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

Su Chen* Timothy D. Johnson† Thomas M. Braun‡

Abstract

We wish to determine if the pair-wise correlations of time-adjacent longitudinal data are homogeneous and can be pooled into a single time-invariant value. Historically, this hypothesis has been addressed using an asymptotic Wald test. As an alternative, we propose to evaluate whether the correlation between two continuous biomarkers measured at several time points are equal by examining the posterior predictive p-values within the Bayesian paradigm. We decompose the variance/covariance matrix to standard deviation elements and correlation elements and run a Metropolis-Hastings with Gibbs algorithm. A replicated dataset is generated conditional on each draw of parameters, and a "test statistic" is calculated for both the original and replicated dataset, by which the posterior predictive p-value is determined. The performance of our method is examined via simulation studies. We demonstrate our method on a periodontal longitudinal data set.

Key Words: longitudinal study, Metropolis-Hastings, posterior p-value, serial correlation

1. Introduction

Our method is motivated by a small longitudinal study of gingivitis, or inflammation of the gums (Salvi et al. (2010)). Eighteen subjects were enrolled in this study and were instructed to refrain from all oral health practices for 21 days so that the natural progression of oral disease could occur. After 21 days, the subjects returned 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) to measure levels of oral pathogens and biomarkers. It has been suggested that the level of oral pathogens may directly trigger an immune response and thereby promote increased levels of inflammatory biomarkers. Our current goal is to assess the association of biomarkers and pathogens expressed by Pearson correlations,and whether the associations changed over the course of the study.

Historically, several published statistical methods exist for assessing homogeneity of correlation coefficients. Examples are Olkin and Siotani (1976), Olkin and Finn (1990) and Dunn and Clark (1969, 1971) that use asymptotic normality of sample correlation coefficients or Fisher's Z-transformed sample correlation coefficients to construct asymptotic Wald test statistics. However, since more controversies arise from Bayesian school that the calculation of frequentist p-value involves violation of the likelihood principle, we are motivated to explore model checking approaches within Bayesian paradigm. We thus fit our problem into a model selection framework. There are different ways of doing Bayesian model selection, the most commonly used one is perhaps Bayes factor. However, it is often difficult to calculate, especially for models that involve large numbers of unknowns or improper priors. Implying the same test statistics used in Wald test, we instead chose the idea of posterior predictive p-value via MCMC, proposed by Meng (1994) and Gelman et

*Biostatistics Department, 1415 Washington Heights,Ann Arbor, Michigan 48109
†Biostatistics Department, 1415 Washington Heights,Ann Arbor, Michigan 48109
‡Biostatistics Department, 1415 Washington Heights,Ann Arbor, Michigan 48109

al. (1996). To perform MCMC, Sampling parameters of the model under null hypothesis involves modeling a variance-covariance structure Σ. A direct variance/covariance decomposition suggested by Barnard et al. (2000) that allows us to work with standard deviations and correlation matrix separately is of most interest. Specifically, the covariance matrix can be written as

$$\Sigma = diag(S)\ R\ diag(S)$$

where S is the vector of m standard deviations and R is the $m \times m$ correlation matrix. Different priors can be put on S and R respectively. In our study, we choose to use marginally uniform priors for correlations and Gamma priors for standard deviations. Detailed procedures of sampling from posterior distribution and obtaining posterior p-values will be described in the following sections.

2. Method

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 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.

Explicitly, if we 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, we assume $\boldsymbol{D}_i$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ in which

$$\boldsymbol{\mu} = \{\mu_{x_1}, \mu_{y_1}, \mu_{x_2}, \mu_{y_2}, \cdots, \mu_{x_m}, \mu_{y_m}\} \tag{1}$$

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 Prior specifications

According to the direct decomposition strategy suggested by Barnard et al.(2000), Σ in equation (2) is decomposed into S=$\{\sigma_1, \tau_1, \sigma_2, \tau_2, \cdots, \sigma_m, \tau_m\}$ and

$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho_1 & \rho_x & \rho_{xy} & \cdots & \rho_x & \rho_{xy} \\ \rho_1 & 1 & \rho_{xy} & \rho_y & \cdots & \rho_{xy} & \rho_y \\ \rho_x & \rho_{xy} & 1 & \rho_2 & \cdots & \rho_x & \rho_{xy} \\ \rho_{xy} & \rho_y & \rho_2 & 1 & \cdots & \rho_{xy} & \rho_y \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_x & \rho_{xy} & \rho_x & \rho_{xy} & \cdots & 1 & \rho_m \\ \rho_{xy} & \rho_y & \rho_{xy} & \rho_y & \cdots & \rho_m & 1 \end{pmatrix} \tag{3}$$

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 uninformative prior for μ: $\pi(\mu) \propto 1$; Gamma(2,2) prior for σ_j^2 and τ_j^2; and uniform(-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 R matrix is positive definite.

2.3 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 $\boldsymbol{Y_{.k}}$. Let $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_m$, $\hat{\rho}_x$, $\hat{\rho}_y$ and $\hat{\rho}_{xy}$ be:

$$\hat{\rho}_j = \frac{\widehat{\rho_j \sigma_j \tau_j}}{\sqrt{\hat{\sigma_j^2}}\sqrt{\hat{\tau_j^2}}} = \frac{\tilde{S}_{XY_j}}{\sqrt{\tilde{S}_{XX_j}\tilde{S}_{YY_j}}} \;\; ; \;\; j = 1, 2, \ldots m \tag{4}$$

$$\hat{\rho}_x = \frac{\sum_{j\neq k} \widehat{\rho_x \sigma_j \sigma_k}}{\sum_{j\neq k} \sqrt{\hat{\sigma}_j^2 \hat{\sigma}_k^2}} = \frac{\sum_{j\neq k} \tilde{S}_{XX_{jk}}}{\sum_{j\neq k} \sqrt{\tilde{S}_{XX_j}\tilde{S}_{XX_k}}} \tag{5}$$

$$\hat{\rho}_y = \frac{\sum_{j\neq k} \widehat{\rho_y \tau_j \tau_k}}{\sum_{j\neq k} \sqrt{\hat{\tau}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j\neq k} \tilde{S}_{YY_{jk}}}{\sum_{j\neq k} \sqrt{\tilde{S}_{YY_j}\tilde{S}_{YY_k}}} \tag{6}$$

$$\hat{\rho}_{xy} = \frac{\sum_{j\neq k} \widehat{\rho_{xy} \sigma_j \tau_k}}{\sum_{j\neq k} \sqrt{\hat{\sigma}_j^2 \hat{\tau}_k^2}} = \frac{\sum_{j\neq k} \tilde{S}_{XY_{jk}}}{\sum_{j\neq k} \sqrt{\tilde{S}_{XY_j}\tilde{S}_{XY_k}}} \tag{7}$$

define θ_{jj} as

$$\theta_{jj} = \frac{1}{n-3} \tag{8}$$

while define θ_{jk} as

$$\theta_{jk} = \frac{1}{n-3} \frac{\frac{1}{2}\hat{\rho}_j\hat{\rho}_k(\hat{\rho}_x^2 + \hat{\rho}_y^2) + \hat{\rho}_{xy}^2(1 + \hat{\rho}_j\hat{\rho}_k) + \hat{\rho}_x\hat{\rho}_y - \hat{\rho}_{xy}(\hat{\rho}_j + \hat{\rho}_k)(\hat{\rho}_x + \hat{\rho}_y)}{(1 - \hat{\rho}_j^2)(1 - \hat{\rho}_k^2)} \tag{9}$$

which is a function of not only $\hat{\rho}_j$ and $\hat{\rho}_k$, but also $\hat{\rho}_x$, $\hat{\rho}_y$, and $\hat{\rho}_{xy}$.

Let $\hat{\Sigma}_z$ be an $m \times m$ matrix with diagonal element (j, j) equal to θ_{jj} as given by Equation (8), and off-diagonal element (j, k) equal to θ_{jk}, as given by Equation (9). Also let L be an $(m-1) \times m$ contrast matrix for the pairwise differences, i.e.

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

To construct a posterior predictive p-value to test the null model $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m$, the test statistic $T_z(D) = (L\hat{\mathbf{z}})^T(L\hat{\Sigma}_z L^T)^{-1}(L\hat{\mathbf{z}})$ is chosen. T_z is almost identical to Wald test statistic χ^2_z. As reviewed in Meng (1994) and Gelman et al. (1996), the posterior predictive p-value that incorporates a classic test statistic is defined as $p_b(D) = P_A[T(D^{rep}) \geq T(D)|H_0, D]$, where $P_A(D^{rep}|H_0, D) = \int P_A(D^{rep}|H_0, \theta)P(\theta|H_0, D)d\theta$.

2.4 Computational details

Denote the parameters $\{\mu_{x_1}, \mu_{y_1}, \mu_{x_2}, \mu_{y_2}, \cdots, \mu_{x_m}, \mu_{y_m}, \sigma_1^{-2}, \tau_1^{-2}, \cdots, \sigma_m^{-2}, \tau_m^{-2}, \rho_0, \rho_x, \rho_y, \rho_{xy}\}$ as ϕ. Given a set of posterior draws of parameters by Metropolis-Hastings(MH) Algorithm within Gibbs sampling, ϕ^j, $j = 1, \cdots, J$, perform the following two steps for each j:

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

2. Calculate $T_z(D)$ and $T_z(D^{repj})$.

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

$\sum_{j=1}^{J} \mathbf{1}[T_z(D^{rep,j}) > T_z(D)]/J$.

3. Application of Methods

3.1 Simulation Study

We now examine the performance of the proposed tests 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. We selected autoregressive nuisance correlations: $\rho_{x_{jk}} = 0.5^{|j-k|}$, $\rho_{y_{jk}} = 0.6^{|j-k|}$, $\rho_{xy_{jk}} = 0.51 \times 0.7^{|j-k|}$ for nuisance correlation parameters. Note that these nuisance correlations are not constant. This setting has introduced some degree of model misspecification. With regard to the correlation parameters of interest, $\{\rho_1, \rho_2, \ldots, \rho_m\}$, a simulation setting was defined by two quantities, $\rho_{min} \in \{0.2, 0.5\}$ 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 (1) and (2). We considered sample sizes of $n \in \{25, 50\}$ and number of time points $m \in \{2, 3, 4, 5\}$.

For each scenario, we simulated 1,000 datasets and run 2000 iterations each. Metropolis-Hastings within Gibbs sampling was used. The steps below are followed:

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

2. Draw each of the components in S one at a time by performing an independent MH by choosing a Gamma distribution as the proposal density for an individual variance parameter. For example, 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 just include $\{\rho_0, \rho_x, \rho_y, \rho_{xy}\}$. While sampling ρ_x and ρ_y, for each 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 were observed between ρ_0 and ρ_{xy}. An extra step before updating the n'th sample with the proposal sample is to check if the proposal sample keeps R matrix to be positive definite. If the condition is not satisfied, the n'th sample takes the value of the n-1'th sample. The proposal variance was tuned every 25 iterations during the burn-in period for truncated normal proposal density to get an acceptance rate of between 30% to 40%. The first 400 samples in burn-in period were discarded. Histograms and summary statistics including mean and 95 % credible interval were obtained based on the rest 1600 samples. Having obtained $T(D)$, $T(D^{repj})$, $j = 1, \cdots, J$, histograms were made to graphically assess the proportion of the J pairs for which $T(D^{repj})$ exceeds $T(D)$. 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.

3.2 Evaluation of performance

Table 1 presents the empirical size, e.g. $\Delta = 0$, of posterior predictive test T_z and Wald Z-test for various combinations of n, m, and ρ_{min}. The first column under each sample size shows the empirical size of posterior predictive test T_z based on 1,000 simulations, while the second column is the empirical size of 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 number of rejections in 1,000 and 5,000 simulations for a nominal test would lie in the interval $(3.65, 6.35)$ and $(4.5, 5.6)$ respectively. Therefore, the Bayesian test has nominal size, regardless of the number of time points and the value of ρ_{min}. As the sample size drops to $n = 25$, posterior predictive test T_z became conservative, although the conservativeness was allowed due to Monte Carlo error. Overall, even if some degree of model mis-specification is present, the size of posterior predictive test T_z still remains nominal. Based on this finding, the new test can be applied to real data that do not have perfectly constant nuisance correlations.

Regarding the empirical power of the Bayesian posterior predictive test, Table 2 shows the simulation results comparing posterior predictive tests T_z 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. Both tests have similar power, but when n goes down, the power of posterior predictive tests drops.

3.3 Motivating example

A longitudinal periodontal study conducted by Kinney et al. (2011) and Ramseier et al. (2009) was analyzed in this section. 79 subjects completed the 12-month study, with samples of saliva-derived biomarkers (TNF-α, metalloproteinase(MMP)-8) and periodontal plaque biofilm pathogens (*P.gingivals*, *T.forsythia*) examined at baseline (Day0), 6 months and 12 months.

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

Table 1: Size of posterior predictive test T_z and Wald Z test with $n = 50, 25$, $m = \{2, 3, 4, 5\}$ at $\rho_{min} = \{0.2, 0.5\}$. T_z=posterior predictive test using test statistic χ^2_z; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation

		$n = 50$		$n = 25$	
m	ρ_{min}	T_z	χ^2_z	T_z	χ^2_z
2	0.2	4.7	5.4	4.4	5.3
	0.5	4.8	5.1	4.7	5.2
3	0.2	5.2	5.2	4.7	5.2
	0.5	5.3	5.1	4.3	5.1
4	0.2	4.9	5.0	4.0	5.0
	0.5	4.2	5.1	4.0	4.8
5	0.2	4.1	5.1	5.6	5.4
	0.5	4.5	5.1	4.7	5.2

Table 2: Power of posterior predictive test T_z and Wald Z test with $n = 50, 25$, $m = \{2, 3, 5\}$ at $\rho_{min} = \{0.2, 0.5\}$. T_z=posterior predictive test using test statistic χ^2_z; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation

		$n = 50$		$n = 25$	
m	ρ_{min}	T_z	χ^2_z	T_z	χ^2_z
2	0.2	50.8	50.7	27.9	28.7
	0.5	88.6	90.4	61.8	61.0
3	0.2	34.4	37.0	19.9	20.0
	0.5	75.5	75.3	42.7	44.4
5	0.2	33.3	32.2	15.8	16.8
	0.5	68.8	66.0	33.9	34.6

Table 3: Sample serial correlations between combinations of biomarkers and pathogens, and testing equality of serial correlations at 0, 6,12 months. T_z=posterior predictive test using test statistic χ^2_z ; χ^2_z=Asymptotic Wald Test based on Fisher's Z-transformation.

Pathogen	Biomarker	Sample serial correlation			p-value	
		0	6	12	T_z	χ^2_z
P.gingivals	TNF-α	-0.17	0.07	-0.07	0.154	0.107
	MMP-8	-0.01	0.28	0.04	0.08	0.059
T.forsythia	TNF-α	-0.19	0.07	-0.19	0.014	0.005
	MMP-8	0	0.22	0.13	0.192	0.182

a log-transformation. Shown in Table 3 is the sample serial correlations for each pair of biomarker and pathogen after the transformation described above. To test the hypothesis whether the correlation between a biomarker and a pathogen at each time point is equal, both of posterior predictive test, T_z, and Wald Z-test were performed. Table 3 summarizes the p-values. Both tests gave comparable results and most pairs have homogeneous serial correlations. It is shown that heterogeneity exists in pairs between TNF-α and *T.forsythia* (max-min difference is 0.22) and MMP-8 and *P.gingivals* (max-min difference is 0.29, p-value close to 0.05).

4. Conclusions and Discussion

This report described a Bayesian approach to perform tests of equality of correlation coefficients for longitudinal studies. We borrowed the classical Wald χ^2_z statistic 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 computed. Bayesian Test using posterior predictive p-value maintains nominal size, and the assumption of equal nuisance correlations (ρ_x, ρ_y and ρ_{xy}) in Σ is generally robust to data without constant nuisance correlation. Posterior predictive tests T_z has similar power as Wald χ^2_z.

We should be aware that Bayesian and Frequentist p-value are two completely different concepts. The Wald test Type I error was listed just for reference, but the p-value should not be compared to the posterior predictive p-value. This report is to propose a new approach to perform tests of equality of correlation coefficients for longitudinal studies from a very different perspective.

REFERENCES

Barnard, J., McCulloch, R. and Meng, X. (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10(4),1281-311.

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

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

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

Kinney, J. S., Morelli, T., Braun, T., Ramseier, C. A., Herr,A. E., Sugai, J. V., Shelburne, C. E., Rayburn, L. A., Singh, A. K., and Giannobile,W. V. (2011), "Saliva/pathogen biomarker signatures and periodontal disease progression," *Journal of Dental Research*, 90, 752–8.

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

Olkin, I., and Siotani, M. (1976), "Asymptotic distribution of functions of a correlation matrix," *Essays in probability and statistics*, 235–251.

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

Ramseier, C. A., Kinney, J. S., Herr, A. E., Braun, T. M, Sugai, J. V., Shelburne, C. A., Rayburn, L. A., Tran, H. M., Singh, A. K., and Giannobile, W. V. (2010), "Identification of pathogen and host-response markers correlated with periodontal disease," *Journal of Periodontol*, 80, 436–46.

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