the biomarkers and pathogens appear to have an association with oral disease, it has been suggested that the level of oral pathogens may directly trigger an immune response and thereby promote increased levels of inflammatory biomarkers, including those listed (Yakob et al., 2012). Thus, one goal of the pilot study was to assess the association of GCF biomarkers and plaque pathogens during the progression and resolution of gingivitis and whether the associations changed over the course of the study. Specifically, investigators were curious to know if the association of pathogens and biomarkers would be weak prior to the development of gingivitis and would become stronger once gingivitis had developed. The data were summarized as Pearson correlation coefficients between each biomarker and each pathogen at each of Days 0, 21, and 35, and we began investigating methods for testing the equality of serially-measured correlation coefficients. Another issue that is of interest is whether there is a summary correlation to describe the overall association between each biomarker and each pathogen, if the correlation between them does not change over time.

The data described above generated two questions of interest. The first question is how to test the equality of serial correlation between two longitudinally measured continuous variables, and the second question aims to look for a pooled correlation coefficient for the serial correlations if lack of homogeneity exists. In Chapters II and III, we provide approaches from both frequentist and Bayesian frameworks, respectively, for inference regarding heterogeneity of longitudinal correlation coefficients, and in Chapter IV we propose methods for both estimation and inference of a pooled

correlation coefficient.

Specifically, in the second chapter, we propose a model for the joint distribution of the serial biomarker measures and the serial pathogen measures and from this model, we derive the asymptotic distribution of the sample correlation coefficient of a biomarker and a pathogen at each time point. To determine if the correlation between a biomarker and a pathogen is homogeneous over time, we use both a Wald test based upon Fisher's Z-transformation and an F-test with estimated degrees of freedom in order to produce a test with valid size. We examine the performance of both tests via Monte Carlo simulation in a variety of settings defined by the number of subjects, the number of time points, and the range of the true correlation coefficients. We also evaluated the validity of our tests when our assumed model does not match the true model of the data.

In the third chapter, we propose to evaluate whether the correlation between two continuous quantities measured at two or more time points are equal using Bayesian credible intervals and Bayesian posterior predictive p-values. We use Bayesian credible interval method with simulated data of two repeated measures, and sample the serial correlations from their posterior distributions. We then evaluate the credible interval of the samples drawn and decide whether to reject our null hypothesis. We use Bayesian posterior predictive p-values with simulated data with more than two time points, and the posterior predictive p-value is evaluated for Wald-like test statistics and discrepancy measures. To sample parameters from their posterior distributions, we run a Metropolis-Hastings algorithm within Gibbs with a set of parameters being updated at one time. We examine the performance of our proposed Bayesian tests via Monte Carlo simulation in a variety of settings defined by the number of subjects, the number of time points, the range of the true correlation coefficients, the prior

setting, and whether there exists model violations. Our simulation results suggest that Bayesian approaches have comparable performance in identifying heterogeneity in medium sized datasets (e.g. $n = 50$) compared to Wald tests that are based on asymptotics.

In the fourth chapter, we further explore how to pool serial correlation estimates after a test of serial heterogeneity lacks statistical significance. We propose a Mantel-Haenszel-like estimator for the pooled correlation coefficient and derive an asymptotic variance estimator when the sample size goes to infinity. Unfortunately, we are still in need of a valid variance formula when the number of time points goes to infinity. We evaluate the bias of our pooled correlation estimator and the variance estimator based on the number of subjects going to infinity in different settings via Monte Carlo simulations.

CHAPTER II

Tests for Time-invariant Correlation of Longitudinally Measured Biomarkers

2.1 Introduction

Interest in assessing the equality of correlation coefficients has been examined in a variety of research settings. Hotelling (1940) proposed methods for determining whether a calf's girth or the calf's length at an early age is a better predictor of the calf's ultimate weight. Elston (1975) examined the homogeneity of intra- and interclass correlation coefficients in a study of the correlation of heights within and between genders. Donner and Zou (2002) applied several methods to study the equality of intra-correlations of two techniques used for measuring ventricle-brain ratio. Olkin and Finn (1990) examined whether the correlation of systolic blood pressure with body-mass index (BMI) was equal for three different age cohorts.

Several published statistical methods exist for assessing homogeneity of correlation coefficients. Olkin and Siotani (1976) propose using the asymptotic normality of sample correlation coefficients, and Olkin and Finn (1990) derive an asymptotic χ^2_{p-1} test of equality of more than two correlated coefficients when assuming a spe-

cific form of the variance-covariance matrix. Dunn and Clark (1969) and Dunn and Clark (1971) present related methods using Fisher's Z-transformation, defined as $z = ln[(1 + r)/(1 - r)]/2$, where r is the sample estimate of ρ (Fisher, 1915, 1921). Raghunathan et al. (1996) compared the power of a statistic based on the difference of two correlations and the difference of their Fisher's Z-transformed values for testing equality and suggested Fisher's Z-transformation should be used to obtain higher power. Meng et al. (1992) extended the work of Dunn and Clark (1969) using Fisher's Z transformation with the goal of comparing correlation coefficients between a dependent random variable and a set of mutually independent random variables. Raghunathan (2003) then extended Meng et al. (1992)'s methods to allow for missing values in the data.

The methods cited above focused upon the comparison of two correlation coefficients. However, given our interest in assessing the homogeneity of several correlation coefficients from longitudinal data, an obvious extension of the methods was needed. We chose to model the data using a more general joint normal distribution that is more appropriate for a longitudinal study than the model used by Olkin and Siotani (1976) and Olkin and Finn (1990). We examine the performance of modified tests using both untransformed correlation coefficients, as well as one using Fisher's Z-transformed correlation coefficients. For certain data patterns, an F-test is introduced. In Section 2.2, we describe our model for the joint distribution of the serial measures of a biomarker and a pathogen and derive the joint asymptotic distribution of the serial sample correlation coefficients. We then present our three test statistics and appropriate null distributions for each. In Section 2.3, we assess the empirical size and power of our proposed tests in a variety of scenarios under two major settings (medium serial correlations and small serial correlations), motivated by the data col-

lected in the pilot study described earlier and also apply our methods to data from another longitudinal periodontal study. Section 2.5 contains our concluding remarks.

2.2 Methods

2.2.1 Notation

We have n subjects who are each examined sequentially at times $t_1 < t_2, \ldots, < t_m$. Let X_{ij} and Y_{ij}, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$, denote the respective values of biomarker X and biomarker Y collected from subject i at time t_j. Marginally, we assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$. We assume that $\boldsymbol{X_{i.}}$=$\{X_{i1}, X_{i2}, \ldots, X_{im}\}$ and $\boldsymbol{Y_{i.}} = \{Y_{i1}, Y_{i2}, \ldots, Y_{im}\}$ have a multivariate normal distribution in which the elements of $\boldsymbol{X_i}$ are assumed to be exchangeably correlated with correlation ρ_x, and the elements of $\boldsymbol{Y_i}$ are exchangeably correlated with correlation ρ_y. A common cross-correlation, ρ_{xy} between X_{ij} and Y_{ik} is also assumed, where $j \neq k$. All of these correlations are nuisance parameters; our primary interest lies in $\rho_1, \rho_2, \ldots, \rho_m$, in which $\rho_j = Corr(X_{ij}, Y_{ij}), j = 1, 2, \ldots m$.

If we then denote $\boldsymbol{D}_i = \{X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, \cdots, X_{im}, Y_{im}\}^t$ as the $(2m \times 1)$ longitudinal vector of pairs of biomarker and pathogen for subject i, $\boldsymbol{D}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ in which

$$\boldsymbol{\mu} = \{\mu_{x1}, \mu_{y1}, \mu_{x2}, \mu_{y2}, \cdots, \mu_{xm}, \mu_{ym}\} \tag{2.1}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\tau_1 & \rho_x\sigma_1\sigma_2 & \rho_{xy}\sigma_1\tau_2 & \cdots & \rho_x\sigma_1\sigma_m & \rho_{xy}\sigma_1\tau_m \\ \rho_1\sigma_1\tau_1 & \tau_1^2 & \rho_{xy}\sigma_2\tau_1 & \rho_y\tau_1\tau_2 & \cdots & \rho_{xy}\sigma_m\tau_1 & \rho_y\tau_1\tau_m \\ \rho_x\sigma_1\sigma_2 & \rho_{xy}\sigma_2\tau_1 & \sigma_2^2 & \rho_2\sigma_2\tau_2 & \cdots & \rho_x\sigma_2\sigma_m & \rho_{xy}\sigma_2\tau_m \\ \rho_{xy}\sigma_1\tau_2 & \rho_y\tau_1\tau_2 & \rho_2\sigma_2\tau_2 & \tau_2^2 & \cdots & \rho_{xy}\sigma_m\tau_2 & \rho_y\tau_2\tau_m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_x\sigma_1\sigma_m & \rho_{xy}\sigma_m\tau_1 & \rho_x\sigma_2\sigma_m & \rho_{xy}\sigma_m\tau_2 & \cdots & \sigma_m^2 & \rho_m\sigma_m\tau_m \\ \rho_{xy}\sigma_1\tau_m & \rho_y\tau_1\tau_m & \rho_{xy}\sigma_2\tau_m & \rho_y\tau_2\tau_m & \cdots & \rho_m\sigma_m\tau_m & \tau_m^2 \end{pmatrix} \tag{2.2}$$

2.2.2 Proposed Tests

We are interested in testing the hypotheses $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m$ versus H_a : two or more of $\rho_1, \rho_2, \ldots, \rho_m$ are unequal. For time j, let $\boldsymbol{X_{.j}} = \{X_{1j}, X_{2j}, \cdots, X_{nj}\}$ and $\boldsymbol{Y_{.j}} = \{Y_{1j}, Y_{2j}, \cdots, Y_{nj}\}$ denote the respective vectors of all subjects' values of biomarker X and pathogen Y. For $j \neq k$, we then denote $\tilde{S}_{XX_j}$ as the sample variance of $\boldsymbol{X_{.j}}$, $\tilde{S}_{YY_j}$ as the sample variance of $\boldsymbol{Y_{.j}}$, $\tilde{S}_{XX_{jk}}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{X_{.k}}$, $\tilde{S}_{YY_{jk}}$ as the sample covariance between $\boldsymbol{Y_{.j}}$ and $\boldsymbol{Y_{.k}}$, $\tilde{S}_{XY_j}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{Y_{.j}}$, and $\tilde{S}_{XY_{jk}}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{Y_{.k}}$. Elston (1975) has shown that the maximum likelihood estimators $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m$,

$\hat{\rho}_x$, $\hat{\rho}_y$ and $\hat{\rho}_{xy}$ for $\rho_1, \rho_2, \ldots, \rho_m$, ρ_x, ρ_y and ρ_{xy}, respectively, are:

$$\hat{\rho}_j = \frac{\widehat{\rho_j \sigma_j \tau_j}}{\sqrt{\hat{\sigma_j^2}}\sqrt{\hat{\tau_j^2}}} = \frac{\tilde{S}_{XY_j}}{\sqrt{\tilde{S}_{XX_j}\tilde{S}_{YY_j}}} \; ; \; j = 1, 2, \ldots m \tag{2.3}$$

$$\hat{\rho}_x = \frac{\sum_{j \neq k} \widehat{\rho_x \sigma_j \sigma_k}}{\sum_{j \neq k} \sqrt{\hat{\sigma}_j^2 \hat{\sigma}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{XX_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{XX_j}\tilde{S}_{XX_k}}} \tag{2.4}$$

$$\hat{\rho}_y = \frac{\sum_{j \neq k} \widehat{\rho_y \tau_j \tau_k}}{\sum_{j \neq k} \sqrt{\hat{\tau}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{YY_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{YY_j}\tilde{S}_{YY_k}}} \tag{2.5}$$

$$\hat{\rho}_{xy} = \frac{\sum_{j \neq k} \widehat{\rho_{xy} \sigma_j \tau_k}}{\sum_{j \neq k} \sqrt{\hat{\sigma}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{XY_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{XY_j}\tilde{S}_{XY_k}}} \tag{2.6}$$

We now derive the variance of each $\hat{\rho}_j$. Let c_{ls} be the usual estimate (i.e., sample covariance) of γ_{ls}, which is the (l, s)th element of the covariance matrix of a multivariate normal distribution with n observations. According to Elston (1975), asymptotically,

$$cov(c_{ls}, c_{rh}) = \frac{1}{n}(\gamma_{lh}\gamma_{sr} + \gamma_{lr}\gamma_{sh}) \tag{2.7}$$

Using this asymptotic property, we are able to derive the asymptotic variances and covariances of $\widehat{\rho_j \sigma_j \tau_j}$, $\hat{\sigma}_j^2$ and $\hat{\tau}_j^2$. After some straightforward algebra,

$$var(\hat{\sigma}_j^2) = \frac{2}{n}\sigma_j^4$$

$$var(\widehat{\rho_j \sigma_j \tau_j}) = \frac{\sigma_j^2 \tau_j^2}{n}(1 + \rho_j^2)$$

$$cov(\hat{\sigma}_j^2, \hat{\tau}_j^2) = \frac{\sigma_j^2 \tau_j^2}{n}(1 + \rho_j^2)$$

$$cov(\hat{\sigma}_j^2, \hat{\sigma}_k^2) = \frac{2}{n}(\rho_x \sigma_j \sigma_k)^2$$

$$cov(\hat{\sigma}_j^2, \hat{\tau}_k^2) = \frac{2}{n}(\rho_{xy}\sigma_j\tau_k)^2$$

$$cov(\widehat{\rho_j\sigma_j\tau_j}, \widehat{\rho_j\sigma_k\tau_k}) = \frac{\sigma_j\sigma_k\tau_j\tau_k}{n}(\rho_x\rho_y + \rho_{xy}^2)$$

$$cov(\hat{\sigma}_j^2, \widehat{\rho_j\sigma_j\tau_j}) = \frac{2}{n}(\rho_j\sigma_j^3\tau_j)^2$$

$$cov(\hat{\sigma}_j^2, \widehat{\rho_k\sigma_j\tau_k}) = \frac{2}{n}(\rho_x\rho_{xy}\sigma_j^2\sigma_k\tau_k)^2$$

The maximum likelihood estimates of $\rho_j's$ can be expressed by Equation 2.3 , a function of elements whose asymptotic variances and covariances are shown above. Therefore the asymptotic variance of $\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_m$ can be easily obtained by Delta method.

Therefore, the variance of each $\hat{\rho}_j$ is

$$Var(\hat{\rho}_j) \approx \frac{1}{n}(1 - \hat{\rho}_j^2)^2, \tag{2.8}$$

which is a function solely of $\hat{\rho}_j$, while the covariance between $\hat{\rho}_j$ and $\hat{\rho}_k$ is

$$Cov(\hat{\rho}_j, \hat{\rho}_k) \approx \frac{1}{n}\left\{\frac{1}{2}\hat{\rho}_j\hat{\rho}_k(\hat{\rho}_x^2 + \hat{\rho}_y^2) + \hat{\rho}_{xy}^2(1 + \hat{\rho}_j\hat{\rho}_k) + \hat{\rho}_x\hat{\rho}_y - \hat{\rho}_{xy}(\hat{\rho}_j + \hat{\rho}_k)(\hat{\rho}_x + \hat{\rho}_y)\right\}, \tag{2.9}$$

which is a function of not only $\hat{\rho}_j$ and $\hat{\rho}_k$, but also $\hat{\rho}_x$, $\hat{\rho}_y$, and $\hat{\rho}_{xy}$. Note that Equations (2.8) and (2.9) are similar in form to those given by Olkin and Siotani (1976), but differ in order to reflect the fact that the covariance matrix given in (2.2) differs from that used by Olkin. We also mention that Yu and Dunn (1982) suggested that the value n in Equations (2.8) and (2.9) be replaced by the value $n-3$ in order to improve the approximation in small samples.

We let $\boldsymbol{\rho} = \{\rho_1, \rho_2, \ldots, \rho_m\}$ with corresponding estimator $\hat{\boldsymbol{\rho}} = \{\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m\}$, so that ρ_x, ρ_y, and ρ_{xy} are viewed as nuisance parameters. Let $\hat{\Sigma}_\rho$ be an $m \times m$

matrix with diagonal element (j, j) equal to $Var(\hat{\rho}_j)$ as given by Equation (2.8), and off-diagonal element (j, k) equal to $Cov(\hat{\rho}_j, \hat{\rho}_k)$, as given by Equation (2.9). Also let L be an $m \times (m-1)$ contrast matrix for the pairwise differences, i.e.

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}$$

Then the null hypothesis $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m$ can be tested using the Wald statistic $\chi^2_\rho = (L\hat{\rho})^T (L\hat{\Sigma}_\rho L^T)^{-1} (L\hat{\rho})$ which has an asymptotic chi-square distribution with $m-1$ degrees of freedom under the null hypothesis.

However, we discovered in simulations with small sample sizes and small serial correlations that comparison of χ^2_ρ to a chi-squared distribution with $m-1$ degrees of freedom led to over-rejection of the null hypothesis whether or not the denominators in Equations (2.8) and (2.9) were equal to n or $n-3$. In order to create a test with a size closer to the nominal level desired, we adopted the idea of Kenward and Roger (1997) as follows. Instead of comparing χ^2_ρ to a chi-squared distribution with $m-1$ degrees of freedom, we will instead compare $F_\rho = \lambda\chi^2_\rho/(m-1)$ to an $\mathcal{F}$-distribution with $m-1$ numerator degrees of freedom and δ denominator degrees of freedom, in which both λ and δ are estimated by equating the first two moments of F_ρ with the first two moments of the reference $\mathcal{F}$-distribution. Over a grid search of possible values of n and m, we found that a scale factor $\lambda = (n+50-(m-1))/(n+49)$ which is a number slightly smaller than 1 when $m > 2$, and $\delta = n+20-(m-1)$ led to very similar moments and a test with improved size.

An alternate test would be based on Fisher's Z-transformation, $\hat{\boldsymbol{z}} = \{\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_m\}$,

leading to the statistic $\chi_z^2 = (L\hat{\boldsymbol{z}})^T(L\hat{\Sigma}_z L^T)^{-1}(L\hat{\boldsymbol{z}})$, in which $\hat{\Sigma}_z$ is the variance of $\hat{\boldsymbol{z}}$ obtained through the Delta method. Specifically, $\hat{\Sigma}_z$ has diagonal element (j,j) equal to $1/n$ and off-diagonal element (j,k) equal to $Cov(\hat{\rho}_j, \hat{\rho}_k)/[(1-\hat{\rho}_j^2)(1-\hat{\rho}_k^2)]$. We note again that Yu and Dunn (1982) suggest replacing the value n with $n-3$ in the variance and covariance expressions. This statistic also has an asymptotic chi-square null distribution with $m-1$ degrees of freedom. We found in simulations with small sample sizes that unlike the test using χ_ρ^2, a test using χ_z^2 had nominal size when the value $n-3$ was used as a replacement to n in the denominators of Equations (2.8) and (2.9).

2.3 Application of Methods

2.3.1 Simulation Study

We first examined the performance of the test with both proposed statistics χ_ρ^2 and χ_z^2 under various settings for hypothetical longitudinal datasets based upon the data from our motivating example. For each subject i, $i = 1, 2, \ldots, n$, biomarker X and pathogen Y are both observed at m time points. We assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$, in which $\mu_{xj} = 2.5$ and $\mu_{yj} = 4.0$, $\sigma_j = 0.3$, and $\tau_j = 0.40 - 0.05(j-1)$. Note that correlation is location and scale invariant, so that our results are generalizable to other values of location and scale. In terms of the joint distribution of the data, we considered two settings. In the first setting, which we call an "autoregressive nuisance" setting, the elements of $\boldsymbol{X_{i\cdot}}$ have an autoregressive correlation structure with correlation $\rho_{x_0}^{|j-k|}$, the correlation between Y_{ij} and Y_{ik} is similarly set to be $\rho_{y_0}^{|j-k|}$, and the cross-correlation between X_{ij} and Y_{ik} or between X_{ik} and Y_{ij} is $c\rho_{xy_0}^{|j-k|}$, where c is a positive constant. In our second setting, which we

call a "constant nuisance" setting, the elements of $\boldsymbol{X_i}$ are exchangeably correlated with correlation ρ_x, and the elements of $\boldsymbol{Y_i}$ are exchangeably correlated with correlation ρ_y. We also assume a common cross-correlation, ρ_{xy} between X_{ij} and Y_{ik}, where $j \neq k$. Thus, the autoregressive nuisance setting violates our assumed model, whereas the constant nuisance setting matches our assumed model.

We first present simulation results for the setting with autoregressive nuisance correlations. We selected the values $\rho_{x_0} = 0.5$, $\rho_{y_0} = 0.6$, $\rho_{xy_0} = 0.7$ and $c = 1.7 \times \rho_{x_0}\rho_{y_0}$. With regard to the correlation parameters of interest, $\{\rho_1, \rho_2, \ldots, \rho_m\}$, a specific set of values was defined by two quantities, $\rho_{min} \in \{0.2, 0.3, 0.4, 0.5\}$ and $\Delta \in \{0.0, 0.1, 0.3\}$. We set $\rho_1 = \rho_{min}$, $\rho_m = \rho_{min} + \Delta$, and all other correlation parameters $\rho_2, \rho_3, \ldots, \rho_{m-1}$ were equally spaced between ρ_1 and ρ_m. Thus, a value of $\Delta = 0$ represents the null hypothesis, while $\Delta > 0$ represents the alternative hypothesis. For each combination of ρ_{min} and Δ, we simulated $\boldsymbol{D_i}$=$\{X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, \cdots, X_{im}, Y_{im}\}$, the data for each subject i, from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance as described above. We considered sample sizes of $n \in \{50, 100, 500\}$ and the number of time points $m \in \{2, 3, 4, 5\}$. We performed a test of equality of correlation coefficients over time by comparing χ^2_ρ and χ^2_z, each computed with $n - 3$ in the denominator of Equations (2.8) and (2.9), to a chi-squared distribution with $m - 1$ degrees of freedom.

Table 2.1 presents the empirical size of tests using either of the two proposed statistics for various combinations of n, m, and ρ_{min}. Based upon a 95% confidence interval around a desired size of 0.05, we would expect the number of rejections in 5,000 simulation for a nominal test would lie in the interval $(4.5, 5.6)$. Overall, with a sample size of $n = 500$, both modified tests have nominal size, regardless of the number of time points and the value of ρ_{min}. With sample sizes of $n = 100$ and

$n = 50$, a test using χ_z^2 continued to remain nominal, while the size of the test using χ_ρ^2 was slightly inflated in some scenarios but not far from the upper bound of the interval. Moreover, the inflated test size when using χ_ρ^2 increased with the number of time points. Thus, tests using either statistic are generally robust under this setting, although we suggest caution when using χ_ρ^2 in small samples with several time points.

Table 2.1: Size of tests with autoregressive nuisance correlation. Each value is the percentage of simulations out of 5,000 in which the null hypothesis is rejected. χ_z^2=Test based on Fisher's Z-transformation; χ_ρ^2=Test based on original $\boldsymbol{\rho_j}$'s.

		$n = 500$		$n = 100$		$n = 50$	
m	ρ_{min}	χ_z^2	χ_ρ^2	χ_z^2	χ_ρ^2	χ_z^2	χ_ρ^2
2	0.2	5.0	5.0	4.7	4.8	5.4	5.6
	0.3	5.0	5.0	4.8	4.8	5.3	5.3
	0.4	5.0	5.0	4.8	4.7	5.1	5.0
	0.5	5.0	4.9	4.8	4.6	5.1	4.6
3	0.2	5.0	5.2	4.9	5.3	5.2	5.9
	0.3	5.2	5.2	4.9	5.2	5.1	5.7
	0.4	5.1	5.2	4.8	5.2	5.1	5.4
	0.5	5.1	5.0	4.9	4.9	5.1	4.9
4	0.2	4.5	4.7	5.4	5.7	5.0	6.2
	0.3	4.6	4.7	5.4	5.7	5.1	5.9
	0.4	4.5	4.5	5.1	5.4	5.2	5.5
	0.5	4.4	4.4	5.1	5.2	5.1	5.0
5	0.2	5.0	5.1	5.3	5.6	5.1	6.3
	0.3	4.9	5.0	5.3	5.8	4.9	6.1
	0.4	4.9	5.0	5.2	5.5	5.0	5.7
	0.5	4.9	5.0	5.3	5.3	5.1	5.3

Table 2.2 presents the power of the tests using χ_z^2 and χ_ρ^2 it the setting with autoregressive nuisance correlation. For reference, Table 2.2 also includes the estimated power of a test of equality of correlation coefficients using the formula proposed by Tu et al. (2006). However, the formula is based upon a correlation structure different

from ours, so we expect the empirical power of our test to deviate slightly from that found via the formula. From the results in Table 2.2, we see first that both tests have comparable power, and the empirical power of each is higher than that predicted by the power formula of Tu et al. (2006), especially in large data. The results suggest that the more the serial correlations depart from zero, we see an increase in power of both tests. For example, when $n = 100$ and $m = 3$ and $\Delta = 0.1$, as ρ_{min} increases from 0.2 to 0.5, the power of both tests increases from about 0.11 to 0.15. With $\Delta = 0.3$, as ρ_{min} increases from 0.2 to 0.5, the power of both tests increases from about 0.65 to 0.97. The results in Table 2.1 also suggest that power decreases as the number of time points increases. For example, with $n = 100$, $\Delta = 0.1$ and $\rho_{min} = 0.2$, the power of χ^2_ρ decreases from just over 0.50 when $m = 2$ to just over 0.30 when $m = 5$.

We now present simulation results with constant nuisance correlation. Given that we have just shown that our tests are robust to model mis-specification, we expect our tests to perform equally as well when the joint distribution of the data matches that assumed in our test statistics. The means and variances are the same as those used in the earlier simulations. For the nuisance correlation parameters, we set $\rho_x = 0.5$, $\rho_y = 0.7$, and $\rho_{xy} = 0$. We once again determined values for the correlation parameters of interest, $\{\rho_1, \rho_2, \ldots, \rho_m\}$, by considering two quantities, $\rho_{min} \in \{-0.2, -0.1, 0, 0.1, 0.2\}$ and $\Delta \in \{0.0, 0.1, 0.3\}$. We set $\rho_1 = \rho_{min}$, $\rho_m = \rho_{min} + \Delta$, and all other correlation parameters $\rho_2, \rho_3, \ldots, \rho_{m-1}$ were equally spaced between ρ_1 and ρ_m. We also considered sample sizes of $n \in \{50, 100, 500\}$ and the number of time points $m \in \{2, 3, 4, 5\}$.

We note that the values of the correlation parameters are of lower magnitude than those used in setting with autoregressive nuisance correlation. This is because the

Table 2.2: Empirical and theoretical power of both tests when Δ=0.1 and 0.3 with autoregressive nuisance correlation. Each value is the number of simulations out of 5,000 in which the p-value is less than 0.05. χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s; PF_{Tu}=Tu's power function.

Δ=0.1		$n=500$			$n=100$			$n=50$		
m	ρ_{min}	χ^2_z	χ^2_ρ	PF_{Tu}	χ^2_z	χ^2_ρ	PF_{Tu}	χ^2_z	χ^2_ρ	PF_{Tu}
2	0.2	51.8	51.9	50.7	14.1	14.4	14.3	10.1	10.4	9.6
	0.3	55.8	55.8	53.8	15.1	15.0	15.1	10.6	10.6	10.0
	0.4	63.1	63.0	57.9	17.0	16.7	16.2	11.6	11.0	10.5
	0.5	75.3	75.1	62.7	20.6	20.0	17.5	13.8	12.5	11.2
3	0.2	37.8	38.1	32.7	10.8	11.2	9.9	8.1	8.6	7.4
	0.3	41.3	41.4	35.8	11.6	11.6	10.4	8.3	8.7	7.6
	0.4	47.9	47.8	40.3	12.7	12.5	11.3	8.9	8.7	8.0
	0.5	58.7	58.5	46.7	15.0	14.5	12.5	10.0	9.2	8.6
4	0.2	34.7	35.0	27.2	10.3	10.7	8.7	7.6	8.7	6.8
	0.3	37.9	38.2	30.4	10.8	11.1	9.2	7.8	8.6	7.0
	0.4	44.5	44.4	35.4	12.0	12.1	10.1	8.4	8.7	7.4
	0.5	55.2	55.2	43.4	14.2	13.7	11.5	9.5	8.8	8.1
5	0.2	31.8	32.0	24.8	9.3	9.9	8.2	7.3	8.6	6.6
	0.3	35.2	35.4	28.1	9.7	10.2	8.7	7.5	8.4	6.8
	0.4	41.1	41.1	33.8	10.4	10.8	9.6	7.8	8.2	7.2
	0.5	51.8	51.7	43.9	12.3	12.2	11.3	8.9	8.6	7.9
Δ=0.3		$n=500$			$n=100$			$n=50$		
m	ρ_{min}	χ^2_z	χ^2_ρ	PF_{Tu}	χ^2_z	χ^2_ρ	PF_{Tu}	χ^2_z	χ^2_ρ	PF_{Tu}
2	0.2	100.0	100.0	100.0	81.1	81.0	80.2	50.7	50.7	51.1
	0.3	100.0	100.0	100.0	87.8	87.5	83.4	58.7	57.6	54.5
	0.4	100.0	100.0	100.0	95.5	95.2	86.5	71.8	69.2	58.2
	0.5	100.0	100.0	100.0	99.8	99.7	88.8	90.4	88.5	61.2
3	0.2	100.0	100.0	100.0	65.4	65.6	61.1	37.0	37.5	33.9
	0.3	100.0	100.0	100.0	73.9	73.6	67.0	43.6	43.1	37.9
	0.4	100.0	100.0	100.0	86.1	85.9	73.9	55.0	53.4	43.3
	0.5	100.0	100.0	100.0	97.1	96.9	80.7	75.3	72.9	49.5
4	0.2	100.0	100.0	99.9	61.5	62.3	53.8	32.6	33.8	28.6
	0.3	100.0	100.0	100.0	70.1	70.2	61.5	38.9	39.0	33.2
	0.4	100.0	100.0	100.0	82.0	81.6	71.6	50.2	48.9	40.4
	0.5	100.0	100.0	100.0	94.9	94.7	83.1	68.7	66.7	50.8
5	0.2	100.0	100.0	99.9	58.6	59.8	51.1	32.2	33.8	26.6
	0.3	100.0	100.0	100.0	68.2	68.5	60.4	37.3	38.0	32.0
	0.4	100.0	100.0	100.0	80.8	80.8	73.8	47.1	46.5	41.3
	0.5	100.0	100.0	100.0	93.6	93.8	89.6	66.0	63.7	58.0

empirical correlations in our motivating data were in the interval $[-0.20, 0.20]$ and we wanted to examine the performance of our tests with data simulated under conditions similar to those we observed. In this setting, since we discovered that comparison of χ^2_ρ to a chi-squared distribution with $m-1$ degrees of freedom led to over-rejection of the null hypothesis, we also compared F_ρ to an $\mathcal{F}$-distribution with $m-1$ numerator degrees of freedom and $n+20-(m-1)$ denominator degrees of freedom. The size and power of the three tests in each setting were estimated from the rejection rates in 5,000 simulated datasets.

Table 2.3 presents the empirical size of tests using the three statistics for various combinations of n, m, and ρ_{min}. Overall, with a sample size of $n = 500$, all three tests have nominal size, regardless of the number of time points and the value of ρ_{min}. With sample sizes of $n = 100$ and $n = 50$, a test using χ^2_z continued to maintain a nominal size, while the size of the test using χ^2_ρ was inflated in some settings. Relative to the results seen in Table 2.1, the magnitude of inflation in Type I error rate when using χ^2_ρ is bigger in our current setting, which has smaller serial correlations than those in the previous setting. Thus, general use of χ^2_ρ in small samples with small serial correlations is not advised. In contrast, the size of our proposed $\mathcal{F}$-test using the statistic F_ρ remains nominal regardless of the sample size.

Table 2.4 presents the power of the tests using χ^2_z and F_ρ; the test using χ^2_ρ was not be examined for power due to its invalid size. From the results in Table 2.4, we see first that the tests using F_ρ and χ^2_z have comparable power, and the empirical power of each is fairly close to that predicted by the power formula of Tu et al. (2006).

Lastly, we investigated the effect of non-normality on the size of our tests. We repeated the simulations summarized in Tables 2.1 and 2.3 on heavier tailed, correlated logistic data. Now for each subject, we first sampled a vector of independent

Table 2.3: Size of tests using χ^2_ρ and χ^2_z under setting II. Each value is the percentage of simulations out of 5,000 in which the null hypothesis is rejected. χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s; F_ρ=$\mathcal{F}$-test with estimated denominator degrees of freedom.

		$n = 500$			$n = 100$			$n = 50$		
m	ρ_{min}	χ^2_z	χ^2_ρ	F_ρ	χ^2_z	χ^2_ρ	F_ρ	χ^2_z	χ^2_ρ	F_ρ
2	-0.2	5.0	5.1	5.0	4.9	5.0	4.8	5.5	5.6	5.2
	-0.1	4.9	4.9	4.9	4.9	5.0	4.9	5.3	5.6	5.2
	0	4.9	4.9	4.9	4.9	5.1	4.8	5.3	5.7	5.3
	0.1	4.9	5.0	4.9	4.6	4.9	4.6	5.4	5.7	5.3
	0.2	4.8	4.8	4.8	4.9	4.9	4.7	5.5	5.7	5.3
3	-0.2	5.1	5.2	5.0	5.3	5.5	5.2	5.3	6.1	5.3
	-0.1	5.0	5.1	4.9	5.2	5.7	5.1	5.4	6.1	5.2
	0	5.0	5.1	5.0	5.3	5.7	5.3	5.5	6.2	5.4
	0.1	5.2	5.3	5.1	5.4	5.7	5.3	5.5	6.2	5.4
	0.2	5.2	5.3	5.1	5.3	5.5	5.1	5.5	6.2	5.2
4	-0.2	4.9	4.9	4.8	5.9	6.3	5.5	5.7	6.3	5.3
	-0.1	4.9	4.9	4.7	5.7	6.2	5.4	5.7	6.7	5.3
	0	4.8	4.9	4.7	5.7	6.1	5.4	5.6	6.6	5.5
	0.1	4.9	5.0	4.8	5.7	6.1	5.4	5.5	6.6	5.4
	0.2	5.0	5.1	4.8	5.7	6.0	5.6	5.5	6.6	5.3
5	-0.2	5.2	5.3	5.1	5.5	6.1	5.1	5.7	6.9	5.2
	-0.1	5.2	5.4	5.0	5.5	5.9	5.1	5.4	6.8	5.0
	0	5.2	5.3	5.1	5.4	6.1	5.1	5.2	6.7	4.8
	0.1	5.2	5.3	5.1	5.5	6.1	5.1	5.3	6.6	5.0
	0.2	5.2	5.3	5.1	5.5	6.1	5.1	5.4	7.0	5.2

Table 2.4: Empirical and theoretical power of both tests when Δ=0.1 and 0.3 with constant nuisance correlation. Each value is the number of simulations out of 5,000 in which the p-value is less than 0.05. χ^2_z=Test based on Fisher's Z-transformation; F_ρ=$\mathcal{F}$-test with estimated denominator degrees of freedom; Tu=Tu's power function.

Δ=0.1		$n = 500$			$n = 100$			$n = 50$		
m	ρ_{min}	χ^2_z	F_ρ	Tu	χ^2_z	F_ρ	Tu	χ^2_z	F_ρ	Tu
2	-0.2	53.0	53.0	51.6	14.3	14.2	14.6	10.4	10.0	9.7
	-0.1	50.2	50.1	50.3	13.4	13.4	14.2	10.0	9.9	9.5
	0	49.9	49.9	50.7	13.3	13.2	14.3	10.2	10.0	9.6
	0.1	53.1	53.0	52.8	14.1	13.9	14.8	10.6	10.5	9.8
3	-0.2	43.3	43.2	41.6	11.7	11.6	11.5	8.6	8.4	8.1
	-0.1	40.8	40.6	40.3	11.2	11.1	11.3	8.5	8.4	8.0
	0	40.3	40.2	40.7	11.2	11.1	11.3	8.5	8.4	8.1
	0.1	43.0	42.9	42.7	11.9	11.8	11.7	8.9	8.5	8.2
4	-0.2	42.4	42.2	39.6	11.7	11.3	10.8	8.6	8.3	7.8
	-0.1	40.1	39.8	38.4	11.3	11.0	10.6	7.9	7.7	7.7
	0	40.2	40.0	38.8	11.2	11.0	10.7	8.1	7.8	7.7
	0.1	43.2	42.9	40.9	11.7	11.4	11.0	8.6	8.2	7.9
5	-0.2	41.3	40.8	39.5	10.8	10.0	10.5	7.9	7.5	7.6
	-0.1	38.3	38.0	38.3	10.2	9.8	10.3	7.7	7.2	7.5
	0	38.5	38.1	38.8	10.4	9.7	10.4	7.6	7.2	7.6
	0.1	41.2	40.7	40.9	10.9	10.3	10.8	8.0	7.4	7.7

Δ=0.3		$n = 500$			$n = 100$			$n = 50$		
m	ρ_{min}	χ^2_z	F_ρ	PF_{Tu}	χ^2_z	F_ρ	PF_{Tu}	χ^2_z	F_ρ	PF_{Tu}
2	-0.2	100.0	100.0	100.0	75.2	75.1	76.5	46.2	46.1	47.5
	-0.1	100.0	100.0	100.0	75.5	75.2	77.7	46.2	46.0	48.6
3	-0.2	100.0	100.0	100.0	65.3	65.0	66.8	36.6	36.2	37.8
	-0.1	100.0	100.0	100.0	65.4	65.1	68.2	36.1	36.1	38.8
4	-0.2	100.0	100.0	100.0	64.5	63.8	65.1	33.8	32.8	35.6
	-0.1	100.0	100.0	100.0	64.5	64.0	66.7	33.7	33.0	36.8
5	-0.2	100.0	100.0	100.0	64.3	63.1	65.7	33.4	32.2	35.4
	-0.1	100.0	100.0	100.0	64.5	63.1	67.7	33.3	31.9	36.7

values from a logistic distribution with mean 0 and variance 1 and then multiplied this vector by the Cholesky decomposition of the true covariance matrix (Equation 2.2) to obtain data with the desired correlations and variances. The same tests for each setting, autoregressive and constant cross-correlation were performed and the resulting sizes of the tests are shown in Tables 2.5 and 2.6, respectively. In Table 2.5, we see that the sizes of both the χ_z^2 and χ_ρ^2 tests are inflated, indicating that mild lack of normality in addition to a model mis-specification (autoregressive nuisance correlations) has some impact on the size of the tests. In Table 2.6, the same χ_z^2, χ_ρ^2 and F_ρ tests were performed. The size of all three tests are inflated, especially in data with ρ_{min} far from 0 and with a larger number of time points. The test using F_ρ produces nominal sizes in most scenarios although the size is inflated in some cases. Nonetheless, the test using F_ρ still performs better than the χ_z^2 and χ_ρ^2 tests. Table 2.6 indicates that as long as there is no model mis-specification (constant nuisance correlations), mild lack of normality has some impact on the size of the tests, but not much.

2.4 Motivating example

Since the original data that motivated our work has only 18 subjects whose observations at the third time point were completely missing, we analyzed a similar set of data from a longitudinal periodontal study described by Kinney et al. (2011) and Ramseier et al. (2009). 79 subjects contributed complete data during 12-month study, including levels of four serum-derived biomarkers: TNF-α, Calprotectin, metalloproteinase-8 (MMP-8), and MMP-9, and four saliva-derived biomarkers: IL-1β, MMP-8, MMP-9 and OPG, and three periodontal plaque biofilm pathogens: *P.gingivalis*, *T.forsythia*, and *T.denticola* examined at baseline (denoted Month 0), 6

Table 2.5: Size of tests using χ^2_ρ and χ^2_z when data follows a multivariate logistic distribution with autoregressive nuisance correlation. Each value is the percentage of simulations out of 5,000 in which the null hypothesis is rejected. χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s.

		$n = 500$		$n = 100$		$n = 50$	
m	ρ_{min}	χ^2_z	χ^2_ρ	χ^2_z	χ^2_ρ	χ^2_z	χ^2_ρ
2	0.2	6.1	6.1	6.1	6.2	5.8	6.0
	0.3	5.8	5.8	6.0	6.1	5.9	5.9
	0.4	6.1	6.1	6.1	6.1	6.0	5.6
	0.5	6.3	6.2	6.8	6.4	6.2	5.6
3	0.2	6.1	6.3	6.5	6.9	6.3	7.0
	0.3	5.8	5.8	6.1	6.2	5.6	5.9
	0.4	5.8	6.0	5.8	5.9	5.5	5.5
	0.5	6.5	6.5	6.2	6.1	6.0	5.6
4	0.2	6.4	6.6	6.4	6.9	6.4	7.5
	0.3	5.7	5.8	5.9	6.3	5.7	6.5
	0.4	5.5	5.6	6.0	6.2	5.8	6.3
	0.5	6.0	6.2	6.4	6.5	6.5	6.2
5	0.2	6.9	7.1	6.6	7.3	6.1	7.3
	0.3	6.3	6.3	5.9	6.4	5.5	6.5
	0.4	6.1	6.2	5.9	6.0	5.5	5.8
	0.5	6.9	6.9	6.3	6.2	6.2	6.0

Table 2.6: Size of tests using χ^2_ρ and χ^2_z with constant nuisance correlations, while the data follows a multivariate logistic distribution. Each value is the percentage of simulations out of 5,000 in which the null hypothesis is rejected. χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s; F_ρ=$\mathcal{F}$-test with estimated degrees of freedom.

		$n = 500$			$n = 100$			$n = 50$		
m	ρ_{min}	χ^2_z	χ^2_ρ	F_ρ	χ^2_z	χ^2_ρ	F_ρ	χ^2_z	χ^2_ρ	F_ρ
2	-0.2	5.8	5.8	5.8	5.2	5.3	5.0	5.9	6.0	5.7
	-0.1	5.1	5.1	5.0	5.1	5.2	5.1	5.3	5.6	5.2
	0	4.8	4.8	4.7	5.2	5.4	5.1	5.2	5.5	5.1
	0.1	5.0	5.0	5.0	5.3	5.5	5.2	5.4	5.7	5.3
	0.2	5.4	5.4	5.4	5.8	5.9	5.8	5.6	5.7	5.3
3	-0.2	6.2	6.2	6.1	6.5	6.9	6.3	5.8	6.4	5.6
	-0.1	5.7	5.8	5.6	6.1	6.5	6.0	4.8	5.3	4.7
	0	4.8	4.9	4.8	5.7	6.0	5.6	4.7	5.6	4.8
	0.1	5.0	5.1	5.0	5.8	6.3	5.6	5.2	5.9	5.0
	0.2	6.0	6.0	5.9	6.4	6.7	6.1	6.1	6.6	5.8
4	-0.2	6.0	6.2	5.9	6.6	7.0	6.3	6.0	7.1	5.8
	-0.1	5.2	5.3	5.1	5.8	6.1	5.4	5.4	6.4	5.2
	0	5.2	5.3	5.2	5.3	5.8	5.0	5.1	6.3	4.9
	0.1	5.5	5.6	5.3	5.4	6.1	5.2	5.7	6.5	5.5
	0.2	6.2	6.3	6.0	6.4	7.0	6.1	6.7	7.5	6.4
5	-0.2	6.0	6.2	6.0	6.3	7.2	5.8	6.1	7.7	5.9
	-0.1	5.4	5.4	5.2	5.4	6.2	5.0	5.3	6.7	5.0
	0	5.1	5.3	5.0	5.0	5.8	4.8	5.2	6.6	4.9
	0.1	5.2	5.3	5.1	5.4	6.1	4.9	5.7	7.0	5.3
	0.2	6.3	6.5	6.3	6.8	7.5	6.4	6.7	8.3	6.5

months and 12 months. Ramseier et al. (2009) found that the concentration levels of MMP-8, MMP-9 and calprotectin were strong predictors of periodontitis and all three plaque pathogens demonstrated stronger associations than the four biomarkers. Considering the ability of both biomarkers and pathogens as predictors of periodontal disease, it is natural to expect some degree of correlation between them. Our primary question is whether or not the correlation between each biomarker and each pathogen is constant over time. Before analyzing the data, we first added 1.0 to all measures (as some had values equal to zero) and then took the natural logarithm of each.

2.4.1 Serum biomarker dataset

Table 2.7 contains the serial correlations of each serum biomarker with each pathogen after the transformation described above. Most serial correlations ranged between -0.2 and 0.2. The range of minimum serial correlation was $[-0.2, 0]$ and the difference between the largest and smallest correlation coefficients over time varied from 0.09 for MMP-8 and *T.denticola* to 0.33 for MMP-9 and *T.gingivalis*. Furthermore, most combinations had both positive and negative correlations over the three time points. To test the hypothesis time-homogeneous correlation between a serum biomarker and a pathogen, χ^2_z, χ^2_ρ and F tests were performed, and Table 2.7 contains the resulting p-values. All three tests gave comparable results and we fail to find evidence for concluding that correlation varies over time for most biomarker-pathogen combinations. However, we do find significant heterogeneity exists between TNF-α and *T.forsythia*, MMP-9 and *P.gingivalis*, and MMP-8 and *P.gingivalis*. In these pathogen/biomarker pairs, the correlation is highest at six months, which is the end of the disease monitoring (no treatment given) period when periodontal damage would be greatest. During the disease recovery period (6 to 12 months), treatment

was given to the patients, thereby repairing periodontal damage. Previous biological findings found the pathogenicity of the red complex, especially *P.gingivalis*, is related to periodontal tissue destruction associated with periodontitis (both soft and hard tissues), while MMP-8 and MMP-9 proteins are triggered by periodontitis to follow an anti-inflammatory process and play an important role in inhibiting periodontal destruction (Kuula et al., 2009; Gamonal et al., 2011). Thus, one might hypothesize that correlation is strongest when periodontal disease is greatest and lowest when periodontal disease is low; this conjecture agrees with our findings, which produce an interesting hypothesis that could be investigated in a larger study.

Table 2.7: Empirical serial correlations between serum biomarkers and pathogens at 0, 6, and12 months and resulting p-values for test of equality; χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s; F_ρ=F-test with estimated denominator degrees of freedom.

Pathogen	Biomarker	Serial correlation			p-value		
		0	6	12	χ^2_z	χ^2_ρ	F_ρ
P.gingivalis	TNF-α	-0.17	0.07	-0.07	0.107	0.103	0.110
	Calprotectin	-0.18	-0.01	0.11	0.147	0.139	0.146
	MMP-8	-0.01	0.28	0.04	0.059	0.052	0.058
	MMP-9	0.12	0.19	-0.14	0.021	0.018	0.022
T.forsythia	TNF-α	-0.19	0.07	-0.19	0.005	0.005	0.007
	Calprotectin	-0.21	-0.02	0.05	0.160	0.151	0.159
	MMP-8	0.00	0.22	0.13	0.182	0.178	0.186
	MMP-9	0.13	0.19	-0.03	0.136	0.133	0.141
T.denticola	TNF-α	-0.13	-0.04	0.03	0.365	0.362	0.368
	Calprotectin	-0.17	0.01	0.13	0.109	0.101	0.109
	MMP-8	-0.01	0.08	0.00	0.785	0.784	0.786
	MMP-9	0.17	0.02	-0.02	0.271	0.264	0.272

2.4.2 Salivary biomarker dataset

Table 2.8 contains the serial correlations of each salivary biomarker with each pathogen after the transformation described above. Most serial correlations ranged between -0.1 and 0.5. The range of minimum serial correlation was $[-0.15, 0.2]$, and the difference between the largest and smallest correlation coefficients over time varied from 0.06 for IL-1β and *T.forsythia* to 0.66 for MMP-8 and *T.forsythia*. The table shows that all three tests gave comparable results and we fail to find evidence for concluding that correlation varies over time for most biomarker-pathogen combinations. However, we do find significant heterogeneity exists in pairs between salivary MMP-8 and *T.forsythia* (max-min difference is 0.66), salivary MMP-8 and *P.gingivalis* (max-min difference is 0.37), salivary MMP-9 and *T.forsythia* (max-min difference is 0.45), salivary OPG and *T.denticola* (max-min difference is 0.44). Here, MMP-8 again shows its strong correlation with different rex complex pathogens, consistent with the findings reviewed by Kuula et al. (2009) and Gamonal et al. (2011).

2.5 Conclusion

In this chapter we examined methods that are a modification to the test of Olkin and Finn (1990), to perform tests of equality of correlation coefficients for longitudinal studies. Our method assumes fewer nuisance parameters that require estimation for our test statistics, thereby reducing computational burden. We describe our model for the joint distribution of a biomarker and a plaque pathogen and derive asymptotic distributions for testing homogeneity of their correlation over time, using both untransformed (χ^2_ρ test) and Fisher's Z-transformed (χ^2_z test) sample correlation coefficients. Since the χ^2_ρ test tends to be liberal in small samples, we proposed an

Table 2.8: Empirical serial correlations between salivary biomarkers and pathogens at 0, 6, and12 months and resulting p-values for test of equality; χ^2_z=Test based on Fisher's Z-transformation; χ^2_ρ=Test based on original $\boldsymbol{\rho_j}$'s; F_ρ=F-test with estimated denominator degrees of freedom.

Pathogen	Biomarker	Serial correlation			p-value		
		0	6	12	χ^2_z	χ^2_ρ	F_ρ
P.gingivalis	IL-1β	0.21	0.04	0.07	0.680	0.674	0.680
	MMP-8	0.38	0.40	0.03	0.099	0.104	0.117
	MMP-9	0.30	0.20	-0.03	0.330	0.324	0.336
	OPG	0.26	0.02	-0.08	0.262	0.245	0.259
T.forsythia	IL-1β	0.23	0.26	0.20	0.948	0.948	0.948
	MMP-8	0.35	0.56	-0.10	<0.001	<0.001	<0.001
	MMP-9	0.30	0.13	-0.15	0.101	0.088	0.101
	OPG	0.27	-0.01	0.16	0.360	0.354	0.366
T.denticola	IL-1β	0.36	-0.03	0.16	0.135	0.123	0.137
	MMP-8	0.41	0.41	0.20	0.351	0.367	0.379
	MMP-9	0.35	0.41	0.07	0.227	0.232	0.246
	OPG	0.38	0.21	-0.06	0.098	0.089	0.102

alternate F_ρ statistic derived from χ^2_ρ to maintain a nominal test size. The empirical size and power of our proposed tests in a variety of settings motivated by the data collected in our motivating study were collected. Conclusions are: (1) When data is under our assumed model, F_ρ tests have nominal size, while χ^2_z test has inflated size in some settings in the small sample, and χ^2_ρ test has more inflated size in small sample and some medium sized sample. Accordingly, χ^2_z and F_ρ tests have similar power while χ^2_z is a little superior. (2) When data is mis-specified such that the cross correlations are not kept constant as assumed, χ^2_z is generally robust. (3) When data are heavier tailed, the size of all tests are inflated in small sample, while the F_ρ is still the closest to the nominal size than χ^2_z. χ^2_ρ is the most liberal. (4) When the absolute value of ρ_{min} increases, the power of all three tests increases as well.

Lastly, we found suitable values for λ and δ needed for our $\mathcal{F}$-test through a grid search of possible values. However, the values could be found directly by equating the first two moments of $\lambda\chi^2_\rho/(m-1)$ and $F(m-1,d)$. However, this approach requires use of the Delta method to obtain the variance of a function of all the correlation parameters, which is computationally intensive. Finding these values and comparing them to those we used would prove interesting to determine if the added computational burden is warranted.

CHAPTER III

A Bayesian Approach of Testing for Serial Homogeneity in the Correlation of Longitudinally Measured Biomarkers

3.1 Introduction

Our method is motivated by studies that measure several biomarkers longitudinally with the goal of predicting for future disease occurrence. An example is the data collected from a small longitudinal study of gingivitis, or inflammation of the gums (gingivae) (Salvi et al., 2010). Our current goal is to determine how correlated the biomarkers are with each other at each time point and if the serial correlations are homogeneous and can be pooled into a single time-invariant value that quantifies the correlation of the biomarkers. The method we introduced in Chapter 2 used asymptotic frequentist methods for inference, which in small samples, may fail to have nominal size or satisfactory power. In this chapter, we explore alternative approaches using Bayesian inferential methods, specifically posterior credible intervals and Bayesian posterior predictive p-values, both of which we introduce next.

3.1.1 Posterior credible intervals

In some simple circumstances such as longitudinal data that have only two time points, we may use a credible interval, or Bayesian confidence interval, to construct a Bayesian test. In Bayesian statistics, a credible interval is an interval within the posterior distribution of a parameter, describing the uncertainty of that parameter. A $100(1-\alpha)\%$ credible interval for a parameter has the property that the posterior probability that the parameter lies in the interval is $1-\alpha$. To test if two correlation parameters, ρ_1 and ρ_2, are equal, we could draw values of ρ_1 and ρ_2 from their joint posterior distribution many times to obtain a number of samples of $\rho_1-\rho_2$ from which we can compute a $100(1-\alpha)\%$ credible interval for $\rho_1-\rho_2$. If the credible interval includes zero, we have evidence for H_0.

A $100(1-\alpha)\%$ credible interval can be defined in several ways. When the marginal posterior distribution is symmetric, we can easily obtain the credible interval by calculating the $(100\alpha/2)$th and $100(1-\alpha/2)$th quantiles of the posterior sample. If a posterior distribution is not symmetric, we can choose the narrowest interval, also called the $100(1-\alpha)\%$ highest posterior density (HPD) interval, since every point covered in the interval has higher probability density than any point outside the interval (Chen and Shao, 1999).

3.1.2 Bayesian predictive p-value

In data with more than two time points, determining an HPD interval is more difficult. Thus, we instead approach our hypothesis testing problem as a model selection problem. Specifically, if we take the null and alternative hypotheses as two different models, we then compare which model is more likely given the data. There are many Bayesian model selection methods; the most commonl of which is the Bayes

factor. However, Bayes factors are often difficult to calculate, especially for models that involve many random effects, large numbers of unknown parameters or improper priors. Another useful tool in checking a model's adequacy is the posterior predictive p-value via MCMC as proposed by Meng (1994) and Gelman et al. (1996). The advantage of this method is the ability to assess the fit of a single model without the need for an alternative model. Let $\boldsymbol{D}$ denote the observed data, H denote the model to be checked, $\boldsymbol{\phi}$ denote the unknown model parameter, and $T(\boldsymbol{D})$ denote a test statistic. If $\boldsymbol{D}^{rep}$ denotes a replication of $\boldsymbol{D}$ that could be observed and has the distribution $P_A[\boldsymbol{D}^{rep}|H, \boldsymbol{\phi}]$, where A represents auxiliary statistics that are functions of the original data and are assumed to be constant in each replication, the classical p-value is

$$p_c(\boldsymbol{D}, \boldsymbol{\phi}) = P_A[T(\boldsymbol{D}^{rep}) \geq T(\boldsymbol{D})|H, \boldsymbol{\phi}].$$

Note that the value of p_c is obtainable only when it does not contain unknown nuisance parameters.

If we denote the posterior distribution of $\boldsymbol{\phi}$ as $P(\boldsymbol{\phi}|H, \boldsymbol{D})$, the posterior predictive distribution of the replicated data $\boldsymbol{D}^{rep}$ is

$$P_A(\boldsymbol{D}^{rep}|H, \boldsymbol{D}) = \int P_A(\boldsymbol{D}^{rep}|H, \boldsymbol{\phi}) P(\boldsymbol{\phi}|H, \boldsymbol{D}) d\boldsymbol{\phi}.$$

The corresponding tail-area probability of the posterior distribution of $T(\boldsymbol{D})$, is an example of posterior predictive p-value

$$p_b(\boldsymbol{D}) = P_A[T(\boldsymbol{D}^{rep}) \geq T(\boldsymbol{D})|H, \boldsymbol{D}] = \int p_c(\boldsymbol{D}, \boldsymbol{\phi}) P(\boldsymbol{\phi}|H, \boldsymbol{D}) d\boldsymbol{\phi}$$

which is the classical p-value, $p_c(\boldsymbol{D}, \boldsymbol{\phi})$ averaged over the posterior distribution of $\boldsymbol{\phi}$.

The posterior predictive p-value is generalized as a tail-area probability of the posterior distribution of a discrepancy measure. The discrepancy measure is an equation that either involves nuisance parameters, denoted as $Q(\boldsymbol{D};\boldsymbol{\phi})$, or does not involve nuisance parameters, such as a test statistic like $T(\boldsymbol{D}$, the example above. The reference distribution for a selected discrepancy measure, $Q(\boldsymbol{D};\boldsymbol{\phi})$, is derived from the joint posterior distribution of $\boldsymbol{D}^{rep}$ and $\boldsymbol{\phi}$:

$$P_A(\boldsymbol{D}^{rep},\boldsymbol{\phi}|H,\boldsymbol{D}) = P_A(\boldsymbol{D}^{rep}|H,\theta)P(\boldsymbol{\phi}|H,\boldsymbol{D})$$

Then the tail-area probability corresponding to the posterior reference distribution of Q is

$$\begin{aligned} p_b(\boldsymbol{D}) &= P_A[Q(\boldsymbol{D}^{rep};\boldsymbol{\phi}) \geq Q(\boldsymbol{D};\boldsymbol{\phi})|H,\boldsymbol{D}] \\ &= \int P_A[Q(\boldsymbol{D}^{rep};\boldsymbol{\phi}) \geq Q(\boldsymbol{D};\boldsymbol{\phi})|H,\boldsymbol{\phi}]P(\boldsymbol{\phi}|H,\boldsymbol{D})d\boldsymbol{\phi} \end{aligned}$$

With this generalized formulation, we are able to compare directly the discrepancy between the observed data and the model when the null hypothesis is true.

3.1.3 Bayesian modeling of multivariate data parameters

Markov Chain Monte Carlo (MCMC) is most often used to obtain samples from a distribution in complex settings. In our setting, sampling parameters of the model under the null hypothesis involves modeling a variance-covariance structure $\boldsymbol{\Sigma}$; Daniels and Kass (1999) reviewed several prior choices for $\boldsymbol{\Sigma}$. The first choice is the inverse Wishart distribution, which is the conjugate prior with a multivariate normal likelihood. However, this prior allows only one precision parameter for all elements in $\boldsymbol{\Sigma}$ and does not have enough flexibility for our setting. Furthermore, when the

sample size is small, the specification of the scale matrix (the parameter defining an Inverse-Wishart prior) can be quite influential. The second prior choice for $\mathbf{\Sigma}$ is a nonconjugate reference prior, such as Jeffreys' prior. However, such a prior can lead to an improper posterior distribution. A third class of priors for $\mathbf{\Sigma}$ are hierarchical priors, some of which are based on different parameterizations. One hierarchical prior is a Wishart prior that considers the degrees of freedom ν to be unknown and to vary uniformly between $m-1$ and a large number, where m is the dimension of $\mathbf{\Sigma}$.

Finally, a direct variance/covariance decomposition suggested by Barnard et al. (2000) that allows us to work with standard deviations and a correlation matrix is of most interest to us. The idea is simple: $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma} = diag(\boldsymbol{S})\ \boldsymbol{R}\ diag(\boldsymbol{S})$, where $\boldsymbol{S}$ is the vector of p standard deviations and $\boldsymbol{R}$ is the $m \times m$ correlation matrix. According to Barnard et al. (2000), their separation strategy "has a strong practical motivation since most practitioners are trained to think in terms of standard deviations and correlations." In addition, different priors can be put on $\boldsymbol{S}$ and $\boldsymbol{R}$. For example, we can put a normal distribution on the Fisher's Z-transformation of the correlations: $Z_\rho = 1/2\ log[(1-\rho)/(1+\rho)] \sim \mathcal{N}(0, \tau^2)$, and as suggested by Daniels (1992), we put another prior on τ^2, i.e. $\pi(\tau^2) \propto (c+\tau^2)^{-2}$, where c is chosen to be $1/(n-3)$, the asymptotic variance of Z_ρ. Barnard et al. (2000) suggested assuming uniform priors for correlations, and they also showed that it is easy to derive constraints on the correlations to keep $\boldsymbol{R}$ positive definite. They also suggested log-normal priors for the parameters in $\boldsymbol{S}$. In our study, we choose to use marginally uniform priors for correlations, and both informative Gamma priors and improper priors for the precision parameters.

In Section 3.2, the specific priors and the full conditional distributions of each parameter are defined with detail. We then present the posterior predictive p-value

method in our data setting, as well as a detailed procedure of sampling from the posterior distribution and obtaining posterior p-values. Since credible intervals are straightforward, details about them are not presented in Section 3.2. In Section 3.3, we assess the empirical size and power of our credible interval method and proposed posterior predictive tests in a variety of settings motivated by the data collected in the pilot study described earlier and also apply our methods to data from another longitudinal periodontal study. Section 3.4 contains our concluding remarks.

3.2 Methods

3.2.1 Notation

The setting is the same as Chapter 2. We have n subjects who are each examined sequentially at times $t_1 < t_2, \ldots, < t_m$. Let X_{ij} and Y_{ij}, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$, denote the respective values of biomarker X and pathogen Y collected from subject i at time t_j. Marginally, we assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$, where μ_{xj} and μ_{yj} are $m \times 1$ vectors of parameters quantifying the means of X_{ij} and Y_{ij}, respectively. The elements of $\boldsymbol{X_i}$ are assumed to be exchangeably correlated with each other with correlation ρ_x, and the elements of $\boldsymbol{Y_i}$ are exchangeably correlated with each other with correlation ρ_y. We also assume a common cross-correlation, ρ_{xy} between X_{ij} and Y_{ik}, where $j \neq k$. The parameters we are interested in are $\rho_1, \rho_2, \ldots, \rho_m$, the within-time correlation of X_{ij} and Y_{ij} defined to be $\rho_j = Corr(X_{ij}, Y_{ij}), j = 1, 2, \ldots m$, while all other parameters are nuisance. Let $\boldsymbol{D}_i$ denote the $(2m \times 1)$ longitudinal vector of pairs of biomarker and pathogen for subject i, and $\boldsymbol{D}$ denote the observations for all subjects. $\boldsymbol{D}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ as defined in Equations (2.1)

and (2.2).

3.2.2 Prior specifications

According to the direct decomposition strategy suggested by Barnard et al. (2000), we decompose $\boldsymbol{\Sigma}$ in Equation (2.2) into $\boldsymbol{S}$=$\{\sigma_1, \tau_1, \sigma_2, \tau_2, \cdots, \sigma_m, \tau_m\}$ and

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho_1 & \rho_x & \rho_{xy} & \cdots & \rho_x & \rho_{xy} \\ \rho_1 & 1 & \rho_{xy} & \rho_y & \cdots & \rho_{xy} & \rho_y \\ \rho_x & \rho_{xy} & 1 & \rho_2 & \cdots & \rho_x & \rho_{xy} \\ \rho_{xy} & \rho_y & \rho_2 & 1 & \cdots & \rho_{xy} & \rho_y \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_x & \rho_{xy} & \rho_x & \rho_{xy} & \cdots & 1 & \rho_m \\ \rho_{xy} & \rho_y & \rho_{xy} & \rho_y & \cdots & \rho_m & 1 \end{pmatrix} \tag{3.1}$$

so that $\boldsymbol{\Sigma} = diag(\boldsymbol{S})\ \boldsymbol{R}\ diag(\boldsymbol{S})$. Since under our null hypothesis, $\rho_1 = \cdots = \rho_m = \rho_0$, where ρ_0 is not specified, all serial correlations from ρ_1 to ρ_m in the above expression are replaced by ρ_0. We set an uninformative prior for mean parameters μ_{xj} and μ_{yj}, $\pi(\mu_{xj}) \propto 1$, $\pi(\mu_{yj}) \propto 1$, and a Unif(-1,1) prior for $\{\rho_0, \rho_x, \rho_y, \rho_{xy}\}$. Let $\mathcal{A}(\rho)$ be the range of all correlation parameters such that the correlations are bounded between -1 and 1 and the $\boldsymbol{R}$ matrix is positive definite. Two sets of priors were specified for the precision parameters: (1) an improper prior: $\pi(\sigma_j^{-2}) \propto \sigma_j^2$, $\pi(\tau_j^{-2}) \propto \tau_j^2$; (2) an informative Gamma prior with both its shape and rate parameters being 2 for the precision parameters in standardized data.

For specification (1), the joint posterior is

$$
\begin{aligned}
& \pi(\boldsymbol{\mu}, \sigma_1^{-2}, \tau_1^{-2}, \sigma_2^{-2}, \tau_2^{-2}, \cdots, \sigma_m^{-2}, \tau_m^{-2}, \rho_0, \rho_x, \rho_y, \rho_{xy} \mid \boldsymbol{D}) \\
\propto \;& \sigma_1^2\ \tau_1^2\ \sigma_2^2\ \tau_2^2\ \cdots\ \sigma_m^2\ \tau_m^2\ I[\rho' s \in \mathcal{A}(\rho)] |\boldsymbol{\Sigma}|^{-\frac{n}{2}} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{D}_i - \boldsymbol{\mu})\} \\
\propto \;& (\sigma_1^{-2}\ \tau_1^{-2}\ \sigma_2^{-2}\ \tau_2^{-2}\ \cdots\ \sigma_m^{-2}\ \tau_m^{-2})^{\frac{n}{2}-1} I[\rho' s \in \mathcal{A}(\rho)] |\boldsymbol{R}|^{-\frac{n}{2}} \\
& exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{D}_i - \boldsymbol{\mu})\}
\end{aligned}
$$

The full conditional distribution for $\boldsymbol{\mu}$ given all other parameters and the data is $[\boldsymbol{\mu}|\cdot] \sim MVN(\bar{\boldsymbol{D}}, \boldsymbol{\Sigma}/n)$, and the full conditional distribution of each σ_j^{-2} or τ_j^{-2} is

$$
\begin{aligned}
\pi(\sigma_j^{-2}|\cdot) \;&\propto\; (\sigma_j^{-2})^{\frac{n}{2}-1} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\sigma_j^2}^{-1}(\boldsymbol{D}_i - \boldsymbol{\mu})\} \\
\pi(\tau_j^{-2}|\cdot) \;&\propto\; (\tau_j^{-2})^{\frac{n}{2}-1} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\tau_j^2}^{-1}(\boldsymbol{D}_i - \boldsymbol{\mu})\}
\end{aligned}
$$

where $\boldsymbol{\Sigma}_{\sigma_j^2}$ and $\boldsymbol{\Sigma}_{\tau_j^2}$ denote covariance matrices in which all parameters are fixed except for σ_j^2 or τ_j^2, respectively.

The full conditional distributions of ρ_0, ρ_x, ρ_y and ρ_{xy} have similar forms. As an example, the full conditional of ρ_0 can be written as

$$
\pi(\rho_0|\cdot) \propto I[\rho_0 \in \mathcal{A}(\rho)] |\boldsymbol{R}_{\rho_0}|^{-\frac{n}{2}} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\rho_0}^{-1}(\boldsymbol{D}_i - \boldsymbol{\mu})\}
$$

where $\boldsymbol{\Sigma}_{\rho_0}$ and $\boldsymbol{R}_{\rho_0}$ denote a covariance matrix and correlation matrix, respectively, in which all parameters are fixed except for ρ_0, in which all parameters are fixed except for ρ_0.

With the Gamma (2,2) prior, we work on the standardized data, denoted as $\boldsymbol{D}^*$ with parameters $\boldsymbol{\mu}^*, \sigma_1^{*2}, \tau_1^{*2}, \sigma_2^{*2}, \tau_2^{*2}, \cdots, \sigma_m^{*2}, \tau_m^{*2}, \rho_0, \rho_x, \rho_y, \rho_{xy}$. Note that $\boldsymbol{\mu}^*$ should

be close to $\mathbf{0}$, and $\sigma_1^{*-2}, \tau_1^{*-2}, \sigma_2^{*-2}, \tau_2^{*-2}, \cdots, \sigma_m^{*-2}, \tau_m^{*-2}$ are centered around $\mathbf{1}$, with a Gamma(2,2) prior assigned to each. The joint posterior is

$$\begin{aligned}
& \pi(\boldsymbol{\mu}^*, \sigma_1^{*-2}, \tau_1^{*-2}, \sigma_2^{*-2}, \tau_2^{*-2}, \cdots, \sigma_m^{*-2}, \tau_m^{*-2}, \rho_0, \rho_x, \rho_y, \rho_{xy} \mid \boldsymbol{D^*}) \\
\propto \; & \{\sigma_1^{*-2}\ \tau_1^{*-2}\ \cdots\ \sigma_m^{*-2}\ \tau_m^{*-2}\} exp(-\sigma_1^{*-2} - \tau_1^{*-2} - \cdots\ \sigma_m^{*-2} - \tau_m^{*-2}) \\
& I[\rho' s \in \mathcal{A}(\rho)] |\boldsymbol{\Sigma}^*|^{-\frac{n}{2}} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}^{*-1}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)\} \\
\propto \; & (\sigma_1^{*-2}\ \tau_1^{*-2}\ \cdots\ \sigma_m^{*-2}\ \tau_m^{*-2})^{\frac{n}{2}+1} I[\rho' s \in \mathcal{A}(\rho)] |\boldsymbol{R}|^{-\frac{n}{2}} \\
& exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}^{*-1}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*) - \sigma_1^{*-2} - \tau_1^{*-2} - \cdots\ \sigma_m^{*-2} - \tau_m^{*-2}\}
\end{aligned}$$

The full conditional distribution for $\boldsymbol{\mu^*}$ given all other parameters and data is

$$[\boldsymbol{\mu}^* | \cdot] \sim MVN(\bar{\boldsymbol{D}}^*, \boldsymbol{\Sigma}^*/n)$$

The full conditional distribution of each σ_j^{*-2} or τ_j^{*-2} is

$$\begin{aligned}
\pi(\sigma_j^{*-2} | \cdot) &\propto (\sigma_j^{*-2})^{\frac{n}{2}+1} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}_{\sigma_j^{*2}}^{*-1}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*) - \sigma_j^{*-2}\} \\
\pi(\tau_j^{*-2} | \cdot) &\propto (\tau_j^{*-2})^{\frac{n}{2}+1} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}_{\tau_j^{*2}}^{*-1}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*) - \tau_j^{*-2}\}
\end{aligned}$$

where $\boldsymbol{\Sigma}^*_{\sigma_j^{*2}}$ and $\boldsymbol{\Sigma}^*_{\tau_j^{*2}}$ denote covariance matrix in which all parameters are fixed but σ_j^{*2} or τ_j^{*2}.

The full conditional distributions of each correlation parameter, including ρ_0, ρ_x, ρ_y and ρ_{xy} have similar expressions. As an example, the full conditional of ρ_0 can be written as

$$\pi(\rho_0 | \cdot) \propto I[\rho_0 \in \mathcal{A}(\rho)] |\boldsymbol{R}_{\rho_0}|^{-\frac{n}{2}} exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)^T \boldsymbol{\Sigma}_{\rho_0}^{*-1}(\boldsymbol{D}_i^* - \boldsymbol{\mu}^*)\}$$

where $\boldsymbol{\Sigma}^*_{\rho_0}$ and $\boldsymbol{R}_{\rho_0}$ denote a covariance matrix and correlation matrix, respectively, in which all parameters are fixed except for ρ_0.

3.2.3 Bayesian posterior predictive methods for longitudinal data with two or more time points

We are interested in testing the hypotheses $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m$ versus H_a : two or more of $\rho_1, \rho_2, \ldots, \rho_m$ are unequal. For time j, let $\boldsymbol{X_{.j}} = \{X_{1j}, X_{2j}, \cdots, X_{nj}\}$ and $\boldsymbol{Y_{.j}} = \{Y_{1j}, Y_{2j}, \cdots, Y_{nj}\}$ denote the respective vectors of all subjects' values of biomarker X and pathogen Y. For $j \neq k$, we then denote $\tilde{S}_{XX_j}$ as the sample variance of $\boldsymbol{X_{.j}}$, $\tilde{S}_{YY_j}$ as the sample variance of $\boldsymbol{Y_{.j}}$, $\tilde{S}_{XX_{jk}}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{X_{.k}}$, $\tilde{S}_{YY_{jk}}$ as the sample covariance between $\boldsymbol{Y_{.j}}$ and $\boldsymbol{Y_{.k}}$, $\tilde{S}_{XY_j}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{Y_{.j}}$, and $\tilde{S}_{XY_{jk}}$ as the sample covariance between $\boldsymbol{X_{.j}}$ and $\boldsymbol{Y_{.k}}$. Let $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m$, $\hat{\rho}_x$, $\hat{\rho}_y$ and $\hat{\rho}_{xy}$ be:

$$\hat{\rho}_j = \frac{\widehat{\rho_j \sigma_j \tau_j}}{\sqrt{\hat{\sigma_j^2}}\sqrt{\hat{\tau_j^2}}} = \frac{\tilde{S}_{XY_j}}{\sqrt{\tilde{S}_{XX_j}\tilde{S}_{YY_j}}} \; ; \; j = 1, 2, \ldots m$$

$$\hat{\rho}_x = \frac{\sum_{j \neq k} \widehat{\rho_x \sigma_j \sigma_k}}{\sum_{j \neq k} \sqrt{\hat{\sigma}_j^2 \hat{\sigma}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{XX_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{XX_j}\tilde{S}_{XX_k}}}$$

$$\hat{\rho}_y = \frac{\sum_{j \neq k} \widehat{\rho_y \tau_j \tau_k}}{\sum_{j \neq k} \sqrt{\hat{\tau}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{YY_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{YY_j}\tilde{S}_{YY_k}}}$$

$$\hat{\rho}_{xy} = \frac{\sum_{j \neq k} \widehat{\rho_{xy} \sigma_j \tau_k}}{\sum_{j \neq k} \sqrt{\hat{\sigma}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j \neq k} \tilde{S}_{XY_{jk}}}{\sum_{j \neq k} \sqrt{\tilde{S}_{XY_j}\tilde{S}_{XY_k}}}$$

Define θ_{jj} as

$$\theta_{jj} = \frac{1}{n}(1 - \rho_j^2)^2 \tag{3.2}$$

with corresponding estimate $\hat{\theta}_{jj} = \frac{1}{n}(1 - \hat{\rho}_j^2)^2$

and define θ_{jk} as

$$\theta_{jk} = \frac{1}{n}\left\{\frac{1}{2}\rho_j\rho_k(\rho_x^2+\rho_y^2) + \rho_{xy}^2(1+\rho_j\rho_k) + \rho_x\rho_y - \rho_{xy}(\rho_j+\rho_k)(\rho_x+\rho_y)\right\} \quad (3.3)$$

with corresponding estimate

$$\hat{\theta}_{jk} = \frac{1}{n}\left\{\frac{1}{2}\hat{\rho}_j\hat{\rho}_k(\hat{\rho}_x^2+\hat{\rho}_y^2) + \hat{\rho}_{xy}^2(1+\hat{\rho}_j\hat{\rho}_k) + \hat{\rho}_x\hat{\rho}_y - \hat{\rho}_{xy}(\hat{\rho}_j+\hat{\rho}_k)(\hat{\rho}_x+\hat{\rho}_y)\right\}.$$

Let $\boldsymbol{\Sigma}_\rho$ be an $m \times m$ matrix with diagonal element (j,j) equal to θ_{jj} as given by Equation (3.2), and off-diagonal element (j,k) equal to θ_{jk}, as given by Equation (3.3). Let $\hat{\boldsymbol{\Sigma}}_\rho$ be an $m \times m$ matrix with diagonal element (j,j) equal to $\hat{\theta}_{jj}$ and off-diagonal element (j,k) equal to $\hat{\theta}_{jk}$. Also let $\boldsymbol{L}$ be an $(m-1)\times m$ contrast matrix for the pairwise differences, i.e.

$$\boldsymbol{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}$$

To construct a posterior predictive p-value to test the null model $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m$, the test statistic $T_\rho(\boldsymbol{D}) = (\boldsymbol{L}\hat{\boldsymbol{\rho}})^T(\boldsymbol{L}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\rho}}\boldsymbol{L}^{\boldsymbol{T}})^{-1}(\boldsymbol{L}\hat{\boldsymbol{\rho}})$ and discrepancy variable $Q_\rho(\boldsymbol{D},\boldsymbol{\phi}) = (\boldsymbol{L}\hat{\boldsymbol{\rho}})^T(\boldsymbol{L}\boldsymbol{\Sigma}_{\boldsymbol{\rho}}\boldsymbol{L}^{\boldsymbol{T}})^{-1}(\boldsymbol{L}\hat{\boldsymbol{\rho}})$ are chosen. $T_\rho(\boldsymbol{D})$ is totally data based and is the same as the Wald test statistic χ^2_ρ we used in Chapter 2. $Q_\rho(\boldsymbol{D},\boldsymbol{\phi})$ contains unknown parameters, including the parameter ρ we are interested, as well as the nuisance correlation parameters ρ_x, ρ_y and ρ_{xy}.

An alternate test would be based on Fisher's Z-transformation, $\hat{\boldsymbol{z}} = \{\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_m\}$,

leading to the statistic $T_z(\boldsymbol{D}) = (\boldsymbol{L}\hat{\boldsymbol{z}})^T(\boldsymbol{L}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}\boldsymbol{L}^{\boldsymbol{T}})^{-1}(\boldsymbol{L}\hat{\boldsymbol{z}})$, in which $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$ has diagonal element (j,j) equal to $1/n$ and off-diagonal element (j,k) equal to $\hat{\theta}_{jk}/[(1-\hat{\rho}_j^2)(1-\hat{\rho}_k^2)]$. Identically, a discrepancy measure $Q_z(\boldsymbol{D},\boldsymbol{\phi}) = (\boldsymbol{L}\hat{\boldsymbol{z}})^T(\boldsymbol{L}\boldsymbol{\Sigma}_{\boldsymbol{z}}\boldsymbol{L}^{\boldsymbol{T}})^{-1}(\boldsymbol{L}\hat{\boldsymbol{z}})$ is also chosen in which $\boldsymbol{\Sigma}_{\boldsymbol{z}}$ has diagonal element (j,j) equal to $1/n$ and off-diagonal element (j,k) equal to $\theta_{jk}/[(1-\hat{\rho}_j^2)(1-\hat{\rho}_k^2)]$.

The difference between a test statistic $T(\boldsymbol{D})$ and a generalized discrepancy variable $Q(\boldsymbol{D},\boldsymbol{\phi})$ is whether it contains unknown parameters.

3.2.4 Computational details

Given a set of posterior draws of parameters using the Metropolis-Hastings (MH) algorithm within Gibbs sampling, $\boldsymbol{\phi}^j$, $j = 1,\cdots,J$, we perform the following two steps for each j:

1. Given $\boldsymbol{\phi}^j$, draw a simulated replicated data set, $\boldsymbol{D}^{rep,j}$, from the sampling distribution, $P_A(\boldsymbol{D}^{rep}|H_0,\boldsymbol{\phi}^j)$.

2. Calculate $T(\boldsymbol{D})$, $T(\boldsymbol{D}^{rep,j})$ and $Q(\boldsymbol{D},\boldsymbol{\phi}^j)$ and $Q(\boldsymbol{D}^{rep,j},\boldsymbol{\phi}^j)$.

Having obtained $T(\boldsymbol{D})$, $T(\boldsymbol{D}^{rep,j})$ and $Q(\boldsymbol{D},\boldsymbol{\phi}^j)$ and $Q(\boldsymbol{D}^{rep,j},\boldsymbol{\phi}^j)$, $j = 1,\cdots,J$, we can make a histogram of $T(\boldsymbol{D}^{rep,j})$ with $T(\boldsymbol{D})$ located on it to make a graphical assessment, and estimate p_b by the proportion of the J pairs for which $T(\boldsymbol{D}^{rep,j})$ exceeds $T(\boldsymbol{D})$, namely

$$p_b = \frac{1}{J}\sum_{j=1}^{J}\mathbf{1}[T(\boldsymbol{D}^{rep,j}) > T(\boldsymbol{D})]$$

and we can plot $Q(\boldsymbol{D}^{rep,j},\boldsymbol{\phi}^j)$ against $Q(\boldsymbol{D},\boldsymbol{\phi}^j)$ and estimate p_b by the proportion of

the J pairs for which $Q(\boldsymbol{D}^{rep,j}, \boldsymbol{\phi}^j)$ exceeds $Q(\boldsymbol{D}, \boldsymbol{\phi}^j)$, namely

$$p_b = \frac{1}{J}\sum_{j=1}^{J} \mathbf{1}[Q(\boldsymbol{D}^{rep,j}, \boldsymbol{\phi}^j) > Q(\boldsymbol{D}, \boldsymbol{\phi}^j)]$$

Since we have derived the full conditional posterior (up to a proportionality constant) of all parameters, it is convenient to compute the posterior using the Gibbs sampler. To get draws from $(\boldsymbol{\mu},\ \boldsymbol{S},\ \boldsymbol{R})$, we use Gibbs Sampler and draw $\mu_{x_1}, \mu_{y_1}, \cdots, \mu_{x_m}$, μ_{y_m} together from its multivariate normal conditional posterior, and draw each of the components from $\boldsymbol{S}$ and $\boldsymbol{R}$ one at a time. However, since the conditional posterior for each component of $\boldsymbol{S}$ and $\boldsymbol{R}$ is not a kernel of any known distribution, we need to do another MH algorithm within Gibbs sampling. While sampling components from $\boldsymbol{S}$, we perform an independent MH by choosing a Gamma distribution as the proposal density. For example, for prior specification (1), the proposal density we use to sample σ_1^{-2} is $\mathcal{G}[n/2, \sum_i(\mu_{x_1} - X_{i1})^2/2]$. Note this is actually the full conditional posterior under the special $\boldsymbol{R}$ structure where all off-diagonal elements are zero. While sampling components from $\boldsymbol{R}$, we do a random walk MH using a normal distribution truncated between -1 and 1. An extra step before updating each sample with the proposal sample is to check if $\boldsymbol{R}$ is positive definite for the proposal sample. If that condition is not satisfied, the sample takes the value of the previous sample.

3.3 Application of Methods

3.3.1 Simulation Study

We now examine the performance of the proposed credible interval methods and the posterior predictive tests under various settings for hypothetical longitudinal datasets based upon the data from our motivating example. Here we define the sim-

ulation setting of the posterior predictive method. For each subject $i = 1, 2, \ldots, n$, biomarker X and pathogen Y are both observed at m time points. We assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$, in which $\mu_{xj} = 2.5$ and $\mu_{yj} = 4.0$, $\sigma_j = 0.3$, and $\tau_j = 0.40 - 0.05(j-1)$. Note that correlation is location and scale invariant, so that our results are generalizable to other values of location and scale. In our first setting, where the model matches the correlation structure of the data, we selected the values $\rho_x = 0.5$, $\rho_y = 0.7$, and $\rho_{xy} = 0$ for the nuisance correlation parameters. These values are also set to attempt to match the true data we will present in the example. With regard to the correlation parameters of interest, $\{\rho_1, \rho_2, \ldots, \rho_m\}$, we defined a simulation setting with two quantities, $\rho_{min} \in \{-0.2, -0.1, 0, 0.1, 0.2\}$ and $\Delta \in \{0.0, 0.3\}$. We set $\rho_1 = \rho_{min}$, $\rho_m = \rho_{min} + \Delta$, and all other correlation parameters $\rho_2, \rho_3, \ldots, \rho_{m-1}$ were equally spaced between ρ_1 and ρ_m. Thus, a value of $\Delta = 0$ represents the null hypothesis, while $\Delta > 0$ represents the alternative hypothesis. For each combination of minimum serial correlation and Δ, we simulated $\boldsymbol{D_i}$=$\{X_{i1}, Y_{i1}, \cdots, X_{im}, Y_{im}\}$, the data for each subject i, from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ defined in Equations (2.1) and (2.2). We considered sample sizes of $n \in \{25, 50, 100\}$ and the number of time points $m \in \{2, 3, 4, 5\}$.

Since our assumption that the nuisance parameters ρ_x, ρ_y and ρ_{xy} are constant over time is rarely met in practice, we introduced model mis-specification in our second simulation setting. We selected the values $\rho_{x_0} = 0.5$, $\rho_{y_0} = 0.6$, $\rho_{xy_0} = 0.7$ and defined a constant $c = 1.7 \times \rho_{x_0}\rho_{y_0} = 0.51$ for the nuisance correlation parameters. Then, the within-X, within-Y and cross-X, Y correlation between time j and j are $\rho_{x_0}^{|j-j'|}$, $\rho_{y_0}^{|j-j'|}$ and $c \times \rho_{xy_0}^{|j-j'|}$ respectively. The simulation setting of $\{\rho_1, \rho_2, \ldots, \rho_m\}$ was defined by $\rho_{min} \in \{0.2, 0.3, 0.4, 0.5\}$ and $\Delta \in \{0.0, 0.3\}$. Similarly, $\rho_1 = \rho_{min}$,

$\rho_m = \rho_{min} + \Delta$, and all other correlation parameters $\rho_2, \rho_3, \ldots, \rho_{m-1}$ were equally spaced between ρ_1 and ρ_m.

The settings used for the posterior credible interval method are the same as those used with the posterior predictive method, and no model violation is assumed. The credible interval method is only applied to settings in which $m = 2$.

To evaluate the posterior predictive method in each setting, we simulated 500 datasets for $n = 100$ and 1,000 datasets for $n \in \{25, 50\}$ and ran 2,000 iterations for each. We also examined the use of both proper and improper priors for the precision parameters. With an improper prior, Metropolis-Hastings within Gibbs sampling was used as follows:

1. Draw $\mu_{x_1}, \mu_{y_1}, \cdots, \mu_{x_m}, \mu_{y_m}$ together from their multivariate normal conditional posterior.

2. Draw each of the components in $\boldsymbol{S}$ one at a time by performing an independent MH step by choosing a Gamma distribution as the proposal density for an individual precision parameter. The proposal density we use to sample σ_1^{-2} is $\mathcal{G}[n/2, \sum_i(\mu_{x_1} - X_{i1})^2/2]$.

3. Assuming our null model: $\rho_1 = \rho_2 = \cdots = \rho_m = \rho_0$, where ρ_0 is undefined, the components of R include $\{\rho_0, \rho_x, \rho_y, \rho_{xy}\}$. For sampling ρ_x and ρ_y, we do a random walk MH using a normal distribution truncated between -1 and 1. While drawing each of ρ_x, ρ_y one at a time, we draw ρ_0 and ρ_{xy} jointly from a bivariate truncated normal proposal distribution with correlation 0.6, since a large cross-correlation was observed between ρ_0 and ρ_{xy}. An extra step before updating each sample with the proposal sample is to check if $\boldsymbol{R}$ is positive definite. If that condition is not satisfied, the sample takes the value of the previous sample.

With a proper prior, each simulated data was first standardized, and then in step

1 above, instead of drawing $\boldsymbol{\mu}$, we draw $\boldsymbol{\mu}^*$ which is the standardized mean. In step 2 above, instead of drawing $\sigma_1^{-2}, \tau_1^{-2}, \sigma_2^{-2}, \tau_2^{-2}, \cdots, \sigma_m^{-2}, \tau_m^{-2}$, we draw $\sigma_1^{*-2}, \tau_1^{*-2}, \sigma_2^{*-2}, \tau_2^{*-2}, \cdots, \sigma_m^{*-2}, \tau_m^{*-2}$ which are the new precision parameters after standardization. Step 3 above is unchanged.

For credible interval method, we only consider the use of improper priors and we do not assume the two serial correlations are equal to each other. Therefore, we will sample the two individual serial correlations instead of sampling a uniform ρ_0.

The proposal variance was tuned every 25 iterations during the burn-in period for the truncated normal proposal density to get an acceptance rate of between 30% to 40%. Trace plots, autocorrelation plots and histogram plots were generated to evaluate convergence. A burn-in sample of the first 400 observations was discarded. Histograms and summary statistics including mean and 95% credible interval were obtained based on the remaining 1,600 samples. Having obtained $T(\boldsymbol{D})$, $T(\boldsymbol{D}^{rep,j})$, $Q(\boldsymbol{D}, \boldsymbol{\phi}^j)$ and $Q(\boldsymbol{D}^{rep,j}, \boldsymbol{\phi}^j)$, p-values were obtained by calculating the proportion of the J pairs for which $T(\boldsymbol{D}^{rep,j})$ exceeds $T(\boldsymbol{D})$ and the proportion of the J pairs for which $Q(\boldsymbol{D}^{rep,j}, \boldsymbol{\phi}^j)$ exceeds $Q(\boldsymbol{D}, \boldsymbol{\phi}^j)$. A small p-value indicates poor fit. The test was rejected at size level α=0.05. The size and power of the tests in each scenario were estimated from the rejection rates in 1,000 simulated datasets for $n \in \{25, 50\}$ and 500 simulated datasets for $n = 100$.

To evaluate the performance of the credible interval method, instead of computing a test statistic/discrepancy variable given each parameter drawn, the credible interval method computed the difference of the two individual serial correlations drawn during each iteration. By evaluating the 95% HPD region or credible interval based on $J = 1,600$ differences, we made a decision about whether to reject the homogeneity hypothesis according to whether the interval covered 0. We simulated 1,000 datasets,

and the size and power of the credible interval "test" were estimated from the rejection rate in these datasets.

3.3.2 Evaluation of Bayesian credible interval method

Table 3.1 presents the empirical size, e.g. $\Delta = 0$, of the Bayesian credible interval method using both 95% HPD region and quantile based (QB) credible interval for various combinations of n, m, and ρ_{min}. The two columns under each sample size show the empirical size of credible interval test based on 1,000 simulations. Based upon a 95% confidence interval around a desired size of 0.05, we would expect the number of rejections in 1,000 simulations for a nominal test would lie in the interval $(3.7, 6.4)$. Overall, the size of tests based on both HPD and QB intervals fall into the interval $(3.7, 6.4)$ while HPD tends to be more conservative and QB tend to be more liberal. HPD is thus preferred than 95% QB.

Table 3.1: Size of test based on Bayesian estimation of 95% Highest Posterior Density (HPD) regions and quantile based credible interval (QB). Each value is the percentage of simulations out of 1,000 in which the null hypothesis is rejected (when HPD or QB does not cover zero).

		$n = 100$		$n = 50$		$n = 25$	
m	ρ_{min}	HPD	QB	HPD	QB	HPD	QB
2	0	5.8	5.0	5.1	5.9	4.1	5.7
	0.1	3.9	4.7	5.1	6.2	3.8	5.3
	0.2	6.0	6.8	4.2	4.1	5.4	5.6

Table 3.2 presents the estimated power of the credible interval method using both 95% HPD region and QB when $\Delta = 0.3$. Since QB interval is more liberal than HPD, here we observe a higher power for QB than HPD. The power drops down as sample size decreases.

Table 3.2: Power of test based on Bayesian estimation of 95% Highest Posterior Density (HPD) regions and 95% quantile based credible interval (QB). Each value is the percentage of simulations out of 1,000 in which the null hypothesis is rejected (when HPD or QB does not cover zero).

		$n = 100$		$n = 50$		$n = 25$	
m	ρ_{min}	HPD	QB	HPD	QB	HPD	QB
2	-0.1	77.1	78.2	47.6	48.7	24.2	25.3
	-0.2	78.1	79.2	47.3	48.6	23.1	25.2

3.3.3 Evaluation of posterior predictive method

Table 3.3 presents the empirical size, e.g. $\Delta = 0$, of various posterior predictive tests using the four statistics T_z, T_ρ, Q_z and Q_ρ defined in Section 3.2.3 for various combinations of $n, m,$ and ρ_{min} when ρ_x, ρ_y and ρ_{xy} are constant over time, an improper prior was used to sample the precision parameters. The first two columns under each sample size show the empirical size of posterior predictive tests T_z and T_ρ, the third and fourth columns are the empirical size of posterior predictive tests Q_z and Q_ρ, obtained from 500 simulations when $n = 100$ and 1,000 simulations when $n = 50$ and 25, while the fifth column is the empirical size of the Wald test, obtained from 5,000 simulations based on the asymptotic distribution of Z-transformation. Based upon a 95% confidence interval around a desired size of 0.05, we would expect the percentage of rejections in 500, 1,000 and 5,000 simulations for a nominal test would lie in the interval $(3.1, 6.9)$, $(3.7, 6.4)$ and $(4.5, 5.6)$, respectively. Overall, all four posterior predictive tests have nominal size, regardless of the number of time points and the value of ρ_{min}. However, as the sample size drops to $n = 25$, posterior predictive tests tend to become conservative, especially when there are fewer time points. Although as all four statistics lead to conservative results, the size of T_ρ is closer to 0.05 compared to T_z. In fact, the conservativeness of the four approaches can be ordered as $T_\rho \approx T_z < Q_\rho < Q_z$. The Wald Z-test, in contrast, becomes more

liberal as the sample size decreases and as the number of time points increases. One last finding is that when the number of time points increases, the size of all four tests increases. Relative to the amount of Monte Carlo error, there is no difference in size between T_ρ, T_z and Q_ρ. Q_z is very conservative when $n = 25$.

Table 3.4 presents the empirical size of posterior predictive tests using the four statistics T_z, T_ρ, Q_z and Q_ρ when ρ_x, ρ_y and ρ_{xy} are constant over time. A Gamma (informative) prior was used to sample the precision parameters of standardized data. The sizes of T_z and T_ρ are now close, and they remain nominal while being conservative as the sample size decreases. The size of Q_z and Q_ρ, however, is much smaller than their corresponding values in Table 3.3 and too small to be explained by the Monte Carlo error. Since a discrepancy variable contains unknown parameters, Q_z and Q_ρ depend largely on the parameters sampled. Therefore, they are more sensitive to how the parameters were sampled, i.e., the parameters sampled may be biased since they are based on the standardized data.

Table 3.5 presents the empirical power of the posterior predictive tests using T_ρ, T_z, Q_ρ and Q_z and Wald Z-test at $\Delta = 0.3$ when ρ_x, ρ_y and ρ_{xy} are constant over time, and an improper prior was used to sample the precision parameters . As a general trend, the power goes down as the number of time points goes up. With a sample size of $n = 100$ or 50, all tests have similar power, but when n goes down, the power of the posterior predictive tests drops quickly. Consistent with the conservativeness level as shown in Table 3.3, it can also be noticed that T_ρ has slightly higher power than T_z, Q_ρ has slightly higher power than Q_z. As the sample size goes down, the power difference between the four statistics gets bigger. As the sample size goes up or the number of time points goes down, the power goes up.

Table 3.6 presents the empirical power of the posterior predictive tests using T_ρ,

Table 3.3: Size of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are constant over time, improper prior used to sample precision parameters. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0	3.6	3.6	3.6	3.6	4.9	5.2	5.1	4.4	4.8	5.3	4.8	4.8	3.4	4.0	5.0
	0.1	4.8	5.0	4.4	4.8	4.6	5.8	5.8	5.1	5.5	5.4	4.0	4.2	2.9	3.3	5.2
	0.2	6.8	6.8	6.6	7.2	4.9	4.8	4.7	4.3	4.4	5.5	4.0	4.1	3.3	3.5	5.3
3	0	3.8	4.0	3.8	4.0	5.3	5.8	5.7	4.9	5.6	5.5	4.8	5.0	3.6	4.3	5.3
	0.1	4.6	5.0	4.0	4.4	5.4	5.7	5.8	4.9	5.8	5.5	4.5	4.8	3.5	4.0	5.4
	0.2	5.0	4.8	4.0	4.6	5.3	4.7	4.9	3.6	4.1	5.5	3.8	4.1	3.3	4.0	5.4
4	0	4.0	3.8	4.2	4.6	5.7	6.1	6.3	5.4	6.0	5.6	4.2	4.6	3.1	3.7	5.5
	0.1	5.4	5.3	4.6	4.9	5.7	4.1	4.8	3.2	4.2	5.5	3.6	3.8	3.4	3.8	5.7
	0.2	4.0	4.4	3.2	3.8	5.7	5.0	5.1	4.1	4.7	5.5	4.4	4.5	3.9	4.8	6.1
5	0	3.8	4.0	3.6	3.8	5.4	3.9	4.2	3.6	4.1	5.2	5.3	5.7	3.4	4.1	5.9
	0.1	5.2	5.2	5.2	5.6	5.5	4.8	4.8	4.4	4.6	5.3	4.1	4.4	3.6	4.1	5.9
	0.2	5.4	5.6	5.4	5.6	5.5	5.2	5.2	3.9	4.5	5.4	4.7	5.0	3.7	3.6	6.1

Table 3.4: Size of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are constant over time, Gamma (informative) prior used to sample the precision parameters of standardized data . Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0	3.8	3.8	3.4	3.6	4.9	5.2	5.3	3.8	4.1	5.3	4.8	4.8	3.0	3.2	5.0
	0.1	4.8	4.8	4.4	4.6	4.6	5.8	5.8	5.0	5.1	5.4	4.4	4.3	2.6	2.5	5.2
	0.2	7.0	7.0	6.6	6.6	4.9	4.7	4.7	4.1	3.9	5.5	4.4	4.3	3.3	3.4	5.3
3	0	4.4	4.4	3.8	4.2	5.3	5.6	5.5	3.5	4.2	5.5	4.5	4.5	3.0	3.3	5.3
	0.1	4.0	4.2	3.4	3.6	5.4	5.6	5.7	4.5	4.8	5.5	4.4	4.7	2.6	3.0	5.4
	0.2	4.4	4.4	3.4	3.6	5.3	4.5	5.0	3.0	3.2	5.5	4.2	3.9	2.3	2.6	5.4
4	0	4.0	4.0	3.6	3.8	5.7	6.1	6.2	4.1	4.6	5.6	4.3	3.9	2.3	2.4	5.5
	0.1	4.2	4.4	3.4	3.6	5.7	5.7	5.6	2.5	3.5	5.5	4.3	4.1	2.1	2.3	5.7
	0.2	4.2	4.2	2.4	2.6	5.7	4.7	4.8	3.2	3.7	5.5	4.6	4.9	2.5	3.4	6.1
5	0	3.2	3.0	2.6	2.6	5.4	4.0	3.8	2.7	3.0	5.2	5.0	4.9	2.3	2.2	5.9
	0.1	5.0	5.2	4.0	4.4	5.5	4.6	4.7	2.9	3.0	5.3	4.3	3.9	2.1	2.1	5.9
	0.2	5.0	5.4	4.8	5.0	5.5	4.9	5.0	3.2	3.3	5.4	4.9	5.4	2.1	2.2	6.1

T_z, Q_ρ and Q_z and Wald Z-test at $\Delta = 0.3$ when ρ_x, ρ_y and ρ_{xy} are constant over time, and a Gamma (informative) prior was used to sample the precision parameters of the standardized data. As expected, the power of Q_ρ and Q_z is much lower than T_ρ and T_z.

Tables 3.7-3.10 evaluates the posterior predictive tests under model mis-specification such that the data are simulated without assuming ρ_x, ρ_y and ρ_{xy} being constant. Table 3.7 compares the empirical size of the posterior predictive tests T_z, T_ρ, Q_ρ and Q_z assuming autoregressive ρ_x, ρ_y under improper prior and Wald Z-test. When model mis-specification is present, the posterior predictive tests T_z, T_ρ and Q_z remain the nominal size while and Q_ρ is too conservative under some scenarios where $n = 50$ and $n = 25$ even having monte carlo error being considered. Table 3.8 presents the size of the posterior predictive tests under model mis-specification through sampling the variance parameters from the standardized data. T_z and T_ρ do not change much, while Q_z and Q_ρ, impacted the same way as shown in Table 3.4, become more conservative. Q_ρ falls out of the monte carlo interval while the other three are closer to nominal size compared to Q_ρ.

Regarding the empirical power under model mis-specification, Table 3.9 and 3.10 show the simulation results comparing the posterior predictive tests T_z, T_ρ, Q_z, Q_ρ, and Wald Z-test at $\Delta = 0.3$. As a general trend, power goes down as the number of time points goes up and ρ_{min} gets closer to 0. Similar to the results shown in Table 3.5, With a sample size of $n = 100$ or 50, T_z, T_ρ tests and Wald test have similar power, but when n goes down, the power of the posterior predictive tests drops. Opposite to results shown in Table 3.5, T_z here has slightly higher power than T_ρ when $n = 100$ or 50, and much higher power than T_ρ when $n = 25$ under mis-specified setting. The power of Q_z is slightly lower than T_ρ, while the power of Q_ρ is the lowest. This is

Table 3.5: Power of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are constant over time, improper prior used to sample precision parameters. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	-0.1	77.6	77.6	76.2	77.4	75.2	45.8	46.1	42.8	44.3	46.2	23.3	23.8	19.2	21.3	25.6
	-0.2	72.6	72.6	71.8	72.2	75.5	45.3	45.4	44.0	44.5	46.2	24.8	25.9	19.8	22.8	25.3
3	-0.1	64.2	64.6	62.4	63.4	65.3	36.5	36.3	35.0	35.5	36.6	18.8	19.4	16.0	17.6	19.4
	-0.2	68.2	68.2	66.8	67.8	65.4	40.6	40.5	37.6	39.4	36.1	17.7	17.5	14.9	16.2	19.6
4	-0.1	65.6	65.8	65.2	66.6	64.5	32.9	33.2	30.7	32.2	33.8	18.2	19.1	15.2	17.1	18.0
	-0.2	62.8	63.4	61.0	61.8	64.5	33.8	34.1	32.4	33.6	33.7	17.4	17.9	15.2	16.4	17.9
5	-0.1	63.0	62.4	62.2	63.0	64.3	28.9	29.1	27.6	28.7	33.4	14.5	15.3	12.6	14.5	18.3
	-0.2	60.2	60.2	59.2	60.6	64.5	35.0	35.3	33.3	35.5	33.3	14.9	15.6	12.6	13.0	18.4

Table 3.6: Power of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are constant over time, Gamma (informative) prior used to sample the precision parameters of standardized data. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	-0.1	77.0	77.2	75.8	75.8	75.2	46.0	46.1	41.4	43.1	46.2	23.7	23.4	18.5	20.1	25.6
	-0.2	72.6	72.6	70.6	71.2	75.5	45.1	45.2	42.6	43.3	46.2	25.2	25.3	17.6	19.6	25.3
3	-0.1	65.0	65.0	61.6	61.6	65.3	36.4	36.2	32.7	33.0	36.6	18.3	18.7	13.4	14.2	19.4
	-0.2	68.6	68.8	66.2	67.2	65.4	40.1	39.8	34.7	36.0	36.1	17.5	17.3	12.9	13.7	19.6
4	-0.1	65.8	65.4	63.2	63.4	64.5	32.7	32.7	26.3	27.3	33.8	18.6	18.5	11.1	12.0	18.0
	-0.2	61.8	62.2	58.2	58.6	64.5	34.4	34.1	28.5	30.4	33.7	17.4	17.2	11.8	12.2	17.9
5	-0.1	63.2	62.4	58.6	59.6	64.3	28.6	28.5	22.8	24.0	33.4	14.5	14.0	8.7	9.6	18.3
	-0.2	60.0	60.2	56.4	57.0	64.5	34.8	34.7	28.5	30.1	33.3	14.9	13.8	9.0	9.4	18.4

Table 3.7: Size of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are autoregressive over time, improper prior used to sample precision parameters. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0.2	3.8	3.6	3.4	3.4	4.7	5.1	5.2	4.7	4.7	5.4	4.5	5.2	4.0	3.5	5.3
	0.3	4.6	4.6	4.0	3.8	4.8	3.2	3.2	2.9	2.7	5.3	5.3	5.2	5.0	4.5	5.2
	0.4	7.6	7.4	6.6	6.2	4.8	6.1	5.9	5.5	4.1	5.1	5.3	5.1	4.2	3.4	5.2
	0.5	5.0	5.0	4.4	4.2	4.8	4.8	4.7	4.4	3.8	5.1	4.4	3.7	3.3	2.1	5.2
3	0.2	4.2	4.2	4.4	4.2	4.9	4.9	5.2	4.7	4.7	5.2	5.2	5.4	4.9	5.0	5.2
	0.3	4.2	4.4	4.2	4.0	4.9	4.6	5.0	4.5	3.9	5.1	4.4	4.7	3.9	3.1	5.2
	0.4	4.4	4.4	4.4	3.8	4.8	4.4	4.0	4.0	3.7	5.1	4.7	4.1	4.3	3.7	5.3
	0.5	5.0	5.4	4.6	3.4	4.9	4.4	4.7	4.3	2.8	5.1	4.6	4.1	4.3	2.2	5.1
4	0.2	3.6	3.4	3.8	4.0	5.4	5.4	5.4	5.4	5.2	5.0	5.4	5.4	4.9	4.8	5.0
	0.3	4.0	4.2	4.0	3.6	5.4	4.8	4.7	4.8	4.3	5.1	4.4	4.5	4.5	4.1	5.0
	0.4	3.4	3.6	3.4	3.0	5.1	4.0	3.8	3.9	3.3	5.2	5.5	5.5	5.6	3.7	4.9
	0.5	5.0	5.4	5.2	4.2	5.1	5.2	5.5	5.1	3.9	5.1	4.3	3.9	4.6	3.0	4.8
5	0.2	4.6	4.0	4.4	4.4	5.3	4.8	4.5	4.7	4.7	5.1	5.8	6.0	5.5	5.5	5.4
	0.3	4.6	5.2	4.8	4.2	5.3	4.3	4.1	4.2	4.0	4.9	5.4	5.5	4.8	4.5	5.3
	0.4	4.6	5.0	4.6	3.8	5.2	5.6	5.5	5.6	5.1	5.0	4.6	4.6	4.5	3.6	5.4
	0.5	4.4	5.4	4.2	3.2	5.3	4.1	4.5	4.1	3.3	5.1	5.5	4.8	5.4	3.5	5.2

Table 3.8: Size of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are autoregressive over time, Gamma (informative) prior used to sample the precision parameters of standardized data. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0.2	3.8	3.8	3.6	3.4	4.7	4.7	4.6	3.8	3.4	5.4	4.4	4.5	3.7	3.1	5.3
	0.3	4.0	4.0	3.6	2.8	4.8	4.5	4.4	3.8	3.4	5.3	4.6	4.5	3.7	3.1	5.2
	0.4	4.0	3.8	3.8	2.8	4.8	4.8	4.8	4.0	3.1	5.1	5.0	4.3	3.6	2.7	5.2
	0.5	3.8	3.8	3.4	2.8	4.8	4.8	4.2	4.1	2.7	5.1	4.7	4.4	3.8	1.9	5.2
3	0.2	5.2	4.8	4.8	5.0	4.9	5.2	4.5	4.0	3.6	5.2	4.7	4.4	3.7	3.7	5.2
	0.3	5.4	5.4	5.2	4.6	4.9	5.2	4.5	4.1	3.6	5.1	4.9	4.5	4.0	3.3	5.2
	0.4	5.4	4.8	5.2	4.2	4.8	4.9	4.3	4.4	2.7	5.1	4.6	4.2	4.0	2.9	5.3
	0.5	5.0	4.4	4.8	3.8	4.9	5.3	4.3	4.5	2.4	5.1	4.3	3.9	3.7	2.3	5.1
4	0.2	4.6	4.6	4.4	3.8	5.4	4.9	4.9	4.4	4.0	5.0	4.0	3.6	3.3	3.2	5.0
	0.3	4.8	4.8	4.6	3.4	5.4	4.6	4.5	4.0	3.6	5.1	4.0	3.3	3.5	2.5	5.0
	0.4	4.8	4.8	4.4	3.0	5.1	4.8	4.1	4.3	3.0	5.2	4.0	3.4	3.5	2.0	4.9
	0.5	4.2	5.0	4.2	3.0	5.1	4.2	3.5	4.0	2.3	5.1	4.0	3.4	3.0	1.4	4.8
5	0.2	6.0	6.0	5.8	5.6	5.3	4.1	4.3	3.6	3.1	5.1	5.6	4.8	4.8	3.9	5.4
	0.3	6.0	6.0	5.6	4.8	5.3	4.3	4.0	3.6	2.7	4.9	5.2	4.2	4.7	3.6	5.3
	0.4	5.2	5.8	5.0	4.4	5.2	4.3	4.2	3.6	2.4	5.0	4.8	3.4	3.8	2.3	5.4
	0.5	5.2	5.2	5.2	3.4	5.3	4.5	4.0	4.0	2.2	5.1	4.7	3.3	4.0	2.1	5.2

consistent with the biased size of Q_ρ and Q_z shown in Table 3.7 and 3.8.

3.3.4 Motivating example

In this section, we illustrate the proposed tests with real data. We analyzed a longitudinal periodontal study conducted by Kinney et al. (2011) and Ramseier et al. (2009). 79 subjects completed the 12-month study, with samples of serum-derived biomarkers (TNF-α, calprotectin, metalloproteinase (MMP)-8, MMP-9) and saliva-derived biomarkers (IL-1β, MMP-9, MMP-9, OPG) and periodontal plaque biofilm pathogens (*P.gingivalis T.forsythia*, *T.denticola*) examined at baseline (Day 0), 6 months and 12 months. Ramseier et al. (2009) found that the concentration levels of salivary biomarkers MMP-8, MMP-9 and calprotectin were associated with stages of periodontal disease, and can be used as good predictors of periodontitis because of large odds ratios; moreover, all the plaque biofilm pathogens listed above demonstrate even higher diagnostic ability than biomarkers.

Considering the ability of both biomarkers and pathogens as periodontol disease predictors, it is natural to expect some degrees of correlation between them. We would now like to assess whether there is a constant correlation between certain combination of biomarker and pathogen. We first add 1 to all the measured values and take a log-transformation. Shown in Table 3.11 and 3.12 are the sample serial correlations for each pair of biomarker and pathogen after the transformation described above. For serum biomarker data, most sample serial correlations range between -0.2 and 0.2. The minimum serial correlation ranges from -0.21 to 0. The maximum-minimum correlation difference is between 0.09 (MMP-8 and *T.denticola*) and 0.33 (MMP-9 and *P.gingivalis*). Most pairs have both positive and negative correlations at the three time points. We also calculated the sample cross-correlations and most of them

Table 3.9: Power of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are autoregressive over time, improper prior used to sample precision parameters. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0.2	80.2	80.2	79.0	78.4	81.1	53.9	53.7	51.8	49.7	50.7	26.8	26.6	24.2	22.6	28.7
	0.3	90.8	90.6	90.0	88.8	87.8	60.9	60.8	58.7	53.8	58.7	30.7	30.2	28.2	24.5	33.0
	0.4	93.4	93.4	93.2	92.4	95.5	71.7	71.0	69.4	64.0	71.8	38.6	36.7	36.0	25.9	42.3
	0.5	100.0	100.0	100.0	99.8	99.8	91.2	91.1	90.1	84.7	90.4	63.0	61.7	58.9	42.7	61.0
3	0.2	65.8	65.6	65.4	64.2	65.4	34.0	34.0	33.4	31.2	37.0	18.0	17.7	17.4	15.5	20.0
	0.3	75.6	76.2	75.0	73.8	73.9	46.6	46.5	45.9	40.6	43.6	23.9	22.8	22.9	17.7	22.9
	0.4	83.8	84.8	83.2	81.0	86.1	55.5	54.7	55.0	47.0	55.0	28.9	27.1	27.9	19.3	29.2
	0.5	97.0	97.0	97.0	95.2	97.1	76.0	75.3	75.3	65.6	75.3	43.5	41.4	41.4	25.1	44.4
4	0.2	61.4	61.4	61.2	58.2	61.5	34.7	34.6	34.8	32.2	32.6	16.1	15.2	15.8	13.1	17.5
	0.3	69.8	69.8	69.6	67.2	70.1	39.9	40.1	40.1	34.9	38.9	17.8	16.2	17.5	13.6	20.3
	0.4	81.4	81.8	81.2	78.4	82.0	47.2	47.8	46.9	40.1	50.2	25.9	25.4	25.7	17.2	26.0
	0.5	96.2	96.4	96.2	94.2	94.9	66.9	67.1	66.0	56.1	68.7	36.5	34.0	36.9	21.5	38.0
5	0.2	57.4	56.8	57.4	56.8	58.6	29.8	29.6	29.4	27.4	32.2	16.8	16.6	17.1	14.8	16.8
	0.3	69.4	70.4	69.2	67.6	68.2	37.4	35.5	37.2	32.1	37.3	17.3	14.9	17.1	12.6	19.3
	0.4	77.6	78.2	78.2	73.4	80.8	47.1	46.6	47.5	38.7	47.1	23.9	22.4	24.1	15.9	24.4
	0.5	93.8	94.0	93.8	90.2	93.6	65.3	63.7	65.1	55.0	66.0	35.6	32.9	35.8	21.7	34.6

Table 3.10: Power of posterior predictive tests using T_ρ and T_z, Q_ρ and Q_z when ρ_x, ρ_y and ρ_{xy} are autoregressive over time, Gamma (informative) prior used to sample the precision parameters of standardized data. Each value is the percentage of simulations out of 500 ($n = 100$) or 1,000 ($n = 50, 25$) in which the null hypothesis is rejected. T_ρ=posterior predictive test using data-based test statistic χ^2_ρ ; T_z=posterior predictive test using data-based test statistic χ^2_z ; Q_ρ=posterior predictive test using discrepancy statistic χ^2_ρ ; Q_z=posterior predictive test using discrepancy statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

		$n = 100$					$n = 50$					$n = 25$				
m	ρ_{min}	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z	T_z	T_ρ	Q_z	Q_ρ	χ^2_z
2	0.2	79.0	79.0	78.6	77.8	81.1	50.8	50.2	48.5	45.6	50.7	27.9	26.4	24.1	19.6	28.7
	0.3	90.2	90.2	89.6	87.8	87.8	60.8	60.3	57.4	50.2	58.7	30.2	28.9	26.9	21.3	33.0
	0.4	93.2	93.2	92.8	91.4	95.5	71.1	70.4	68.6	59.6	71.8	38.6	35.8	34.2	21.3	42.3
	0.5	100.0	100.0	100.0	99.8	99.8	88.6	88.3	87.2	80.5	90.4	61.8	59.0	57.9	33.8	61.0
3	0.2	66.8	67.0	66.6	61.8	65.4	34.4	33.9	33.0	29.0	37.0	19.9	18.7	18.2	13.7	20.0
	0.3	76.2	75.4	74.8	70.4	73.9	46.2	45.1	44.9	36.0	43.6	23.8	20.9	21.6	14.2	22.9
	0.4	83.4	83.4	82.8	78.2	86.1	55.3	53.7	53.4	40.5	55.0	29.3	25.5	26.3	14.2	29.2
	0.5	96.8	96.8	96.6	94.2	97.1	75.5	75.2	73.5	55.0	75.3	42.7	38.5	39.1	17.5	44.4
4	0.2	61.4	61.2	60.6	56.4	61.5	35.5	33.8	33.3	28.5	32.6	15.7	14.0	14.0	10.6	17.5
	0.3	70.0	69.2	69.0	64.2	70.1	39.6	39.1	38.4	29.9	38.9	18.0	14.7	16.4	10.0	20.3
	0.4	81.4	81.4	81.0	74.6	82.0	47.1	46.6	46.0	32.8	50.2	26.4	21.8	24.5	13.2	26.0
	0.5	95.8	96.4	95.6	91.8	94.9	66.9	66.6	64.5	46.7	68.7	36.3	31.3	33.8	14.4	38.0
5	0.2	58.2	56.6	56.2	53.2	58.6	33.3	32.5	32.2	25.8	32.2	15.8	13.5	14.9	11.2	16.8
	0.3	69.6	69.4	69.2	64.8	68.2	36.9	35.2	35.1	26.8	37.3	16.8	13.4	15.0	8.8	19.3
	0.4	77.8	77.2	77.2	70.4	80.8	46.2	44.3	45.5	31.6	47.1	22.3	18.4	20.4	9.5	24.4
	0.5	93.8	93.8	93.4	89.2	93.6	68.8	66.0	67.0	45.3	66.0	33.9	28.5	32.0	13.8	34.6

are close to 0 so that they can be considered as equal. For salivary biomarker data, most sample serial correlations range between -0.1 and 0.5. The maximum-minimum correlation difference is between 0.06 (IL-1β and *T.forsythia*) and 0.66 (MMP-8 and *T.forsythia*).

To test the hypothesis whether the correlation between a biomarker and a pathogen at each time point is equal, posterior preditive tests, T_ρ, T_z, Q_z and Wald Z-test (in Chapter 2) were performed. Since Q_ρ does not achieve nominal size, the result is not presented here. Improper prior for the original data (non-standardized) was used to sample variance parameters. Table 3.11 and Table 3.12 summarize the posterior predictive p-values for serum and salivary data respectively. Three tests gave comparable results and most pairs have homogenous serial correlations based on our tests. As shown in Table 3.11, using 0.1 as a critical value, heterogeneity exists in pairs between serum TNF-α and *T.forsythia* (max-min difference is 0.22), serum MMP-9 and *P.gingivalis* (max-min difference is 0.33), serum MMP-8 and *P.gingivalis* (max-min difference is 0.29). Consistent with simulation studies with setting I (Table 3.3 and 3.5), regarding conservativeness, $Q_z > T_z \approx T_\rho > \chi^2_z$. However the same conclusion is drawn from each test. As shown in Table 3.12, heterogeneity exists in pairs between salivary MMP-8 and *T.forsythia* (max-min difference is 0.66), salivary MMP-8 and *P.gingivalis* (max-min difference is 0.37), salivary MMP-9 and *T.forsythia* (max-min difference is 0.45), salivary OPG and *T.denticola* (max-min difference is 0.44). Consistent with simulation studies with setting I (Table 3.3 and 3.5), regarding conservativeness, $Q_z > T_z \approx T_\rho > \chi^2_z$. However the same conclusion is drawn from each test. We conclude that the serial correlation in the pairs mentioned above changes over the twelve months.

Table 3.11: Sample serial correlations between combinations of serum biomarkers and pathogens, and testing equality of serial correlations at 0, 6,12 months. T_ρ=posterior predictive test using test statistic χ^2_ρ; T_z=posterior predictive test using test statistic χ^2_z; Q_z=posterior predictive test using discrepancy statistic χ^2_z

Pathogen	Biomarker	Sample serial correlation			p-value		
		0	6	12	T_z	T_ρ	Q_z
P.gingivalis	TNF-α	-0.17	0.07	-0.07	0.110	0.111	0.153
	Calprotectin	-0.18	-0.01	0.11	0.149	0.148	0.155
	MMP-8	-0.01	0.28	0.04	0.060	0.059	0.077
	MMP-9	0.12	0.19	-0.14	0.025	0.026	0.036
T.forsythia	TNF-α	-0.19	0.07	-0.19	0.006	0.007	0.010
	Calprotectin	-0.21	-0.02	0.05	0.162	0.160	0.166
	MMP-8	0	0.22	0.13	0.181	0.182	0.189
	MMP-9	0.13	0.19	-0.03	0.137	0.139	0.149
T.denticola	TNF-α	-0.13	-0.04	0.03	0.377	0.377	0.404
	Calprotectin	-0.17	0.01	0.13	0.106	0.106	0.112
	MMP-8	-0.01	0.08	0.00	0.788	0.788	0.796
	MMP-9	0.17	0.02	-0.02	0.275	0.273	0.299

Table 3.12: Sample serial correlations between combinations of salivary biomarkers and pathogens, and testing equality of serial correlations at 0, 6,12 months. T_ρ=posterior predictive test using test statistic χ^2_ρ; T_z=posterior predictive test using test statistic χ^2_z; Q_z=posterior predictive test using discrepancy statistic χ^2_z

Pathogen	Biomarker	Sample serial correlation			p-value		
		0	6	12	T_z	T_ρ	Q_z
P.gingivalis	IL-1β	0.21	0.04	0.07	0.668	0.665	0.675
	MMP-8	0.38	0.40	0.03	0.094	0.105	0.097
	MMP-9	0.30	0.20	-0.03	0.319	0.324	0.335
	OPG	0.26	0.02	-0.08	0.257	0.252	0.271
T.forsythia	IL-1β	0.23	0.26	0.20	0.946	0.946	0.945
	MMP-8	0.35	0.56	-0.10	0.001	0.000	0.000
	MMP-9	0.30	0.13	-0.15	0.098	0.099	0.108
	OPG	0.27	-0.01	0.16	0.360	0.360	0.358
T.denticola	IL-1β	0.36	-0.03	0.16	0.134	0.133	0.162
	MMP-8	0.41	0.41	0.20	0.348	0.365	0.377
	MMP-9	0.35	0.41	0.07	0.213	0.224	0.221
	OPG	0.38	0.21	-0.06	0.090	0.093	0.109

3.4 Conclusions and Discussion

In this chapter, we described two Bayesian approaches to perform tests of equality of correlation coefficients for longitudinal studies. In Bayesian confidence interval method, we evaluated the tests based on two types of intervals: highest posterior density region (HPD) and quantile based credible interval (QB). Both of them have a nominal size, while HPD tend to be more conservative and QB tend to be more liberal. When there are only two time points, HPD is suggested to use.

We borrowed the classical Wald χ^2_ρ and χ^2_z statistics to construct posterior predictive p-values. The empirical size and power of our proposed tests in a variety of settings motivated by the data collected in our motivating study were collected. Conclusions are: (1) Posterior predictive tests T_z and T_ρ are conservative compared to Wald Z-test, yet after considering monte carlo errors the three tests have comparable size. (2) Posterior predictive tests T_z and T_ρ have similar power compared to Wald χ^2_z in medium sample size (n=100, 50), and χ^2_z is superior in small sample size (n=25). The size and power of T_z and T_ρ are not significantly impacted by sampling methods (i.e., prior setting and whether the data is standardized). (3) Posterior predictive tests Q_z and Q_ρ are conservative compared with T_z and T_ρ. Q_z and Q_ρ depend largely on the parameters sampled from standardized data. (4) Our assumption of equal cross correlation (ρ_x, ρ_y and ρ_{xy}) in $\mathbf{\Sigma}$ is generally robust to data without equal cross correlation, yet the posterior predictive approach is still more sensitive to mis-specification than Wald test. (5) T_z and T_ρ are preferable than Q_z and Q_ρ in our setting, since they reach a nominal size when model is not mis-specified, and are robust to model mis-specification and different prior choices.

Although in our particular problem, Bayesian posterior predictive method does not seem to provide a better power than Wald test, it offers an alternative way of

examining the hypothesis testing problem. We introduce discrepancy variables into our test, and this allows us to incorporate unknown nuisance parameters into test statistics, rather than plugging an estimator (e.g., maximum-likelihood estimator) obtained from the original data for the parameters.

CHAPTER IV

Pooled Correlation Coefficients for Longitudinally Measured Biomarkers

4.1 Introduction

In order to measure the association between two categorical variables X and Y, both of which are measured on n_j subjects in stratum $j = 1, 2, \ldots m$, a 2×2 table is made for each stratum, in which a_j and d_j are the number of subjects in strata j, who have $X = Y = 1$ and $X = Y = 0$, respectively, b_j is the number of subjects with $X = 0$ and $Y = 1$, and c_j is the number of subjects with $X = 1$ and $Y = 0$. Let ψ_j denote the population odds ratio in strata j, whose estimate is $\hat{\psi}_j = R_j/S_j = a_j d_j/(b_j c_j)$.

A test of homogeneity of the stratum-specific odds ratios attributed to Breslow and Day (1980) is usually applied to determine if there is evidence that the stratum-specific odds ratios are equal. If the test fails to provide evidence of heterogeneity, then it is often assumed the association of X and Y is constant among all strata, necessitating an estimator for the pooled odds ratio for X and Y.

The Mantel-Haenszel (Mantel and Haenszel, 1959) estimator for ψ is

$$\widehat{\psi}_{MH} = \frac{R}{S} = \frac{\sum_{j=1}^{m} R_j}{\sum_{j=1}^{m} S_j} = \frac{\sum_{j=1}^{m} a_j d_j / n_j}{\sum_{j=1}^{m} b_j c_j / n_j}$$

Historically, several variance estimators have been proposed for $\widehat{\psi}_{MH}$, which can be divided into two classes. In the first class, the number of tables remains fixed while the counts in the individual cells increase and have no bound ($n \to \infty$). One estimator based on this class was proposed by Hauck (1979):

$$V_H = \frac{\hat{\psi}_{MH}^2 (\sum_{j=1}^{m} S_j^2 / w_j)}{(\sum_{j=1}^{m} S_j)^2} \tag{4.1}$$

in which $w_j^{-1} = (1/a_j + 1/b_j + 1/c_j + 1/d_j)$. This estimator can be easily derived from the mean and variance of the individual $\hat{\psi}_j$'s, which are MLE and consistent estimators of ψ_j's, by re-writing $\hat{\psi}_{MH}$ as a weighted average of the $\hat{\psi}_j$'s (Hauck, 1979; Silcocks, 2005), i.e.,

$$\widehat{\psi}_{MH} = \frac{\sum_{j=1}^{m} R_j}{\sum_{j=1}^{m} S_j} = \sum_{j=1}^{m} \frac{R_j}{S_j} \frac{S_j}{\sum_{j=1}^{m} S_j} = \sum_{j=1}^{m} \hat{\psi}_j \frac{S_j}{\sum_{j=1}^{m} S_j}$$

In the second class, the number of tables increases while the cell sizes are bounded ($m \to \infty$). Many variance estimators in this class exist (Robins et al., 1986; Breslow and Liang, 1982) and are derived from the expression

$$\sqrt{m}(\hat{\psi}_{MH} - \psi) = \frac{\sqrt{m}(R - \psi S)/m}{S/m} = \frac{\sqrt{m} \sum_{j=1}^{m} (R_j - \psi S_j)/m}{\sum_{j=1}^{m} S_j / m} \tag{4.2}$$

Since $E(Rj - \psi S_j) = 0$, the Central Limit Theorem states that the numerator of

Equation (4.2) is asymptotically normal with mean zero and variance

$$lim_{m\to\infty} Var(\sum_{j=1}^{m}(R_j - \psi S_j))/m$$

The denominator of Equation (4.2) converges to its mean $\sum_{j=1}^{m} E(S_j)/m$, as $m \to \infty$. Applying Slutsky's theorem such that

$$lim_{m\to\infty} mVar(\hat{\psi}_{MH}) = \frac{lim_{m\to\infty} Var(\sum_{j=1}^{m}(R_j - \psi S_j))/m}{[lim_{m\to\infty} \sum_{j=1}^{m} E(S_j)/m]^2} \tag{4.3}$$

Different versions of the numerator and the denominator of Equation (4.3) have been suggested. For example, Breslow (1981) proposed an empirical Mantel-Haenszel variance, that is

$$V_B = \frac{\sum_{j=1}^{m}(R_j - \hat{\psi}_{MH} S_j)^2}{(\sum_{j=1}^{m} S_j)^2} \tag{4.4}$$

Robins et al. (1986) proposed another estimator with the numerator for (4.4), or $Var(\sum_{j=1}^{m}(R_j - \psi S_j))$ replaced by an unbiased estimator. With some algebra, their MH variance estimator became:

$$V_U = [\frac{\sum_{j=1}^{m} P_j R_j}{2R^2} + \frac{\sum_{j=1}^{m}(P_j S_j + Q_j R_j)}{2RS} + \frac{\sum_{j=1}^{m} Q_j S_j}{2S^2}](\hat{\psi}_{MH})^2 \tag{4.5}$$

where $P_j = (a_j + d_j)/n_j$, $Q_j = (b_j + c_j)/n_j$.

Lastly, there are some "hybrid" versions of Mantel-Haenszel variance estimators that fit both models above. An example is that proposed by Breslow and Liang

(1982):

$$V_C = \frac{nV_H + m^2 V_B}{n + m^2}$$

where $n = \sum_{j=1}^{m} n_j$. This estimator incorporate the variance estimators in Equations (4.1) and (4.4), and allows either estimator to dominate according to total counts in all cells and number of strata.

In the previous two chapters, we discussed two ways to test if the correlation between each biomarker and each pathogen is homogeneous over time. After testing the homogeneity hypothesis, and lack of significance is found, a decision can be made to pool the various correlation coefficients into a single time-invariant value that quantifies the correlation between a biomarker and a pathogen. In this study, we are interested in finding a pooled serial correlation estimator for longitudinal data. Our proposed pooled correlation coefficient estimator, r_{MH} is based on Mantel-Haenszel methods. In Section 4.2 we will derive two variance estimators of r_{MH} analogous to the methods just described ($m \to \infty$; $n \to \infty$). We use Monte Carlo simulations to evaluate the bias of these estimators in Section 4.3.

4.2 Methods

4.2.1 Notation and Definition of Pooled Correlation Coefficient Estimate

We have n subjects who are each examined sequentially at times $t_1 < t_2, \ldots, < t_m$. Let X_{ij} and Y_{ij}, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$, denote the respective values of biomarker X and pathogen Y collected from subject i at time t_j. Marginally, we assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$, where μ_{xj} and μ_{yj} are parameters quantifying the means of X_{ij} and Y_{ij}, respectively. The elements of $\boldsymbol{X_i} = (X_{i1}, X_{i2}, \cdots, X_{im})$, are assumed to be exchangeably correlated with each other with

correlation ρ_x, and the elements of $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2}, \cdots, Y_{im})$ are exchangeably correlated with each other with correlation ρ_y. We also assume a common cross-correlation, ρ_{xy} between X_{ij} and Y_{ik}, for $j \neq k$. The time-specific correlation of X_{ij} and Y_{ij} is defined to be $\rho_j = Corr(X_{ij}, Y_{ij})$ and are the parameters of interest. Explicitly, let $\boldsymbol{D}_i = \{X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, \cdots, X_{im}, Y_{im}\}^t$ denote the $(2m \times 1)$ longitudinal vector of pairs of biomarker and pathogen for subject i, and assume $\boldsymbol{D}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ in which

$$\boldsymbol{\mu} = \{\mu_{x_1}, \mu_{y_1}, \mu_{x_2}, \mu_{y_2}, \cdots, \mu_{x_m}, \mu_{y_m}\} \tag{4.6}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_1\sigma_1\tau_1 & \rho_x\sigma_1\sigma_2 & \rho_{xy}\sigma_1\tau_2 & \cdots & \rho_x\sigma_1\sigma_m & \rho_{xy}\sigma_1\tau_m \\ \rho_1\sigma_1\tau_1 & \tau_1^2 & \rho_{xy}\sigma_2\tau_1 & \rho_y\tau_1\tau_2 & \cdots & \rho_{xy}\sigma_m\tau_1 & \rho_y\tau_1\tau_m \\ \rho_x\sigma_1\sigma_2 & \rho_{xy}\sigma_2\tau_1 & \sigma_2^2 & \rho_2\sigma_2\tau_2 & \cdots & \rho_x\sigma_2\sigma_m & \rho_{xy}\sigma_2\tau_m \\ \rho_{xy}\sigma_1\tau_2 & \rho_y\tau_1\tau_2 & \rho_2\sigma_2\tau_2 & \tau_2^2 & \cdots & \rho_{xy}\sigma_m\tau_2 & \rho_y\tau_2\tau_m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_x\sigma_1\sigma_m & \rho_{xy}\sigma_m\tau_1 & \rho_x\sigma_2\sigma_m & \rho_{xy}\sigma_m\tau_2 & \cdots & \sigma_m^2 & \rho_m\sigma_m\tau_m \\ \rho_{xy}\sigma_1\tau_m & \rho_y\tau_1\tau_m & \rho_{xy}\sigma_2\tau_m & \rho_y\tau_2\tau_m & \cdots & \rho_m\sigma_m\tau_m & \tau_m^2 \end{pmatrix} \tag{4.7}$$

same as Equations 2.1 and 2.2.

The correlation of X_{ij} and Y_{ij} at time t_j is defined to be

$$\rho_j = \frac{\sum_{i=1}^{n}(X_{ij} - \mu_X)(Y_{ij} - \mu_Y)}{\sqrt{\sum_{i=1}^{n}(X_{ij} - \mu_X)^2 \sum_{i=1}^{n}(Y_{ij} - \mu_Y)^2}}$$

whose corresponding sample estimate is

$$\begin{aligned}\hat{\rho}_j &= \frac{\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)(Y_{ij}-\overline{Y}_j)}{\sqrt{\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2\sum_{i=1}^{n}(Y_{ij}-\overline{Y}_j)^2}} \\ &= \frac{\sum_{i=1}^{n}\widetilde{X}_{ij}\widetilde{Y}_{ij}}{\sqrt{\sum_{i=1}^{n}\widetilde{X}_{ij}^2\sum_{i=1}^{n}\widetilde{Y}_{ij}^2}} = \frac{C_j}{B_j}\end{aligned}$$

After testing for homogeneity and if we fail to reject H_0, we pool the values of $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m$ into a single time-invariant estimate of the correlation of X and Y by applying the idea of Mantel and Haenszel (1959) to a setting with continuous data. Specifically, we define

$$r_{MH} = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n}\widetilde{X}_{ij}\widetilde{Y}_{ij}}{\sum_{j=1}^{m}\sqrt{\sum_{i=1}^{n}\widetilde{X}_{ij}^2\sum_{i=1}^{n}\widetilde{Y}_{ij}^2}} = \frac{\sum_{j=1}^{m}C_j}{\sum_{j=1}^{m}B_j} = \frac{C}{B} \tag{4.8}$$

Although the computation of r_{MH} is straightforward, a variance estimate for r_{MH} is not. We now derive two asymptotic variance estimates for r_{MH}, one in which the number of time points, m, goes to infinity, and one in which the number of subjects, n, goes to infinity.

4.2.2 Asymptotic variance as $n \to \infty$

Similar to the work of Hauck (1979) and Silcocks (2005) for $Var(\hat{\psi}_{MH})$, r_{MH} can be rewritten as a weighted average of each $\hat{\rho}_j$:

$$\begin{aligned}r_{MH} &= \sum_{j=1}^{m}\frac{\sum_{i=1}^{n}\widetilde{X}_{ij}\widetilde{Y}_{ij}}{\sqrt{\sum_{i=1}^{n}\widetilde{X}_{ij}^2\sum_{i=1}^{n}\widetilde{Y}_{ij}^2}}\cdot\frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\widetilde{X}_{ij}^2\cdot\frac{1}{n-1}\sum_{i=1}^{n}\widetilde{Y}_{ij}^2}}{\sum_{j=1}^{m}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\widetilde{X}_{ij}^2\cdot\frac{1}{n-1}\sum_{i=1}^{n}\widetilde{Y}_{ij}^2}} \\ &= \sum_{j=1}^{m}\hat{\rho}_j\omega_j/\sum_{j=1}^{m}\omega_j\end{aligned} \tag{4.9}$$

where $\omega_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ij}^2 . \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_{ij}^2} = \hat{\sigma}_j\hat{\tau}_j$.

Notice that $(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_m)$ are MLEs (and therefore consistent estimators) of the respective serial correlations $\rho_1, \rho_2, \ldots, \rho_m$, and $\hat{\sigma}_j^2$ and $\hat{\tau}_j^2$ are consistent estimators of σ_j^2 and τ_j^2. As a result, r_{MH} converges in probability to $\rho_{MH} = \sum_{j=1}^{m}\rho_j\nu_j / \sum_{j=1}^{m}\nu_j$, where $\nu_j = \sigma_j\tau_j$.

Recall that in Chapter 2 we derived the asymptotic joint distribution:

$$\sqrt{n}\{(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_m)^T - (\rho_1, \rho_2, \cdots, \rho_m)^T\} \rightarrow_d MVN(\mathbf{0}, \boldsymbol{\Sigma_\rho})$$

Yu and Dunn (1982) suggested that the value n be replaced by the value $n-3$ in order to improve the approximation in small samples. $\boldsymbol{\Sigma_\rho}$ is an $m \times m$ matrix with diagonal element θ_{jj} and off-diagonal element θ_{jk} equal to

$$\theta_{jj} = (1-\rho_j^2)^2 \tag{4.10}$$

$$\theta_{jk} = \left\{\frac{1}{2}\rho_j\rho_k(\rho_x^2+\rho_y^2) + \rho_{xy}^2(1+\rho_j\rho_k) + \rho_x\rho_y - \rho_{xy}(\rho_j+\rho_k)(\rho_x+\rho_y)\right\} \tag{4.11}$$

By Slutsky's Theorem,

$$\begin{aligned}\sqrt{n-3}(r_{MH} - \rho_{MH}) &= \sqrt{n-3}\{\boldsymbol{\omega}(\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_m)^T - \boldsymbol{\nu}(\rho_1, \rho_2, \cdots, \rho_m)^T\} \\ &\rightarrow_d MVN(\mathbf{0}, \boldsymbol{\nu}\boldsymbol{\Sigma_\rho}\boldsymbol{\nu}^T)\end{aligned} \tag{4.12}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_m)/\sum_{j=1}^{m}(\omega_j)$, and $\boldsymbol{\nu} = (\nu_1, \nu_2, ..., \nu_m)/\sum_{j=1}^{m}(\nu_j)$.

When there is a common serial correlation ρ, r_{MH} is a consistent estimator of ρ.

Therefore, r_{MH} has approximate variance

$$\begin{aligned} & Var_n(MH) = \boldsymbol{\omega}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\rho}}\boldsymbol{\omega}^T/(n-3) \\ = \; & \{\frac{\sum_{j\neq k}\omega_j\omega_k\left\{\frac{1}{2}r_{MH}^2(\hat{\rho}_x^2+\hat{\rho}_y^2)+\hat{\rho}_{xy}^2(1+r_{MH}^2)+\hat{\rho}_x\hat{\rho}_y-2\hat{\rho}_{xy}r_{MH}(\hat{\rho}_x+\hat{\rho}_y)\right\}}{(\Sigma_j\omega_j)^2} \\ & +\frac{\sum_{j=1}^m\omega_j^2(1-r_{MH}^2)^2}{(\sum_{j=1}^m\omega_j)^2}\}/(n-3) \end{aligned} \tag{4.13}$$

4.2.3 Asymptotic variance as $m \to \infty$

Let ρ denote the true common serial correlation under null hypothesis. r_{MH} can be expressed as

$$\sqrt{m}(r_{MH}-\rho) \;=\; \frac{\sqrt{m}(C/m-\rho B/m)}{B/m} \tag{4.14}$$

In order to derive the asymptotic variance for r_{MH}, we will show that $\sqrt{m}(C/m-\rho B/m)$ has a limiting normal distribution and that B/m converges in probability to a constant so that the variance of r_{MH} is simply the variance of the limiting variance of $\sqrt{m}(C/m-\rho B/m)$ scaled by the square of the limiting constant for B/m.

We now derive the limiting value for B/m. If we define $U_j=\sum_{i=1}^n\widetilde{X}_{ij}^2/\sigma_j$ and $V_j=\sum_{i=1}^n\widetilde{Y}_{ij}^2/\tau_j$, B can be expressed as $\sum_{j=1}^m\sigma_j\tau_j\sqrt{U_jV_j}$. Note that U_j and V_j follow χ^2 with $(n-1)$ degrees of freedom.

Let $W_j=\sqrt{U_jV_j}$, the square root of the product of two χ^2 random variables. Then B can be written as $\sum_{j=1}^m\sigma_j\tau_jW_j$. A second-order Taylor expension around $E(U_j)=n-1$ and $E(V_j)=n-1$ gives us an approximation for $E(W_j)$:

$$
\begin{aligned}
W_j &\approx \sqrt{\mu_U\mu_V} + \frac{U_j-\mu_U}{2}\sqrt{\frac{\mu_V}{\mu_U}} + \frac{V_j-\mu_V}{2}\sqrt{\frac{\mu_U}{\mu_V}} \\
&\quad -\frac{(U_j-\mu_U)^2}{8}\sqrt{\frac{\mu_V}{\mu_U^3}} + \frac{(U_j-\mu_U)(V_j-\mu_V)}{4\sqrt{\mu_U\mu_V}} - \frac{(V_j-\mu_V)^2}{8}\sqrt{\frac{\mu_U}{\mu_V^3}} \\
&= \sqrt{(n-1)(n-1)} + \frac{U_j-(n-1)}{2}\sqrt{\frac{n-1}{n-1}} + \frac{V_j-(n-1)}{2}\sqrt{\frac{n-1}{n-1}} \\
&\quad -\frac{(U_j-(n-1))^2}{8}\sqrt{\frac{n-1}{(n-1)^3}} + \frac{(U_j-(n-1))(V_j-(n-1))}{4\sqrt{(n-1)^2}} \\
&\quad -\frac{(V_j-(n-1))^2}{8}\sqrt{\frac{n-1}{(n-1)^3}} \\
&= \frac{U_j+V_j}{2} - \frac{(U_j-(n-1))^2+(V_j-(n-1))^2}{8(n-1)} + \frac{(U_j-(n-1))(V_j-(n-1))}{4(n-1)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E(W_j) &\approx n-1-\frac{Var(U_j)+Var(V_j)}{8(n-1)} + \frac{Cov(U_j,V_j)}{4(n-1)} \\
&= n-\frac{3}{2}+\frac{Cov(U_j,V_j)}{4(n-1)}
\end{aligned}
$$

where

$$
\begin{aligned}
Cov(U_j,V_j) &= Cov(\sum_{i=1}^n \widetilde{X}_{ij}^2, \sum_{i=1}^n \widetilde{Y}_{ij}^2)/\sigma_j^2\tau_j^2 \\
&= \sum_{i=1}^n Cov(\widetilde{X}_{ij}^2, \widetilde{Y}_{ij}^2)/\sigma_j^2\tau_j^2 + \sum_{j\neq j'}^n Cov(\widetilde{X}_{ij}^2, \widetilde{Y}_{i'j}^2)/\sigma_j^2\tau_j^2 \\
&= nCov(\widetilde{X}_{ij}^2, \widetilde{Y}_{ij}^2)/\sigma_j^2\tau_j^2 + n(n-1)Cov(\widetilde{X}_{ij}^2, \widetilde{Y}_{i'j}^2)/\sigma_j^2\tau_j^2 \\
&= \{nE(\widetilde{X}_{ij}^2\widetilde{Y}_{ij}^2) + n(n-1)E(\widetilde{X}_{ij}^2\widetilde{Y}_{i'j}^2) - n^2E(\widetilde{X}_{ij}^2)E(\widetilde{Y}_{ij}^2)\}/\sigma_j^2\tau_j^2
\end{aligned}
$$

After some manipulations, we find that $Cov(U_j,V_j) = 2(n-1)\rho_j^2$.

so that

$$E(W_j) \approx n - \frac{3}{2} + \frac{\rho^2}{2} \approx n - a \tag{4.15}$$

where a is a value ranging between 1 and 3/2. We also derived the third order term of the Taylor expansion, and found it contained a negligible amount (approximately $1/4 \sum_{j=1}^{m} \sigma_j \tau_j$). Although higher order terms in the Taylor series expansion are not completely ignorable, their magnitude is small enough to allow us to ignore them in our approximation. Therefore, based on a second-order Taylor expansion, as m gets large B/m gets close to $\sum_{j=1}^{m} (n - 3/2 + \rho^2/2)\sigma_j \tau_j / m$.

To compute the limiting distribution for $\sqrt{m}(C/m - \rho B/m)$, which appears in the numerator of Equation (4.14), we need to derive $Var(C - \rho B) = Var(C) + \rho^2 Var(B) - 2\rho Cov(C, B)$.

In the setting of a pooled odds ratio, one crucial assumption for deriving the limiting distribution for the pooled odds ratio as $m \to \infty$ was the data for each stratum (each time point in our setting) was independent. However, due to a lack of independence of data from different time points, deriving both $Var(B)$ and $Cov(B, C)$ has proved impossible after a number of months of various attempts to approximate the joint distribution of B and C. As a result, we do not have an explicit asymptotic variance estimate for the pooled correlation coefficient estimate. Nonetheless, we do present our derivation for $Var(C)$ and the form the asymptotic variance would take if and when values for $Var(B)$ and $Cov(B, C)$ are found.

To derive $Var(C)$, we need to derive $E(C_j)$, $Var(C_j)$, and $Cov(C_j, C_{j'})$. To derive

$E(C_j)$, we need to first derive $Cov(\widetilde{X}_{ij}, \widetilde{Y}_{ij})$.

$$\begin{aligned} Cov(\widetilde{X}_{ij}, \widetilde{Y}_{ij}) &= Cov(X_{ij} - \bar{X}_j, Y_{ij} - \bar{Y}_j) \\ &= Cov(X_{ij}, Y_{ij}) + Cov(\bar{X}_j, \bar{Y}_j) - Cov(X_{ij}, \bar{Y}_j) - Cov(\bar{X}_j, Y_{ij}) \\ &= \rho_j \sigma_j \tau_j + \frac{1}{n^2} \sum_{i=1}^{n} Cov(X_{ij}, Y_{ij}) - Cov(X_{ij}, \bar{Y}_j) - Cov(\bar{X}_j, Y_{ij}) \\ &= \rho_j \sigma_j \tau_j + \frac{n}{n^2} \rho_j \sigma_j \tau_j - \frac{1}{n} Cov(X_{ij}, Y_{ij}) - \frac{1}{n} Cov(X_{ij}, Y_{ij}) \\ &= \frac{n-1}{n} \rho_j \sigma_j \tau_j \end{aligned} \tag{4.16}$$

Based on the equation above,

$$\begin{aligned} E(C_j) &= \sum_{i=1}^{n} E(\widetilde{X}_{ij} \widetilde{Y}_{ij}) = nE(\widetilde{X}_{ij} \widetilde{Y}_{ij}) \\ &= n\{Cov(\widetilde{X}_{ij}, \widetilde{Y}_{ij}) + E(\widetilde{X}_{ij}) E(\widetilde{X}_{ij})\} \\ &= (n-1) \rho_j \sigma_j \tau_j \end{aligned} \tag{4.17}$$

$$\begin{aligned} Var(C_j) &= Var(\sum_{i=1}^{n} \widetilde{X}_{ij} \widetilde{Y}_{ij}) \\ &= \sum_{i=1}^{n} Var(\widetilde{X}_{ij} \widetilde{Y}_{ij}) + \sum_{i \neq i'} Cov(\widetilde{X}_{ij} \widetilde{Y}_{ij}, \widetilde{X}_{i'j} \widetilde{Y}_{i'j}) \\ &= nVar(\widetilde{X}_{ij} \widetilde{Y}_{ij}) + n(n-1) Cov(\widetilde{X}_{ij} \widetilde{Y}_{ij}, \widetilde{X}_{i'j} \widetilde{Y}_{i'j}) \\ &= n\{E(\widetilde{X}_{ij}^2 \widetilde{Y}_{ij}^2) - E^2(\widetilde{X}_{ij} \widetilde{Y}_{ij})\} + n(n-1)\{E(\widetilde{X}_{ij} \widetilde{Y}_{ij} \widetilde{X}_{i'j} \widetilde{Y}_{i'j}) \\ &\quad - E(\widetilde{X}_{ij} \widetilde{Y}_{ij}) E(\widetilde{X}_{i'j} \widetilde{Y}_{i'j})\} \\ &= nE(\widetilde{X}_{ij}^2 \widetilde{Y}_{ij}^2) + n(n-1)\{E(\widetilde{X}_{ij} \widetilde{Y}_{ij} \widetilde{X}_{i'j} \widetilde{Y}_{i'j})\} - n^2 E^2(\widetilde{X}_{ij} \widetilde{Y}_{ij}) \end{aligned}$$

To derive $E(\widetilde{X}_{ij}^2 \widetilde{Y}_{ij}^2)$ and $E(\widetilde{X}_{ij} \widetilde{Y}_{ij} \widetilde{X}_{i'j} \widetilde{Y}_{i'j})$, we need to first derive $Var(\widetilde{X}_{ij})$, $Var(\widetilde{Y}_{ij})$,

$Cov(\widetilde{X}_{ij}, \widetilde{X}_{i'j})$, $Cov(\widetilde{Y}_{ij}, \widetilde{Y}_{i'j})$, $Cov(\widetilde{X}_{ij}, \widetilde{Y}_{i'j})$, and $Cov(\widetilde{Y}_{ij}, \widetilde{X}_{i'j})$. First, we have

$$\begin{aligned} Var(\widetilde{X}_{ij}) &= Var(X_{ij} - \bar{X}_j) \\ &= Var(X_{ij}) + Var(\bar{X}_j) - 2Cov(X_{ij}, \bar{X}_j) \\ &= \frac{n+1}{n}\sigma_j^2 - 2Cov(X_{ij}, \bar{X}_j) \\ &= \frac{n+1}{n}\sigma_j^2 - \frac{2}{n}Var(X_{ij}) \\ &= \frac{n-1}{n}\sigma_j^2 \end{aligned} \tag{4.18}$$

and similarly to Equation (4.18),

$$Var(\widetilde{Y}_{ij}) = \frac{n-1}{n}\tau_j^2 \tag{4.19}$$

We also find

$$\begin{aligned} Cov(\widetilde{X}_{ij}, \widetilde{X}_{i'j}) &= Cov(X_{ij} - \bar{X}_j, X_{i'j} - \bar{X}_j) \\ &= Cov(X_{ij}, X_{i'j}) - Cov(X_{ij}, \bar{X}_j) - Cov(X_{i'j}, \bar{X}_j) + Cov(\bar{X}_j, \bar{X}_j) \\ &= 0 - \frac{1}{n}Cov(X_{ij}, X_{ij}) - \frac{1}{n}Cov(X_{i'j}, X_{i'j}) + Var(\bar{X}_j) \\ &= -\frac{2}{n}\sigma_j^2 + \frac{1}{n}\sigma_j^2 \\ &= -\frac{1}{n}\sigma_j^2 \end{aligned} \tag{4.20}$$

and similarly,

$$Cov(\widetilde{Y}_{ij}, \widetilde{Y}_{i'j}) = -\frac{1}{n}\tau_j^2 \tag{4.21}$$

$$Cov(\widetilde{X}_{ij}, \widetilde{Y}_{i'j}) = Cov(\widetilde{Y}_{ij}, \widetilde{X}_{i'j}) = -\frac{1}{n}\rho_j\sigma_j\tau_j \tag{4.22}$$

Derivation of $E(\widetilde{X}_{ij}^2\widetilde{Y}_{ij}^2)$ and $E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j}\widetilde{Y}_{i'j})$ takes advantage of the fourth order moment information for the multivariate normal distribution (Isserlis, 1918). Note that $(\widetilde{X}_{ij}, \widetilde{Y}_{ij}, \widetilde{X}_{i'j}, \widetilde{Y}_{i'j})$ form a multivariate normal vector with covariance matrix based on Equations (4.18), (4.19), (4.16), (4.20), (4.21) and (4.22):

$$\mathbf{\Sigma^*} = \begin{pmatrix} \frac{n-1}{n}\sigma_j^2 & \frac{n-1}{n}\rho_j\sigma_j\tau_j & -\frac{1}{n}\sigma_j^2 & -\frac{1}{n}\rho_j\sigma_j\tau_j \\ \frac{n-1}{n}\rho_j\sigma_j\tau_j & \frac{n-1}{n}\tau_j^2 & -\frac{1}{n}\rho_j\sigma_j\tau_j & -\frac{1}{n}\tau_j^2 \\ -\frac{1}{n}\sigma_j^2 & -\frac{1}{n}\rho_j\sigma_j\tau_j & \frac{n-1}{n}\sigma_j^2 & \frac{n-1}{n}\rho_j\sigma_j\tau_j \\ -\frac{1}{n}\rho_j\sigma_j\tau_j & -\frac{1}{n}\tau_j^2 & \frac{n-1}{n}\rho_j\sigma_j\tau_j & \frac{n-1}{n}\tau_j^2 \end{pmatrix} \tag{4.23}$$

As a result, we have:

$$\begin{aligned} E(\widetilde{X}_{ij}^2\widetilde{Y}_{ij}^2) &= \Sigma_{11}^*\Sigma_{22}^* + 2(\Sigma_{12}^*)^2 \\ &= \frac{(n-1)^2}{n^2}(\sigma_j^2\tau_j^2 + 2\rho_j^2\sigma_j^2\tau_j^2) \end{aligned} \tag{4.24}$$

and

$$\begin{aligned} E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j}\widetilde{Y}_{i'j}) &= \Sigma_{12}^*\Sigma_{34}^* + \Sigma_{13}^*\Sigma_{24}^* + \Sigma_{14}^*\Sigma_{23}^* \\ &= \frac{(n-1)^2}{n^2}\rho_j\sigma_j\tau_j\rho_j\sigma_j\tau_j + \frac{1}{n^2}\sigma_j^2\tau_j^2 + \frac{1}{n^2}\rho_j^2\sigma_j^2\tau_j^2 \\ &= \frac{(n-1)^2+1}{n^2}\rho_j^2\sigma_j^2\tau_j^2 + \frac{1}{n^2}\sigma_j^2\tau_j^2 \end{aligned} \tag{4.25}$$

Therefore,

$$\begin{aligned} Var(C_j) &= nE(\widetilde{X}_{ij}^2\widetilde{Y}_{ij}^2) + n(n-1)\{E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j}\widetilde{Y}_{i'j})\} - n^2E^2(\widetilde{X}_{ij}\widetilde{Y}_{ij}) \\ &= n\frac{(n-1)^2}{n^2}(\sigma_j^2\tau_j^2 + 2\rho_j^2\sigma_j^2\tau_j^2) + n(n-1)\frac{(n-1)^2+1}{n^2}\rho_j^2\sigma_j^2\tau_j^2 \\ &\quad + \frac{1}{n^2}\sigma_j^2\tau_j^2 - n^2\frac{(n-1)^2}{n^2}\rho_j^2\sigma_j^2\tau_j^2 \\ &= (n-1)\rho_j^2\sigma_j^2\tau_j^2 \end{aligned} \tag{4.26}$$

$$\begin{aligned} Cov(C_j, C_{j'}) &= Cov(\sum_{i=1}^{n}\widetilde{X}_{ij}\widetilde{Y}_{ij}, \sum_{i=1}^{n}\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) \\ &= nCov(\widetilde{X}_{ij}\widetilde{Y}_{ij}, \widetilde{X}_{ij'}\widetilde{Y}_{ij'}) + n(n-1)Cov(\widetilde{X}_{ij}\widetilde{Y}_{ij}, \widetilde{X}_{i'j'}\widetilde{Y}_{i'j'}) \\ &= nE(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) - nE(\widetilde{X}_{ij}\widetilde{Y}_{ij})E(\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) \\ &\quad + n(n-1)\{E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j'}\widetilde{Y}_{i'j'}) - E(\widetilde{X}_{ij}\widetilde{Y}_{ij})E(\widetilde{X}_{i'j'}\widetilde{Y}_{i'j'})\} \\ &= nE(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) + n(n-1)E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j'}\widetilde{Y}_{i'j'}) \\ &\quad - n^2E(\widetilde{X}_{ij}\widetilde{Y}_{ij})E(\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) \end{aligned}$$

To derive $E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{ij'}\widetilde{Y}_{ij'})$ and $E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{i'j'}\widetilde{Y}_{i'j'})$, we need to further derive cross-time covariances. Without showing the derivation details, we have the following equations:

$$Cov(\widetilde{X}_{ij}, \widetilde{X}_{ij'}) = \frac{n-1}{n}\rho_x\sigma_j\sigma_{j'} \tag{4.27}$$

$$Cov(\widetilde{Y}_{ij}, \widetilde{Y}_{ij'}) = \frac{n-1}{n}\rho_y\tau_j\tau_{j'} \tag{4.28}$$

$$Cov(\widetilde{X}_{ij}, \widetilde{Y}_{ij'}) = \frac{n-1}{n}\rho_{xy}\sigma_j\tau_{j'} \tag{4.29}$$

$$Cov(\widetilde{X}_{ij'}, \widetilde{Y}_{ij}) = \frac{n-1}{n}\rho_{xy}\sigma_{j'}\tau_j \tag{4.30}$$

$$Cov(\widetilde{X}_{ij}, \widetilde{X}_{i'j'}) = -\frac{1}{n}\rho_x\sigma_j\sigma_{j'} \tag{4.31}$$

$$Cov(\widetilde{Y}_{ij}, \widetilde{Y}_{i'j'}) = -\frac{1}{n}\rho_y\tau_j\tau_{j'} \tag{4.32}$$

$$Cov(\widetilde{X}_{ij}, \widetilde{Y}_{i'j'}) = -\frac{1}{n}\rho_{xy}\sigma_j\tau_{j'} \tag{4.33}$$

$$Cov(\widetilde{Y}_{ij}, \widetilde{X}_{i'j'}) = -\frac{1}{n}\rho_{xy}\sigma_{j'}\tau_j \tag{4.34}$$

Note that $(\widetilde{X}_{ij}, \widetilde{Y}_{ij}, \widetilde{X}_{ij'}, \widetilde{Y}_{ij'})$ form a multivariate normal vector whose covariance matrix Σ^{**} can be extracted from Equations (4.18)-(4.19), (4.27)-(4.30):

$$\mathbf{\Sigma^{**}} = \begin{pmatrix} \frac{(n-1)}{n}\sigma_j^2 & \frac{(n-1)}{n}\rho_j\sigma_j\tau_j & \frac{(n-1)}{n}\rho_x\sigma_j\sigma_{j'} & \frac{(n-1)}{n}\rho_{xy}\sigma_j\tau_{j'} \\ \frac{(n-1)}{n}\rho_j\sigma_j\tau_j & \frac{(n-1)}{n}\tau_j^2 & \frac{(n-1)}{n}\rho_{xy}\sigma_{j'}\tau_j & \frac{(n-1)}{n}\rho_y\tau_j\tau_{j'} \\ \frac{(n-1)}{n}\rho_x\sigma_j\sigma_{j'} & \frac{(n-1)}{n}\rho_{xy}\sigma_{j'}\tau_j & \frac{(n-1)}{n}\sigma_{j'}^2 & \frac{(n-1)}{n}\rho_{j'}\sigma_{j'}\tau_{j'} \\ \frac{(n-1)}{n}\rho_{xy}\sigma_j\tau_{j'} & \frac{(n-1)}{n}\rho_y\tau_j\tau_{j'} & \frac{(n-1)}{n}\rho_{j'}\sigma_{j'}\tau_{j'} & \frac{(n-1)}{n}\tau_{j'}^2 \end{pmatrix} \tag{4.35}$$

As a result, we have:

$$\begin{aligned} E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) &= \Sigma^*_{12}\Sigma^*_{34} + \Sigma^*_{13}\Sigma^*_{24} + \Sigma^*_{14}\Sigma^*_{23} \\ &= \frac{(n-1)^2}{n^2}(\rho_j\sigma_j\tau_j\rho_{j'}\sigma_{j'}\tau_{j'} + \rho_x\sigma_j\sigma_{j'}\rho_y\tau_j\tau_{j'} + \rho_{xy}\sigma_j\tau_{j'}\rho_{xy}\sigma_{j'}\tau_j) \\ &= \frac{(n-1)^2}{n^2}\{\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_j\rho_{j'} + \rho_x\rho_y + \rho^2_{xy})\} \end{aligned}$$

Also, $(\widetilde{X}_{ij}, \widetilde{Y}_{ij}, \widetilde{X}_{i'j'}, \widetilde{Y}_{i'j'})$ form a multivariate normal vector whose covariance matrix Σ^{***} can be extracted from Equations (4.16), (4.18), (4.19), (4.31)-(4.34):

$$\mathbf{\Sigma^{***}} = \begin{pmatrix} \frac{n-1}{n}\sigma^2_j & \frac{n-1}{n}\rho_j\sigma_j\tau_j & -\frac{1}{n}\rho_x\sigma_j\sigma_{j'} & -\frac{1}{n}\rho_{xy}\sigma_j\tau_{j'} \\ \frac{n-1}{n}\rho_j\sigma_j\tau_j & \frac{n-1}{n}\tau^2_j & -\frac{1}{n}\rho_{xy}\sigma_{j'}\tau_j & -\frac{1}{n}\rho_y\tau_j\tau_{j'} \\ -\frac{1}{n}\rho_x\sigma_j\sigma_{j'} & -\frac{1}{n}\rho_{xy}\sigma_{j'}\tau_j & \frac{n-1}{n}\sigma^2_{j'} & \frac{n-1}{n}\rho_{j'}\sigma_{j'}\tau_{j'} \\ -\frac{1}{n}\rho_{xy}\sigma_j\tau_{j'} & -\frac{1}{n}\rho_y\tau_j\tau_{j'} & \frac{n-1}{n}\rho_{j'}\sigma_{j'}\tau_{j'} & \frac{n-1}{n}\tau^2_{j'} \end{pmatrix} \tag{4.36}$$

then we have:

$$\begin{aligned} E(\widetilde{X}_{ij}\widetilde{Y}_{ij}\widetilde{X}_{ij'}\widetilde{Y}_{ij'}) &= \Sigma^*_{12}\Sigma^*_{34} + \Sigma^*_{13}\Sigma^*_{24} + \Sigma^*_{14}\Sigma^*_{23} \\ &= \frac{(n-1)^2}{n^2}\rho_j\sigma_j\tau_j\rho_{j'}\sigma_{j'}\tau_{j'} + \frac{1}{n^2}\rho_x\sigma_j\sigma_{j'}\rho_y\tau_j\tau_{j'} + \frac{1}{n^2}\rho_{xy}\sigma_j\tau_{j'}\rho_{xy}\sigma_{j'}\tau_j \\ &= \frac{1}{n^2}\{\sigma_j\tau_j\sigma_{j'}\tau_{j'}[(n-1)^2\rho_j\rho_{j'} + \rho_x\rho_y + \rho^2_{xy}]\} \end{aligned}$$

Therefore,

$$\begin{aligned} Cov(C_j, C_{j'}) &= n\frac{(n-1)^2}{n^2}\{\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_j\rho_{j'} + \rho_x\rho_y + \rho^2_{xy})\} \\ &\quad + n(n-1)\frac{1}{n^2}\{\sigma_j\tau_j\sigma_{j'}\tau_{j'}[(n-1)^2\rho_j\rho_{j'} + \rho_x\rho_y + \rho^2_{xy}]\} \\ &\quad - n^2\frac{(n-1)^2}{n^2}(\rho_j\sigma_j\tau_j)(\rho_{j'}\sigma_{j'}\tau_{j'}) \\ &= (n-1)\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_x\rho_y + \rho^2_{xy}) \end{aligned} \tag{4.37}$$

Let ρ denote the true common serial correlation, $\rho_j = \rho$ under H_0. We substitute ρ for ρ_j and have the following:

$$E(C_j) = (n-1)\rho\sigma_j\tau_j \tag{4.38}$$

$$Var(C_j) = (n-1)(1+\rho^2)\sigma_j^2\tau_j^2 \tag{4.39}$$

$$Cov(C_j, C_{j'}) = (n-1)\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_x\rho_y + \rho_{xy}^2) \tag{4.40}$$

Therefore, $C = \sum_{j=1}^{m} C_j$ has variance

$$\{\sum_{j=1}^{m}(n-1)(1+\rho^2)\sigma_j^2\tau_j^2 + \sum_{j\neq j'}(n-1)\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_x\rho_y + \rho_{xy}^2)\}$$

Once we obtain the approximations to the values for $Var(B)$ and $Cov(C, B)$, which we denote as v and ϕ, respectively, then $\sqrt{m}(C/m - \rho B/m)$ has asymptotic variance

$$\{\sum_{j=1}^{m}(n-1)(1+\rho^2)\sigma_j^2\tau_j^2 + \sum_{j\neq j'}(n-1)\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_x\rho_y + \rho_{xy}^2) + \rho^2 v - 2\rho\phi)\}/m.$$

And given that B/m converges in probability to $\sum_{j=1}^{m}(n - 3/2 + \rho^2/2)\sigma_j\tau_j/m$, the asymptotic variance for r_{MH} as $m \to \infty$, denoted as $Var_m(r_{MH})$, is

$$\frac{\sum_{j=1}^{m}(n-1)(1+\rho^2)\sigma_j^2\tau_j^2 + \sum_{j\neq j'}(n-1)\sigma_j\tau_j\sigma_{j'}\tau_{j'}(\rho_x\rho_y + \rho_{xy}^2) + \rho^2 v - 2\rho\phi}{\{\sum_{j=1}^{m}(n - 3/2 + \rho^2/2)\sigma_j\tau_j\}^2}$$

We estimate the denominator of B with the sums of squares and plug in r_{MH} for ρ, giving us the estimate $\widehat{Var_m(r_{MH})}$ that has the following expression:

$$\frac{\sum_{j=1}^{m}(n-1)(1+r_{MH}^2)\hat{\sigma}_j^2\hat{\tau}_j^2 + \sum_{j\neq j'}(n-1)\hat{\sigma}_j\hat{\tau}_j\hat{\sigma}_{j'}\hat{\tau}_{j'}(\hat{\rho}_x\hat{\rho}_y + \hat{\rho}_{xy}^2) + r_{MH}^2 v - 2r_{MH}\phi}{(\sum_{j=1}^{m}\sqrt{\sum_{i=1}^{n}\widetilde{X}_{ij}^2\sum_{i=1}^{n}\widetilde{Y}_{ij}^2})^2} \tag{4.41}$$

4.3 Numerical Examples

4.3.1 Simulation Study

In this section, simulations were performed under settings similar to those examined in Chapter II and III. We examined the bias of our pooled correlation estimate as well as its variance estimate as $n \to \infty$; our simulations do not include a variance estimate when $m \to \infty$ due to the complications described in Section 4.2.3. For each subject i, $i = 1, 2, \ldots, n$, biomarker X and pathogen Y are both observed at m time points. We assume $X_{ij} \sim \mathcal{N}(\mu_{xj}, \sigma_j^2)$ and $Y_{ij} \sim \mathcal{N}(\mu_{yj}, \tau_j^2)$, in which $\mu_x = 2.5$ and $\mu_y = 4.0$, $\sigma_j = 0.3$, and $\tau_j = 0.40 - 0.05(j-1)$. In terms of the joint distribution of the data, we still considered two settings, the constant nuisance correlation setting that matches our assumed model and autoregressive nuisance correlation setting that has a model violation. However, we also considered several new scenarios in each setting that were not examined in previous chapters.

We first present simulation settings with constant nuisance correlation. The means and variances are the same as those used in the earlier simulations. For the nuisance correlation parameters, three sets of settings were considered: I. zero cross-correlations, in which we set $\rho_x = \rho_y = \rho_{xy} = 0$; II: medium cross-correlations, in which we set $\rho_x = \rho_y = 0.3$ and $\rho_{xy} = 0$; III. large cross-correlations, in which we set $\rho_x = 0.5$, $\rho_y = 0.7$, and $\rho_{xy} = 0$. We chose values for the common correlation of interest, $\rho \in \{0, 0.2, 0.4, 0.6\}$ in settings I and II, and $\rho \in \{0, 0.2\}$ in setting III. The reason we did not let the common serial correlation be larger than 0.2 in sub-setting III was that larger values led to a non-positive definite covariance ma-

trix for the data. For each combination of ρ and nuisance setting, we simulated $\boldsymbol{D_i}$=$\{X_{i1}, Y_{i1}, X_{i2}, Y_{i2}, \cdots, X_{im}, Y_{im}\}$, the data for each subject i, from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance as described above. We considered sample sizes of $n \in \{50, 100, 500\}$ and the number of time points $m \in \{3, 4, 5\}$. We computed the Mantel-Haenszel pooled correlation and its variance based on two variance formula, and also the empirical variance of the 1,000 Mantel-Haenszel correlation estimates.

We then present simulation settings for the setting with autoregressive nuisance correlation. The elements of $\boldsymbol{X_{i\cdot}}$ have an autoregressive correlation structure with correlation $\rho_{x_0}^{|j-k|}$, the correlation between Y_{ij} and Y_{ik} is similarly set to be $\rho_{y_0}^{|j-k|}$, and the cross-correlation between X_{ij} and Y_{ik} or between X_{ik} and Y_{ij} is $c\rho_{xy_0}^{|j-k|}$, where c is a positive constant. For the nuisance correlation parameters, two sets of settings were considered: I: medium cross-correlations, in which we selected the values $\rho_{x_0} = \rho_{y_0} = 0.3$, $\rho_{xy_0} = 0.7$ and $c = 1.7 \times \rho_{x_0}\rho_{y_0}$; II. large cross-correlations, in which we set $\rho_{x_0} = 0.5$, $\rho_{y_0} = 0.6$, $\rho_{xy_0} = 0.7$ and $c = 1.7 \times \rho_{x_0}\rho_{y_0}$. The correlation parameter of interest, ρ, was chosen to be $\rho \in \{0.2, 0.4, 0.6\}$. We also considered sample sizes of $n \in \{50, 100, 500\}$ and the number of time points $m \in \{3, 4, 5\}$.

4.3.2 Simulation Results

Table 4.1 presents the true common serial correlation and its empirical standard error, as well as the Mantel-Haenszel serial correlation estimate and its proposed theoretical (formula based as $n \to \infty$) variances (presented in terms of standard error) when there is zero cross correlations for various combinations of n, m, and ρ. The performance was evaluated out of 1,000 simulations. Each theoretical value is the average of 1,000 formula-based values.

Overall, the Mantel-Haenszel pooled correlation is close to the true common correlation. Regarding standard error estimates, when ρ is small (≤ 0.2), with a sample size of $n = 500$, the estimates are very close to the true value $ES(MH)$, regardless of the number of time points. With sample sizes of $n = 50$, the standard errors obtained by formula derived based on $n \to \infty$ are slightly inflated, but still within an acceptable range. As ρ increases, the standard error developed based on the asymptotics of n remains almost the same as the true value or inflates just by a small amount in small samples ($n = 50$).

A similar trend can be seen when a medium level of constant cross correlations exist, as shown in Table 4.2. Finally, Table 4.3 shows the performance of the proposed variance estimator in data with large constant cross correlations. With a large or medium sample size ($n = 500$ or 100), the variance estimates are close to the true value, while with sample sizes of $n = 50$, our variance estimate tends to underestimate the true value slightly when $m = 4$ or $m = 5$.

We should expect that our variance estimator performs the best when sample size is large, according to its asymptotic features. In general, that is what we have observed. We repeated the simulations with $n = 25$ (results not shown) and the proposed variance estimator $Var_n(MH)$ still performs well.

We then examined a situation when there is model mis-specification. Table 4.4 and Table 4.5 contain r_{MH} and the estimated standard error $S_n(MH)$ for simulated data with medium and large autoregressive nuisance correlations. The Mantel-Haenszel pooled correlation remains close to the true common correlation in our settings. Table 4.4 indicates that if medium autoregressive nuisance correlations exist and ρ is as small as 0.2, the formula based $n \to \infty$ gave relatively accurate estimation with a sample size of $n = 500$. As ρ goes up or the sample size decreases, $S_n(MH)$ remains close to

Table 4.1: The estimated pooled serial correlation r_{MH} and the corresponding standard error estimator with zero nuisance correlation (1000 replications). $S_n(MH)$ = standard error obtained by formula derived based on $n \to \infty$; $ES(MH)$ = empirical standard error of r_{MH}.

		$n = 500$		
m	ρ	r_{MH}	$S_n(MH)$	$ES(MH)$
3	0	-0.002	0.026	0.026
	0.2	0.198	0.025	0.025
	0.4	0.398	0.022	0.022
	0.6	0.599	0.017	0.017
4	0	-0.001	0.023	0.023
	0.2	0.199	0.022	0.022
	0.4	0.400	0.019	0.019
	0.6	0.600	0.015	0.014
5	0	-0.001	0.021	0.021
	0.2	0.199	0.020	0.020
	0.4	0.400	0.017	0.017
	0.6	0.600	0.013	0.013
		$n = 100$		
3	0	0.000	0.059	0.056
	0.2	0.201	0.057	0.054
	0.4	0.401	0.050	0.047
	0.6	0.601	0.038	0.036
4	0	-0.002	0.052	0.052
	0.2	0.198	0.050	0.050
	0.4	0.399	0.043	0.044
	0.6	0.599	0.033	0.033
5	0	-0.002	0.047	0.047
	0.2	0.199	0.045	0.045
	0.4	0.400	0.039	0.040
	0.6	0.600	0.030	0.030
		$n = 50$		
3	0	0.000	0.085	0.082
	0.2	0.201	0.082	0.079
	0.4	0.402	0.071	0.069
	0.6	0.602	0.054	0.052
4	0	-0.002	0.074	0.071
	0.2	0.199	0.071	0.068
	0.4	0.400	0.063	0.060
	0.6	0.600	0.048	0.045
5	0	0.002	0.067	0.066
	0.2	0.203	0.065	0.063
	0.4	0.404	0.057	0.055
	0.6	0.604	0.043	0.042

Table 4.2: The estimated pooled serial correlation r_{MH} and the corresponding standard error estimator with medium constant nuisance correlation (1000 replications).$S_n(MH)$ = standard error obtained by formula derived based on $n \to \infty$; $ES(MH)$ = empirical standard error of r_{MH}.

		$n = 500$		
m	ρ	r_{MH}	$S_n(MH)$	$ES(MH)$
3	0	-0.002	0.028	0.029
	0.2	0.198	0.027	0.028
	0.4	0.398	0.025	0.025
	0.6	0.598	0.021	0.021
4	0	-0.001	0.026	0.025
	0.2	0.199	0.025	0.025
	0.4	0.399	0.023	0.023
	0.6	0.600	0.020	0.020
5	0	-0.001	0.024	0.024
	0.2	0.200	0.023	0.023
	0.4	0.400	0.022	0.021
	0.6	0.600	0.019	0.019
		$n = 100$		
3	0	-0.001	0.064	0.063
	0.2	0.199	0.062	0.061
	0.4	0.400	0.056	0.056
	0.6	0.600	0.048	0.048
4	0	-0.002	0.058	0.058
	0.2	0.198	0.056	0.057
	0.4	0.399	0.052	0.052
	0.6	0.600	0.045	0.046
5	0	-0.003	0.054	0.054
	0.2	0.198	0.053	0.052
	0.4	0.399	0.049	0.049
	0.6	0.599	0.044	0.044
		$n = 50$		
3	0	-0.001	0.092	0.091
	0.2	0.200	0.089	0.088
	0.4	0.401	0.081	0.080
	0.6	0.601	0.068	0.068
4	0	-0.004	0.083	0.080
	0.2	0.198	0.081	0.077
	0.4	0.400	0.074	0.071
	0.6	0.601	0.064	0.062
5	0	0.002	0.077	0.078
	0.2	0.203	0.075	0.076
	0.4	0.404	0.070	0.070
	0.6	0.604	0.062	0.062

Table 4.3: The estimated pooled serial correlation r_{MH} and the corresponding standard error estimator with large constant nuisance correlation (1000 replications). $S_n(MH)$ = standard error obtained by formula derived based on $n \to \infty$; $ES(MH)$ = empirical standard error of r_{MH}.

		$n = 500$		
m	ρ	r_{MH}	$S_n(MH)$	$ES(MH)$
3	0	-0.003	0.034	0.034
	0.2	0.197	0.033	0.034
4	0	-0.001	0.032	0.032
	0.2	0.199	0.032	0.032
5	0	0.000	0.031	0.031
	0.2	0.200	0.031	0.031
		$n = 100$		
3	0	-0.003	0.076	0.077
	0.2	0.198	0.075	0.076
4	0	-0.002	0.073	0.073
	0.2	0.199	0.072	0.073
5	0	-0.004	0.070	0.070
	0.2	0.197	0.070	0.069
		$n = 50$		
3	0	-0.002	0.109	0.110
	0.2	0.200	0.107	0.108
4	0	-0.005	0.103	0.101
	0.2	0.198	0.103	0.100
5	0	0.002	0.100	0.104
	0.2	0.204	0.100	0.103

the true value. With large autoregressive nuisance correlations (Table 4.5), in some data scenarios $S_n(MH)$ is smaller than the true standard error, but only by a very small amount. Therefore, $S_n(MH)$ is fairly robust to model mis-specification.

4.3.3 Real Data Example

We analyzed the same two sets of data as that in previous chapters described by Kinney et al. (2011) and Ramseier et al. (2009). For the serum biomarker dataset, the concentration levels of MMP-8, MMP-9 and calprotectin were found to be strong predictors of periodontitis and all three plaque pathogens demonstrated stronger associations than the four biomarkers (Ramseier et al. (2009)). After checking whether or not the correlation between each biomarker and each pathogen were constant over time, we found significant heterogeneity exists between TNF-α and *T.forsythia*, MMP-9 and *P.gingivalis* at $\alpha = 0.05$ level, and MMP-8 and *P.gingivalis* $\alpha = 0.10$ level. In the salivary biomarker dataset, we found significant heterogeneity exists between MMP-8 and *T.forsythia* at $\alpha = 0.05$ level, OPG and *T.denticola* , between MMP-9 and *T.forsythia*, MMP-8 and *P.gingivalis*, OPG and *T.denticola* at $\alpha = 0.10$ level. For the rest pairs in the two datasets, there is no evidence of heterogeneity, leading to a decision of pooling serial correlations such that the association between a pathogen and a biomarker can be summarized by one single value for all time points.

4.3.3.1 Serum biomarker dataset

Table 4.6 contains the serial correlations of each log-transformed serum biomarker with each log-transformed pathogen. The table also summarizes the Mantel-Haenszel pooled serial correlation between combinations of biomarkers and pathogens, the standard error (se) of r_{MH} computed from the formula assuming the number of subjects

Table 4.4: The estimated pooled serial correlation r_{MH} and the corresponding standard error estimator with medium autoregressive nuisance correlation (1000 replications). $S_n(MH)$ = standard error obtained by formula derived based on $n \to \infty$; $ES(MH)$ = empirical standard error of r_{MH}.

		$n = 500$		
m	ρ	r_{MH}	$S_n(MH)$	$ES(MH)$
3	0.2	0.198	0.026	0.027
	0.4	0.398	0.023	0.023
	0.6	0.598	0.018	0.018
4	0.2	0.199	0.023	0.023
	0.4	0.399	0.020	0.020
	0.6	0.600	0.015	0.016
5	0.2	0.199	0.021	0.021
	0.4	0.400	0.018	0.019
	0.6	0.600	0.014	0.015
		$n = 100$		
3	0.2	0.199	0.059	0.058
	0.4	0.400	0.052	0.051
	0.6	0.600	0.041	0.041
4	0.2	0.198	0.052	0.053
	0.4	0.399	0.045	0.046
	0.6	0.599	0.035	0.036
5	0.2	0.199	0.047	0.048
	0.4	0.399	0.041	0.042
	0.6	0.600	0.032	0.033
		$n = 50$		
3	0.2	0.200	0.085	0.085
	0.4	0.401	0.075	0.075
	0.6	0.602	0.058	0.059
4	0.2	0.198	0.075	0.072
	0.4	0.399	0.065	0.063
	0.6	0.600	0.051	0.050
5	0.2	0.203	0.067	0.068
	0.4	0.404	0.059	0.060
	0.6	0.604	0.045	0.047

Table 4.5: The estimated pooled serial correlation r_{MH} and the corresponding standard error estimator with large autoregressive nuisance correlation (1000 replications). $S_n(MH)$ = standard error obtained by formula derived based on $n \to \infty$; $ES(MH)$ = empirical standard error of r_{MH}.

		$n = 500$		
m	ρ	r_{MH}	$S_n(MH)$	$ES(MH)$
3	0.2	0.199	0.030	0.031
	0.4	0.399	0.026	0.026
	0.6	0.599	0.019	0.020
4	0.2	0.200	0.027	0.028
	0.4	0.400	0.023	0.023
	0.6	0.600	0.017	0.018
5	0.2	0.199	0.024	0.026
	0.4	0.399	0.020	0.021
	0.6	0.600	0.015	0.016
		$n = 100$		
3	0.2	0.199	0.068	0.068
	0.4	0.399	0.058	0.058
	0.6	0.600	0.044	0.044
4	0.2	0.199	0.061	0.062
	0.4	0.399	0.051	0.052
	0.6	0.600	0.039	0.040
5	0.2	0.200	0.055	0.057
	0.4	0.400	0.046	0.048
	0.6	0.601	0.035	0.037
		$n = 50$		
3	0.2	0.199	0.097	0.098
	0.4	0.400	0.083	0.083
	0.6	0.601	0.064	0.064
4	0.2	0.199	0.087	0.087
	0.4	0.400	0.073	0.074
	0.6	0.602	0.056	0.056
5	0.2	0.199	0.079	0.082
	0.4	0.401	0.067	0.069
	0.6	0.602	0.050	0.052

goes to infinity, and the p-value of Z-test of whether the pooled Mantel-Haenszel correlation is zero. The test statistic Z was computed by $Z = r_{MH}/se$. se_n is the standard error obtained by the formula derived based on $n \to \infty$; $p.n$ is the p-value obtained by se_n. Since we found significant heterogeneity between TNF-α and *T.forsythia* and MMP-9/MMP-8 and *P.gingivalis*, we did not compute r_{MH} for these combinations.

Since our data indicate that most serial correlations are close to 0 and ranged between -0.2 and 0.2, and the cross correlations (nuisance correlation) are nearly constant, which is similar to our first simulation data setting. As we expected, Table 4.6 shows that the Mantel-Haenszel pooled correlation estimate is close to zero. However, none of the p-values ($p.n$) is less than 0.05, showing no evidence that the pooled correlation is significantly different from zero in any pathogen-biomarker combination.

4.3.3.2 Salivary biomarker dataset

Table 4.7 contains the serial correlations of each salivary biomarker with each pathogen after the transformation and Mantel-Haenszel pooled serial correlation together with the standard error and p-values. Our results show that among the pathogen/biomarker pairs demonstrating a homogeneous serial correlation, the pooled correlation is significantly different from zero between MMP-8/MMP-9 and *T.denticola*. As shown in Table 4.7, the greatest r_{MH} values (> 0.3) exist between MMP-8/MMP-9 and *T.denticola*, while the pooled correlations obtained from all the other pairs are below 0.3. These findings, along with the previous biological findings in Gamonal et al. (2011) about the progressive increase in MMP-8 and MMP-9 levels in saliva samples as the degree of periodontitis develops, have provided both statistical and biological evidence of a homogeneous and substantial correlation between MMP-8/MMP-9 and

Table 4.6: Mantel-Haenszel pooled serial correlation between combinations of serum biomarkers and pathogens, and test whether the pooled correlation is zero. se_n=standard error obtained by formula derived based on $n \to \infty$; $p.n$=p-value obtained by se_n; n/a=serial heterogeneity found (test of homodeneity has a p-value less than 0.1).

Pathogen	Biomarker	Sample serial correlation			p-value of testing $r_{MH} = 0$		
		0	6	12	r_{MH}	se_n	p.n
P.gingivalis	TNF-α	-0.17	0.07	-0.07	-0.057	0.095	0.551
	calprotectin	-0.18	-0.01	0.11	-0.085	0.083	0.303
	MMP-8	-0.01	0.28	0.04	n/a	n/a	n/a
	MMP-9	0.12	0.19	-0.14	n/a	n/a	n/a
T.forsythia	TNF-α	-0.19	0.07	-0.19	n/a	n/a	n/a
	calprotectin	-0.21	-0.02	0.05	-0.120	0.085	0.159
	MMP-8	0.00	0.22	0.13	0.110	0.090	0.222
	MMP-9	0.13	0.19	-0.03	0.126	0.093	0.177
T.denticola	TNF-α	-0.13	-0.04	0.03	-0.060	0.094	0.523
	calprotectin	-0.17	0.01	0.13	-0.068	0.081	0.401
	MMP-8	-0.01	0.08	0.00	0.030	0.086	0.729
	MMP-9	0.17	0.02	-0.02	0.064	0.088	0.471

red complex pathogens.

4.4 Conclusion

In this Chapter, we propose a Mantel-Haenszel-type estimator of the pooled correlation coefficient and the corresponding variance estimators. Two variance estimators are proposed based on the asymptotics as number of time points or number of subjects goes to infinity. Our proposed Mantel-Haenszel pooled correlation and the variance estimator based on $n \to \infty$ perform well according to the bias evaluation in different settings via simulations. Conclusions are: **(i)** Mantel-Haenszel pooled correlation r_{MH} is close to the true common correlation. **(ii)** $S_n(MH)$, the standard error given by our variance formula is close to $ES(MH)$ obtained empirically from the 1000 simulated data. **(iii)** The simulation setting indicates that even if autoregressive nuisance correlations exist such that the true data violates our model assumption of constant nuisance correlations, our formula (4.13) still maintains an accurate estimation. **(iv)** Even with a smaller sample size, $\widehat{Var_n}(r_{MH})$ still stays close to the true value.

Table 4.7: Mantel-Haenszel pooled serial correlation between combinations of salivary biomarkers and pathogens, and test whether the pooled correlation is zero. se_n=standard error obtained by formula derived based on $n \to \infty$; $p.n$=p-value obtained by se_n; n/a=serial heterogeneity found (test of homodeneity has a p-value less than 0.1).

Pathogen	Biomarker	Sample serial correlation			p-value of testing $r_{MH} = 0$		
		0	6	12	r_{MH}	se_n	p.n
P.gingivalis	IL-1β	0.21	0.04	0.07	0.121	0.137	0.378
	MMP-8	0.38	0.40	0.03	n/a	n/a	n/a
	MMP-9	0.30	0.20	-0.03	0.203	0.126	0.108
	OPG	0.26	0.02	-0.08	0.098	0.133	0.460
T.forsythia	IL-1β	0.23	0.26	0.20	0.229	0.140	0.102
	MMP-8	0.35	0.56	-0.10	n/a	n/a	n/a
	MMP-9	0.30	0.13	-0.15	n/a	n/a	n/a
	OPG	0.27	-0.01	0.16	0.149	0.141	0.291
T.denticola	IL-1β	0.36	-0.03	0.16	0.201	0.137	0.142
	MMP-8	0.41	0.41	0.20	0.366	0.133	0.006
	MMP-9	0.35	0.41	0.07	0.328	0.123	0.008
	OPG	0.38	0.21	-0.06	n/a	n/a	n/a

CHAPTER V

Summary and Future Work

Motivated by a periodontal study that longitudinally measured serum and salivary biomarkers and levels of bacterial pathogens in the oral cavity, this dissertation aimed to look at two issues in longitudinal studies: developing tests for the equality of serial correlation between two longitudinally measured continuous variables over time, and pooling serial correlation coefficients after they are tested to be homogeneous. In Chapter II, we constructed a χ^2_ρ test and a χ^2_z test based on asymptotic distributions for both untransformed and Fisher's Z-transformed sample correlation coefficients, respectively. We also proposed an F_ρ statistic that is derived from χ^2_ρ. We evaluated the empirical size and power of our proposed tests in a variety of settings. We found that F_ρ and χ^2_z tests have nominal sizes when the data fit our assumed model, and they are generally robust to mis-specified models and heavier tailed data. The χ^2_ρ test has a more inflated size in small samples and some medium sized samples. In Chapter III, we introduced Bayesian posterior credible intervals and Bayesian posterior predictive p-values to perform tests of equality of correlation coefficients for longitudinal studies. We evaluate two types of credible intervals–HPD and QB in a simple data setting with only two repeated measures. When there are only two time points, HPD is preferred to QB. When the number of time points is bigger than two, the credible interval

is hard to obtain. Therefore, we explored Bayesian posterior predictive p-values in data with more than two repeated measures. We borrowed the classical Wald χ^2_ρ and χ^2_z statistics from Chapter II to construct test statistics T_ρ and T_z, and discrepancy measures Q_ρ and Q_z. Our simulation results suggest that the posterior predictive p-values based on T_ρ and T_z have comparable performance in identifying heterogeneity in medium sized datasets (e.g. $n = 50$) compared to Wald tests that are based on asymptotics. In Chapter IV, we proposed a Mantel-Haenszel-type estimator of the pooled correlation coefficient, denoted as r_{MH}, and developed the corresponding asymptotic variance estimate, denoted as $Var_n(MH)$, as the sample size goes to infinity. We observed that the Mantel-Haenszel estimator for the pooled correlation is close to the true common correlation, and that $Var_n(MH)$ also performs well even with a small sample size. In addition, $Var_n(MH)$ remains stable and robust to model mis-specification.

Our proposed tests for homogeneity and estimators for pooled correlation coefficient have been applied to a longitudinal periodontal study to investigate the association of pathogenic bacteria and serum/salivary biomarkers in gingivitis. We found heterogeneity between certain red complex pathogens and inflammatory biomarkers, suggesting for these particular pairs, the time-specific correlation, especially at the time point with the biggest correlation is of most interest, rather than the overall correlation. Our findings suggested that we can identify the highest time-specific correlation (such as month six in our example, which is interestingly the period when periodontal disease is the most severe), and the pair of pathogen/biomarker at this particular time point can be jointly examined to quantify disease severity. On the other hand, among the pathogen/biomarker pairs with homogeneous serial correlation, the correlation between *T. denticola* and salivary MMP8/MMP9 is significantly

different from 0. The take-home message from this finding is that *T. denticola* and salivary MMP8/MMP9 are constantly correlated over time and they can be used jointly as a stable predictor of oral inflammatory/disease progression. The particular pairs with constant, non-zero correlations may be very valuable predictors and can save the cost of performing bioassays on many other biomarkers that demonstrate less relevance with periodontitis-pathogenic bacteria.

In Chapter II, a limitation of our method is that we found suitable values for λ and δ needed for our $\mathcal{F}$-test through a grid search of possible values. However, it should be noted that the values could be found directly by equating the first two moments of $\lambda\chi^2_\rho/(m-1)$ and $F(m-1,d)$. However, this approach requires use of the Delta method to obtain the variance of a function of all the correlation parameters, which is computationally intensive. Finding these values and comparing them to those we used would prove interesting to determine if the added computational burden is warranted.

Another limitation is that our tests have poor power when the sample size decreases or the number of time points increases. Furthermore, when the distribution of the data does not follow an exact multivariate normal distribution, simulations showed that our tests have inflated size. An asymptotically exact test maybe developed to assess homogeneity in small samples or non-normal samples. Sakaori (2002a) used an asymptotically exact permutation test for the equality of two correlation coefficients for data from a trivariate normal distribution. He also developed an asymptotically exact permutation test for equality of correlation coefficients in two independent populations (Sakaori, 2002b), which was extended by Omelka and Pauly (2012) to incorporate non-normal data. Although our setting is more complicated than the settings of the cited works, we may also develop a permutation test to determine if serial correlation is time-invariant to address the issue of inflated Type I

error rates of our asymptotic tests.

In Chapter III, the posterior predictive tests in our simulations are generally conservative. Meng (1994) suggested that under the prior predictive distribution, posterior predictive p-value puts less weight to the extreme values, thus it is more closely centered around 1/2 compared with a uniform variable. This indicates that there exists an α_0 small enough such that for all $\alpha \in [0, \alpha_0]$ the probability of observing a p-value less than α given the null hypothesis will never exceed α. Thus, our finding that the posterior predictive tests in our simulations are conservative is supported by Meng (1994) who stated that a posterior predictive test often will have a level close to or less than that desired. Although in our particular problem, Bayesian posterior predictive method does not seem to provide a better power than Wald test, it offers an alternative way of examining the hypothesis testing problem. Besides, the concept of discrepancy variables has been introduced to our test, which allows us to incorporate unknown nuisance parameters into test statistics, rather than plugging an estimator (e.g., maximum-likelihood estimator) obtained from the original data for the parameters.

The biggest challenge we were faced with in Chapter III was that our methods involved sampling parameters from their posterior distribution and generating replicated data in each iteration, and several thousand iterations were needed to get a posterior predictive p-value. The computation time is much longer than performing a Wald test in the R environment. One future step to improve current method is to perform MCMC sampling using another programming language, for example, C, to shorten the computation time.

In Chapter IV, we have yet to derive a variance formula as the number of time points goes to infinity, which is analogous to the methods of Breslow (1981). However,

unlike Breslow (1981)'s 2×2 table settings in which each stratum is independent, the assumption of independence is not appropriate for longitudinal data since within-subject correlation exists among time points. As a future direction, we will look for solutions to obtain the asymptotic variance of $C - \rho B$ as the number of time points goes to infinity. An alternative approach would be to bootstrap the data and then directly compute the variance of r_{MH} over the bootstrap samples.

Our estimators have been only applied to normal data. It is desired to evaluate the estimators in skewed (non-normal) data. Moreover, our current test statistics in Chapter II and Chapter III and estimators in Chapter IV do not accommodate missing data and a common sample size is assumed for all time points. However, in real studies with repeated measures, dropouts is a common issue and needs to be considered. We will do sensitivity analyses in our future simulations to evaluate our methods assuming data are missing at random. Furthermore, it may be of interest to develop a variation form of test statistics or estimators to allow the sample size to vary from one time point to another.

BIBLIOGRAPHY

BIBLIOGRAPHY

J. Barnard, R. McCulloch, and X. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–311, 2000.

N.E. Breslow. Odds ratio estimators when the data are sparse. *Biometrika*, 68:73–84, 1981.

N.E. Breslow and N.E Day. *Classical methods of analysis of grouped data. In: Statistical Methods in Cancer Research: The analysis of case-control studies*, volume 1. International Agency for Research on Cancer, Lyon, 1980.

N.E. Breslow and K.E. Liang. The variance of the mantel-haenszel estimator. *Biometrics*, 38:943–952, 1982.

M.H. Chen and Q.M. Shao. Monte carlo estimation of bayesian credible and hpd intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92, 1999.

M.J. Daniels. A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, 27:567–578, 1992.

M.J. Daniels and R.E. Kass. Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94:1254–1263, 1999.

A. Donner and G. Zou. Testing the equality of dependent intraclass correlation coefficients. *Journal of the Royal Statistical Society*, 51:367–379, 2002.

O. J. Dunn and V. Clark. Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, 64:366–377, 1969.

O. J. Dunn and V. Clark. Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association*, 66:904–908, 1971.

R. C. Elston. On the correlation between correlations. *Biometrika*, 62:133–140, 1975.

R. A. Fisher. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10:507–521, 1915.

R. A. Fisher. On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1:3–32, 1921.

J. Gamonal, N. Silva, M. Hernández, N Dutzan, J. Garcia-Sesnich, L. Abusleme, A. Dezerega, and R. Vernal. *Gingival Diseases - Their Aetiology, Prevention and Treatment*, chapter Components of Host Response to Pathogenic Bacteria in Gingivitis. InTech, 2011.

A. Gelman, X. Meng, and H. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–807, 1996.

W. W. Hauck. The large sample variance of the mantel-haenszel estimator of a common odds ratio. *Biometrics*, 35:817–819, 1979.

H. Hotelling. The selection of variates for use in prediction with some comments on the general problem of nuissance parameters. *Annals of Mathematical Statistics*, 11:271–283, 1940.

L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–139, 1918.

M. Kenward and J. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53:983–997, 1997.

J. S. Kinney, T. Morelli, T. Braun, C. A. Ramseier, A. E. Herr, J. V. Sugai, C. E.

Shelburne, L. A. Rayburn, A. K. Singh, and W. V. Giannobile. Saliva/pathogen biomarker signatures and periodontal disease progression. *Journal of Dental Research*, 90:752–8, 2011.

H. Kuula, T. Salo, E. Pirila, A. Tuomainen, M. Jauhiainen, V. Uitto, L Tjaderhane, P.J. Pussinen, and T. Sorsa. Local and systemic responses in matrix metalloproteinase 8-deficient mice during porphyromonas gingivalis-induced periodontitis. *Infection and Immunity*, 77:850–859, 2009.

N. Mantel and W. Haenszel. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22:719–48, 1959.

X. Meng. Posterior predictive p-values. *Annals of Statistics*, 22:1142–1160, 1994.

X. L. Meng, R. Rosenthal, and D. B. Rubin. Comparing correlated correlation coefficients. *Psychological Bulletin*, 111:172–175, 1992.

I. Olkin and J. D. Finn. Testing correlated correlations. *Psychological Bulletin*, 108: 330–333, 1990.

I. Olkin and M. Siotani. Asymptotic distribution of functions of a correlation matrix. *Essays in Probability and Statistics*, pages 235–251, 1976.

M. Omelka and M. Pauly. Testing equality of correlation coefficients in two populations via permutation methods. *Journal of Statistical Planning and Inference*, 142: 1396–1406, 2012.

T. Raghunathan. An approximate test for homogeneity of correlated correlation coefficients. *Quality and Quantity*, 37:99–110, 2003.

T. Raghunathan, R. Rosenthal, and D. B. Rubin. Comparing correlated but nonoverlapping correlations. *Psychological Methods*, 1:178–183, 1996.

B. Rai, S. Kharb, R. Jain, and S. C. Anand. Biomarkers of periodontitis in oral fluids.

Journal of Oral Science, 50:53–56, 2008.

C. A. Ramseier, J. S. Kinney, A. E. Herr, T. M. Braun, J. V. Sugai, C. A. Shelburne, L. A. Rayburn, H. M. Tran, A. K. Singh, and W. V. Giannobile. Identification of pathogen and host-response markers correlated with periodontal disease. *Journal of Periodontology*, 80:436–46, 2009.

J. Robins, S. Greenland, and N.E. Breslow. A general estimator for the variance of the mantel-haenzel odds ratio. *American Journal of Epidemiology*, 124:719–723, 1986.

F. Sakaori. A nonparametric test for the equality of dependent correlation coefficients under normality. *Communications in Statistics - Theory and Methods*, 31:2379–2389, 2002a.

F. Sakaori. Permutation test for equality of correlation coefficients in two populations. *Communications in Statistics - Simulation and Computation*, 31(4):641–651, 2002b.

G. E. Salvi, L. M. Franco, T. M. Braun, A. Lee, G. R. Persson, N. P. Lang, and W. V. Giannobile. 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, 2010.

P. Silcocks. An easy approach to the robins-breslow-greenland variance estimator. *Epidemiologic Perspectives and Innovations*, 2:9–14, 2005.

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

S. S. Socransky, C. Smith, and A. D. Haffajee. Subgingival microbial profiles in refractory periodontal disease. *Journal of Clinical Periodontology*, 29:260–268, 2002.

X. M. Tu, J. Kowalski, P. Crits-Christoph, and R. Gallop. Power analysis for corre-

lations from clustered study designs. *Statistics in Medicine*, 25:2587–2606, 2006.

M. Yakob, K. Kari, T. Tervahartiala, T. Sorsa, P.-O. Soder, J.H. Meurman, and Soder B. Associations of periodontal microorganisms with salivary proteins and mmp-8 in gingival crevicular fluid. *Journal of Clinical Periodontology*, 39:256–263, 2012.

M. C. Yu and O. J. Dunn. Robust tests for the equality of two correlation coefficients: A monte carlo study. *Educational and Psychological Measurement*, 42:987–1004, 1982.

O. O. Yucel, E. Berker, S. Gariboglu, and H. Otlu. Interleukin-11, interleukin-1beta, interleukin-12 and the pathogenesis of inflammatory periodontal diseases. *Journal of Clinical Periodontology*, 35:365–370, 2008.

J. Zhang, S. Kashket, and P. Lingstrom. Evidence for the early onset of gingival inflammation following short-term plaque accumulation. *Journal of Clinical Periodontology*, 29:1082–1085, 2002.