METHOD FOR ASSESSING CORRELATION COEFFICIENTS AMONG MULTIPLE REPEATEDLY MEASURED BIOMARKERS

by

Kwonho Jeong

B.A., SungKyunKwan University, Korea, 1999

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis is submitted

by

Kwonho Jeong

on

July 30, 2015

and approved by

Thesis Advisor:
Chung-Chou H. Chang, Ph.D.,
Professor
Department of Medicine
School of Medicine
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Jong-Hyeon Jeong, Ph.D.
Professor and Vice Chair
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Jonathan Yabes, Ph.D.
Assistant Professor
Department of Medicine
School of Medicine
University of Pittsburgh

METHOD FOR ASSESSING CORRELATION COEFFICIENTS AMONG MULTIPLE REPEATEDLY MEASURED BIOMARKERS

Kwonho Jeong, M.S.

University of Pittsburgh, 2015

ABSTRACT

Hypoglycemia and hyperglycemia in critically ill patients have been considered as being closely tied to mortality. However, clinical researchers and physicians have suspected that this relationship may be confounded by other time-updated biomarkers such as the acute physiology and the chronic health evaluation (APACHE) score which is a measurement of disease severity. To investigate the effect of dysglycemia on mortality while adjusting for time-dependent biomarkers, we first need to ensure that these time-updated biomarkers are associated with the main exposure variable, i.e., the time-dependent glucose levels.

Several researchers have proposed methods to estimate the true correlation coefficient between two repeatedly measured continuous variables using the maximum likelihood method via the mixed effects modeling with an assumption that these two variables are measured at the same time points. In this study, we further extend the methods proposed by these researchers, and proposed a method that can be used to examine the correlation coefficients among multiple (two or more) variables measured repeatedly. The method we proposed can be applied to linked cases where repeated measurements are linked over time and to unlinked cases where measurements are not necessarily measured concurrently.

The dataset we used for demonstration is the HighDensity ICU dataset, an electronic dataset from an eight-year observational cohort of more than 54,000 admissions recorded in twelve Intensive care units (ICUs) of a large tertiary care center. Several risk factors of mortality were recorded daily for the ICU admissions including glucose level, disease severity score, organ dysfunction scores, cumulative daily insulin doses, and caloric intake.

Simulation studies were conducted to examine the empirical features of our proposed method under different underlying scenarios. In application, we compared our method with the methods proposed by previous researchers on correlation coefficients between any two risk factors.

Public Health Significance: One of the objectives of the current on-going study is to find the relationship between the glucose level and the APACHE III score, both measured repeatedly, while controlling for other repeatedly measured biomarkers. Findings in this study not only can provide helpful information for physicians in the ICUs to optimize treatments in critically ill patients with septic shock, but also can provide biostatisticians a better statistical tool to estimate correlation between two longitudinally measured biomarkers that are adjusted for within-subject correlation while controlling for repeatedly measured confounding factors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

There are many situations that require us to estimate the correlation coecient between biomarkers or factors that are commonly measured repeatedly, since the relationship between them provides new perspectives on disease progression in clinical studies or certain phenomena in social studies.

Under the repeated measures design, each patient will have multiple measurements on a number of markers. The observations taken from the same patient will no longer be independent and this dependency should be incorporated into the subsequent analyses. For example, researchers may want to assess the association between blood pressures and heart rates that are repeatedly measured. Although the Pearson correlation coecient or Spearman correlation coecient are currently the most-used statistical tool for describing the linear relationship among biomarkers that are commonly measured on continuous or ordinal scales, they are not applicable when dealing with repeated measurements.

The interest in estimating correlation coecients in multiple measurements has been raised in the field of biostatistics, and several ad hoc approaches have been developed. Bland and Altman (1995a and 1995b) published two articles on the issue. Their approach consisted of two parts. First, they suggested using the analysis of covariance (ANCOVA) method if a within-subject correlation between two variables is of interest. They used an example of gastric intramural pH and partial pressure of carbon dioxide in the arterial blood (PaCO2), collected

from 8 patients in order to investigate whether an increase in the pH value within a patient was associated with that in PaCO2 (Table 4.1). In their method, two variables of interest were assigned to either the outcome variable or the predictor variable in an ANCOVA model. In addition, they included subject dummy variables as other predictors in the model. The within-subject correlation between the two repeatedly measured variables will then be estimated by the partial correlation coecient of the ANCOVA model. When those subjects in a study are many, there will be a loss of power caused by the increase of parameters related to the subject-specific dummy variables that are needed to be estimated. In their second paper (Bland and Altman [1995b]), they described how to analyze such data if the interest lies in a correlation between subjects, for example, they were interested in investigating whether subjects with high values of pH also tend to have high values of PaCO2. They proposed a weighted correlation coefficient calculated from the mean values for each subject and used the number of repeated measurements within each subject as a weight. It has been found that this method tends to underestimate the true between-subject correlation and that it can only be applied when the number of repeated measurements per variable is the same for a given subject.

Several researchers have proposed more advanced approaches to estimate the true correlation coecient between two variables measured repeatedly. Lam et al. (1999) used the maximum likelihood method to measure it via the mixed eects modeling approach with an assumption that two variables are measured at the same time points. Unfortunately, their method requires a specialized software. Hamlett et al. (2003) later generalized Lam el al.s approach by utilizing the procedure *Proc Mixed* of the SAS software in which the correlation structure of the repeated measurements can be accounted for.

Roy (2006) extended the concept of Hamlett et al. to establish a general framework to estimate the correlation coecient between two variables measured repeatedly by having the choice of dierent correlation structure within a subject. He demonstrated the use of compound symmetry (CS) and the first order autoregressive (AR[1]) correlation structures and extended the model by including random intercept and random slope (of time) terms for each subject.

With all the eorts thus far, the methods of estimating correlation coecients have not been

yet extended to more than two repeatedly measured variables. In this study, we further extend the methods proposed by Hamlett et al. (2003) and Roy (2006) and propose a method that can be used to examine the correlation coecients among multiple (two or more) variables measured repeatedly.

Moreover, we also explore the method under two separate situations: data with linked case and data with unlinked case. In a linked case, repeated measurements have to be measured at the same time points for all subjects, while in a unlinked case, repeated measurements are not necessarily measured at the same time points for all subjects.

We applied our method to estimate correlation coecients among multiple repeatedly measured biomarkers in the HighDensity ICU dataset, which is a rich electronic dataset from an eight-year observational cohort of more than 54,000 admissions recorded in twelve intensity care units (ICUs) of a large tertiary care center.

Hypoglycemia and hyperglycemia in critically ill patients have been considered to be closely tied to mortality. However, clinical researchers have suspected that this relationship may be confounded by other time-updated biomarkers such as the acute physiology, the chronic health evaluation (APACHE) score which is a measurement of disease severity. To investigate the eect of dysglycemia on mortality while adjusting for time-dependent biomarkers, we first need to assure that the time-updated biomarkers are associated with the main exposure variable, i.e., the time-dependent glucose levels. Our goal was to estimate correlation coecients among three risk factors of mortality with continuous type: glucose level, APACHE III score, and the level of vasopressor.

Simulation studies were conducted to evaluate the empirical features of our proposed method under dierent underlying scenarios. We compared the performance of our method with the methods proposed by Hamlett et al. and Roy if only two repeatedly measured variables were considered.

Chapter 2

THE PROPOSED METHODS

In this chapter, methods for estimating correlation coefficients of multiple continuous biomarkers will be constructed under both linked and unlinked cases. In a linked case, the repeated measurements are measured at the same time points for all subjects. In an unlinked case, repeated measurements may not be measured concurrently. Therefore only one variable can be observed in some subjects.

2.1 For the Linked Case

Let $\boldsymbol{y_{it}} = (y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(K)})'$ be a $K \times 1$ vector of measurements for K different markers on the i-th patient at the t-th time point and $\boldsymbol{y}_i$ be the repeated responses of interest for the i-th subject by stacking initial responses at the first time point and then the responses at the second time point, and so on, $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, p$.

We consider the situation in which each subject may experience at most p consecutive measurements over K continuous variables and the number of repeated measurements of the variables for each patient may not be equal. Suppose, for the i-th subject each variable is measured on m_i number of time points. For subject i, $\boldsymbol{y_i}$ is ($n_i \times 1$)-dimensional, $1 \leq n_i \leq Kp$, where $n_i = Km_i$. We assume that $\boldsymbol{y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and with a maximum of $Kp \times Kp$ positive definite variance-covariance matrix $\boldsymbol{\Omega}$. The

$K \times K$ block diagonal matrix in $\boldsymbol{\Omega}$ gives the covariance matrix between the K variables.

For the i-th patient, we consider the linear mixed model that satisfies the standard notation of Laird and Ware (1982) given by:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \ , \ i = 1, 2, \ldots, N$$

where $\boldsymbol{X}_i$ is a ($n_i \times l$)-dimensional design matrix for the fixed effects, $\boldsymbol{\beta}$ is a l-dimensional vector of fixed but unknown parameters, $\boldsymbol{Z}_i$ is a ($n_i \times m$)-dimensional design matrix for the random effects, $\boldsymbol{b}_i$ is a m-dimensional vector of random effects, and $\boldsymbol{\epsilon}_i$ is a n_i-dimensional vector of random errors. In addition, the random effects vector $\boldsymbol{b}_i$ is assumed to be normally distributed with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{D}$ and the errors are also normally distributed with expectation $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{R}_i$. Vectors $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are assumed to be independent.

It can be shown that, under the assumption that the mixed linear model is correctly specified, the marginal density function of $\boldsymbol{y}_i$ is n_i-dimensional normal distribution with mean $E(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, and with variance-covariance matrix $Cov(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z'}_i + \boldsymbol{R}_i$. While $\boldsymbol{R}_i$ here represents the partial variance-covariance matrix corresponding to the i-th individual, the $K \times K$ block diagonal of this gives the partial variance-covariance matrix among the K variables.

However, when the number of subjects N is less than Kp, the maximum likelihood estimation of unstructured $\boldsymbol{R}_i$ is not achievable and we need an additional assumption of Kronecker product structure on $\boldsymbol{R}_i$. It is also parsimonious to assume a Kronecker product structure on $\boldsymbol{R}_i$. We thus assume $\boldsymbol{R}_i = \dim_{n_i}(\boldsymbol{V} \otimes \boldsymbol{\Sigma})$, where $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ respectively are $p \times p$ and $K \times K$ dimensional positive definite matrices and $\otimes$ denotes the Kronecker product. The notation $\dim_{n_i}(\boldsymbol{V} \otimes \boldsymbol{\Sigma})$, represents a $n_i \times n_i$ dimensional submatrix obtained from a $Kp \times Kp$ dimensional matrix $(\boldsymbol{V} \otimes \boldsymbol{\Sigma})$, by appropriately keeping the columns and rows corresponding to the n_i dimensional response vector $\boldsymbol{y}$.

The matrix $\boldsymbol{\Sigma}$ represents the variance-covariance matrix among the response variables at a given time point. It is assumed that $\boldsymbol{\Sigma}$ does not depend on a particular time point and is the same for all time points. The correlation matrix $\boldsymbol{V}$ of the repeated measures on a given

response variable is assumed to be the same for both response variables. For data collected from a longitudinal study, measurements taken for each subject are usually correlated and such correlation should be taken into account in order to produce a valid inference. Misspecification of the correlation structure or incorrect assumption on the correlation structure may have serious impacts on the valid inference of the data (Roy and Khattree, 2006). Various error structures may be assumed for repeated measures correlations. Often, this correlation matrix is assumed to have a compound symmetric (CS) or autoregressive of order one (AR[1]) structure. The compound symmetry correlation structure assumes equal correlation among all the repeated measurements. Typically, the correlation between successive observations decays fairly rapidly as the time points become more widely separated. Modeling of these serial effects can be implemented under the assumption of AR[1] correlation structure.

If $\boldsymbol{V}$ has a compound symmetric structure, it can be rewritten as the form $\boldsymbol{V} = (1-\varrho)I_p + \varrho 1_p 1_p'$, where I_p represents the $p \times p$ identity matrix and 1_p is a $p \times p$ vector containing all elements as unity. For $\boldsymbol{V}$ to be positive definite, we require $-\dfrac{1}{p-1} < \varrho < 1$. If $\boldsymbol{V}$ has the AR[1] structure, then:

$$\boldsymbol{V} = \begin{bmatrix} 1 & \varrho & \varrho^2 & \dots & \varrho^{p-1} \\ \varrho & 1 & \varrho & \dots & \varrho^{p-2} \\ \varrho^2 & \varrho & 1 & \dots & \varrho^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varrho^{p-1} & \varrho^{p-2} & \varrho^{p-3} & \dots & 1 \end{bmatrix}.$$

The variance-covariance matrix of the K variables at a single time point is denoted by:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1K} \\ & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2K} \\ & & \sigma_3^2 & \dots & \sigma_{3K} \\ & & & \ddots & \vdots \\ & & & & \sigma_K^2 \end{bmatrix},$$

where σ_k^2 is the variances of each variable $y^{(k)}$, and $\sigma_{kk'}$ is the covariance between $y^{(k)}$ and $y^{(k')}$. Therefore, $\tilde{\varrho}_{kk'} = \dfrac{\sigma_{kk'}}{\sigma_k \sigma_{k'}}$ is the partial correlation coefficient between two variables at a single time point after controlling for the effects of other variables and is one of our parameters of interest. Note that the number of random effects and the form of Z_i can be chosen to fit the observed covariance matrix for the i-th individual as:

$$\text{Cov } (\boldsymbol{y}_i) = \boldsymbol{\Omega}_i = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z'}_i + \dim_{n_i}(\boldsymbol{V} \otimes \boldsymbol{\Sigma}).$$

Thus, the covariance matrix has the same structure for each subject, except that of the dimension. Suppose that subject 1 has four replicates ($m_i = 4$) on three variables ($K = 3$) and that $3m_i = 12$. Then, the covariance matrix for subject 1 will have 12×12 dimension. If subject 2 has the maximum number (saying $p = K$) of repeated measures, the covariance matrix for subject 2 will have $3K \times 3K$ dimension. We assume that correlation between measurements taken at two different time points, t and t' where $t \neq t'$, are given by:

$$\begin{aligned}
Corr(y_{it}^{(k)}, y_{it'}^{(k)}) &= \varrho_k \\
Corr(y_{it}^{(k')}, y_{it'}^{(k')}) &= \varrho_{k'} \\
Corr(y_{it}^{(k)}, y_{it}^{(k')}) &= \varrho_{kk'} \\
Corr(y_{it}^{(k)}, y_{it'}^{(k')}) &= \delta \varrho_{kk'}
\end{aligned}$$

The main advantage of *Proc Mixed* of SAS is that it can analyze and calculate $\boldsymbol{\Omega}_i$ and $\boldsymbol{R}_i$, $n_i \times n_i$ dimensional variance-covariance matrices for each individual i. The purpose of the present paper is to extend the methods of Lam et al. (1999) and Roy (2006) to multiple (three or more) variables measured repeatedly by using the correlation structure on the repeated measures of each variable that is present in the data. We will show that our model gives a very similar result after controlling for other variables.

We use *Proc Mixed* to get the maximum likelihood estimates of $\boldsymbol{\beta}$, $\boldsymbol{D}$, $\boldsymbol{R}_i$ and $\boldsymbol{\Omega}_i$. Random and repeated statements specify the structure of the covariance matrices $\boldsymbol{D}$ and $\boldsymbol{R}_i$. Another advantage of *Proc Mixed* is that it can handle the Kronecker product structure of the variance-covariance matrix $\boldsymbol{R}_i = \dim_{n_i}(\boldsymbol{V} \otimes \boldsymbol{\Sigma})$ and calculate $n_i \times n_i$ dimensional submatrix $\boldsymbol{R}_i$, from a $Kp \times Kp$ dimensional matrix $(\boldsymbol{V} \otimes \boldsymbol{\Sigma})$. Currently, *Proc Mixed* can only have option $\boldsymbol{\Sigma}$ as

unstructured and $\boldsymbol{V}$ as unstructured, AR[1] or CS structure. Besides possible random effects, choosing $\boldsymbol{V}$ as AR[1] or CS leads to a parsimonious covariance structure and has the potential to provide a correct representation of the correlation structure on the repeated measurements for each variable that is often present in the data. SAS estimates $\boldsymbol{\Sigma}$, the same $K \times K$ positive definite matrix for all individuals and the same correlation coefficient ϱ, corresponding to the correlation matrix $\boldsymbol{V}$ for all individuals. As noted above, *Proc Mixed* is used to estimate the variance covariance matrices $\boldsymbol{\Omega}_i$, $\boldsymbol{\Sigma}$, and the correlation coefficient ϱ for the correlation structure $\boldsymbol{V}$ on the repeated measures.

2.2 For the Unlinked Case

A model for the unlinked repeated-measure design is easily obtained through a simple revision of the model for the linked case. The structural difference between the linked and the unlinked case is that $y_{it}^{(1)}, y_{it}^{(2)}, \ldots, y_{it}^{(K)}$ are no longer linked together. That is, there is no "time" effect in the problem and we have the same correlation between any two of $y^{(1)}, y^{(2)}, \ldots, y^{(K)}$. The correlation structure is thus:

$$Corr(y_{it}^{(j)}, y_{it'}^{(j)}) = \varrho_{y^{(j)}}$$
$$Corr(y_{it}^{(j')}, y_{it'}^{(j')}) = \varrho_{y^{(j')}}$$
$$Corr(y_{it}^{(j)}, y_{it'}^{(j')}) = \varrho_{y^{(j)}y^{(j')}} = Corr(y_{it}^{(j)}, y_{it}^{(j')})$$

One can think of the unlinked case as a special case of the linked setting, with δ set equal to 1. The covariance matrix for the i-th subject in this setting is given by:

$$
cov\begin{pmatrix} y_{i1}^{(1)} \\ y_{i1}^{(2)} \\ y_{i1}^{(3)} \\ \hdashline y_{i2}^{(1)} \\ y_{i2}^{(2)} \\ y_{i2}^{(3)} \\ \hdashline \vdots \\ \hdashline y_{im_i}^{(1)} \\ y_{im_i}^{(2)} \\ y_{im_i}^{(3)} \end{pmatrix} = \left[\begin{array}{cccccccccc} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} & \dots & \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} & \dots & \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 & \dots & \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 \\ \hdashline \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} & \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} & \sigma_{12} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 & \sigma_{13} & \sigma_{23} & \sigma_3^2 & \dots & \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 \\ \hdashline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \hdashline \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} & \sigma_1^2\varrho_1 & \sigma_{12} & \sigma_{13} & \dots & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} & \sigma_{12} & \sigma_2^2\varrho_2 & \sigma_{23} & \dots & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 & \sigma_{13} & \sigma_{23} & \sigma_3^2\varrho_3 & \dots & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{array}\right].
$$

It is important to note here that there are no terms involved in δ in this unlinked version of the covariance matrix and that the blocks on the off-diagonal are now constant.

Chapter 3

SIMULATION

In this chapter, we illustrate the finite-sample performance of the proposed method. We first defined the mean vector μ and the variance-covariance matrix Σ. Having the form:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} & \sigma_1\sigma_3\varrho_{13} \\ & \sigma_2^2 & \sigma_2\sigma_3\varrho_{23} \\ & & \sigma_3^2 \end{bmatrix}$$

$$= \begin{bmatrix} 6.02 & \sqrt{6.02}\sqrt{10.11}(-0.36) & \sqrt{6.02}\sqrt{13.87}(-0.17) \\ & 10.11 & \sqrt{10.11}\sqrt{13.87}(0.13) \\ & & 13.87 \end{bmatrix} = \begin{bmatrix} 6.02 & -2.81 & -1.57 \\ & 10.11 & 1.55 \\ & & 13.87 \end{bmatrix}.$$

Then, we created the random variable that follows multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

$$\begin{bmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ y_{it}^{(3)} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} & \sigma_1\sigma_3\varrho_{13} \\ & \sigma_2^2 & \sigma_2\sigma_3\varrho_{23} \\ & & \sigma_3^2 \end{bmatrix} \right)$$

In addition, we assume compound symmetry correlation matrix with setting not only the correlation coefficient between two observations at different time points on the same variable,

calling it ϱ_1 , ϱ_2, and ϱ_3 but also the kind of abrasion rates between two different variables that measures at different time points, calling it δ_{12} , δ_{13}, and δ_{23}. Thus, the full true variance-covariance matrix $\boldsymbol{\Sigma}$ of the three variables is set to:

$$\begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} & \sigma_1\sigma_3\varrho_{13} & \sigma_1^2(0.7) & \sigma_1\sigma_2\varrho_{12}(0.3) & \sigma_1\sigma_3\varrho_{13}(0.55) \\ \sigma_1\sigma_2\varrho_{12} & \sigma_2^2 & \sigma_2\sigma_3\varrho_{23} & \sigma_1\sigma_2\varrho_{12}\delta_{12} & \sigma_2^2(0.4) & \sigma_2\sigma_3\varrho_{23}(0.75) \\ \sigma_1\sigma_3\varrho_{13} & \sigma_2\sigma_3\varrho_{23} & \sigma_3^2 & \sigma_1\sigma_3\varrho_{13}\delta_{13} & \sigma_2\sigma_3\varrho_{23}\delta_{23} & \sigma_3^2(0.25) \\ \sigma_1^2\varrho_1 & \sigma_1\sigma_2\varrho_{12}\delta_{12} & \sigma_1\sigma_3\varrho_{13}\delta_{13} & \sigma_1^2 & \sigma_1\sigma_2\varrho_{12} & \sigma_1\sigma_3\varrho_{13} \\ \sigma_1\sigma_2\varrho_{12}\delta_{12} & \sigma_2^2\varrho_2 & \sigma_2\sigma_3\varrho_{23}\delta_{23} & \sigma_1\sigma_2\varrho_{12} & \sigma_2^2 & \sigma_2\sigma_3\varrho_{23} \\ \sigma_1\sigma_3\varrho_{13}\delta_{13} & \sigma_2\sigma_3\varrho_{23}\delta_{23} & \sigma_3^2\varrho_3 & \sigma_1\sigma_3\varrho_{13} & \sigma_2\sigma_3\varrho_{23} & \sigma_3^2 \end{bmatrix}$$

$$= \begin{bmatrix} 6.02 & -2.81 & -1.57 & 6.02 \times 0.7 & -2.81 \times 0.3 & -1.57 \times 0.55 \\ & 10.11 & 1.55 & & 10.11 \times 0.4 & 1.55 \times 0.75 \\ & & 13.87 & & & 13.87 \times 0.25 \\ 6.02 \times 0.7 & -2.81 \times 0.3 & -1.57 \times 0.55 & 6.02 & -2.81 & -1.57 \\ & 10.11 \times 0.4 & 1.55 \times 0.75 & & 10.11 & 1.55 \\ & & 13.87 \times 0.25 & & & 13.87 \end{bmatrix}$$

The true corelation matrix of our interest is set to:

$$\boldsymbol{R} = \begin{bmatrix} 1 & \varrho_{12} & \varrho_{13} & \varrho_1 & \varrho_{12}\delta_{12} & \varrho_{13}\delta_{13} \\ \varrho_{12} & 1 & \varrho_{23} & \varrho_{12}\delta_{12} & \varrho_2 & \varrho_{23}\delta_{23} \\ \varrho_{13} & \varrho_{23} & 1 & \varrho_{13}\delta_{13} & \varrho_{23}\delta_{23} & \varrho_3 \\ \varrho_1 & \varrho_{12}\delta_{12} & \varrho_{13}\delta_{13} & 1 & \varrho_{12} & \varrho_{13} \\ \varrho_{12}\delta_{12} & \varrho_2 & \varrho_{23}\delta_{23} & \varrho_{12} & 1 & \varrho_{23} \\ \varrho_{13}\delta_{13} & \varrho_{23}\delta_{23} & \varrho_3 & \varrho_{13} & \varrho_{23} & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & -0.36 & -0.17 & 0.7 & -0.36 \times 0.3 & -0.17 \times 0.55 \\ & 1 & 0.13 & & 0.4 & 0.13 \times 0.75 \\ & & 1 & & & 0.25 \\ 0.7 & -0.36 \times 0.3 & -0.17 \times 0.55 & 1 & -0.36 & -0.17 \\ & 0.4 & 0.13 \times 0.75 & & 1 & 0.13 \\ & & 0.25 & & & 1 \end{bmatrix}$$

Table 3.1 presents the estimates, biases, coverage rates, and standard errors from a sample of 100 subjects, which is generated in each of the 100 simulated datasets. Table 3.2 shows the results from 300 subjects in 300 simulated datasets. Absolute and relative bias are sufficiently

tiny and most of estimates were included in the range of 2 times of standard deviation (2 x SD) when we have large sample (Table 3.2). All of the estimates are as follows:

Table 3.1: Simulation results for the estimates from 100 simulated dataset with 100 samples

Parameters	TRUE	Estimate	Abs. Bias	Rel. Bias	SE	Coverage rate
μ_1	6	5.94	0.06	1.01%	0.031	70%
μ_2	8	8.00	0.00	0.00%	0.029	70%
μ_3	10	10.01	0.01	0.13%	0.027	64%
σ_1^2	6.02	5.97	0.05	0.91%	0.087	64%
σ_2^2	10.1	10.02	0.09	0.90%	0.084	70%
σ_3^2	13.9	13.80	0.07	0.47%	0.118	66%
ϱ_1	0.7	0.69	0.01	0.77%	0.004	70%
ϱ_2	0.4	0.39	0.01	1.58%	0.006	64%
ϱ_3	0.25	0.26	0.01	4.30%	0.005	60%
ϱ_{12}	-0.36	-0.37	0.01	-3.56%	0.007	62%
ϱ_{13}	-0.17	-0.19	0.02	-8.27%	0.006	70%
ϱ_{23}	0.13	0.13	0.00	1.20%	0.006	72%
δ_{12}	0.3	0.29	0.01	1.87%	0.017	68%
δ_{13}	0.55	0.55	0.00	0.59%	0.019	74%
δ_{23}	0.75	0.69	0.06	8.85%	0.024	74%

Table 3.2: Simulation results for the estimates from 300 simulated dataset with 300 samples

Parameters	TRUE	Estimate	Abs. Bias	Rel. Bias	SE	Coverage rate
μ_1	6	6.00	0.00	0.04%	0.007	96%
μ_2	8	8.01	0.01	0.10%	0.007	96%
μ_3	10	10.00	0.00	0.02%	0.007	96%
σ_1^2	6.02	6.03	0.01	0.17%	0.021	97%
σ_2^2	10.1	10.08	0.03	0.26%	0.023	96%
σ_3^2	13.9	13.82	0.05	0.33%	0.026	97%
ϱ_1	0.7	0.70	0.00	0.11%	0.001	96%
ϱ_2	0.4	0.40	0.00	0.40%	0.001	98%
ϱ_3	0.25	0.25	0.00	1.30%	0.001	96%
ϱ_{12}	-0.36	-0.36	0.00	-0.98%	0.002	97%
ϱ_{13}	-0.17	-0.17	0.00	-1.38%	0.002	96%
ϱ_{23}	0.13	0.13	0.00	2.54%	0.001	96%
δ_{12}	0.3	0.30	0.00	0.06%	0.004	95%
δ_{13}	0.55	0.54	0.01	2.78%	0.006	96%
δ_{23}	0.75	0.76	0.01	0.79%	0.005	95%

Chapter 4

APPLICATION

In this chapter, we apply our proposed method to the linked and unlinked data from Roy (2006) and Hamlett et al. (2003). They assessed the correlation in the presence of replication only in two variables. But we add a third variable with hypothetical values and compare our results with theirs to see whether our method preserves the accuracy after controlling for the third variable. we also apply our method to the HighDensity ICU dataset.

4.1 The Linked Case (Intramural pH and PaCO2)

Both Hamlett et al (2003) and Roy (2006) used the same dataset to assess the correlation between intramural pH and PaCO2 from the critically ill patients, as shown in Table 4.1 and their results are given in Table 4.2. We created a new variable, called HCO3 in Table 4.1 so that we could compare the result after controlling for HCO3.

Figure 4.1 represents the profiles of three variables pH, PaCO2, and HCO3 for each individual. The behaviors of the three variables are very different for each individual and the intercepts and the slopes are significantly different for each individual in each variable. There is also a considerable difference in intercepts within the variable PaCO2. These observations lead us to fit the model with the intercept and group variable (i.e., pH, PaCO2, and HCO3) as random effects.

Our proposed method produces a result very similar to the previous studies except δ_{12}, which is a partial penalty factor between pH and PaCO2 measured at different time points.

Table 4.1: Repeated measurement of intramural pH and PaCO2 for 8 critically ill paients

Patient #	Repeat	pH	PaCO2	HCO3	Patient #	Repeat	pH	PaCO2	HCO3
1	1	6.68	3.97	4.48	5	3	7.30	4.32	6.85
1	2	6.53	4.12	4.99	5	4	7.37	3.23	6.75
1	3	6.43	4.09	5.39	5	5	7.27	4.46	6.37
1	4	6.33	3.97	6.09	5	6	7.28	4.72	7.44
2	1	6.85	5.27	4.49	5	7	7.32	4.75	7.50
2	2	7.06	5.37	4.55	5	8	7.32	4.99	9.51
2	3	7.13	5.41	5.66	6	1	7.38	4.78	4.63
2	4	7.17	5.44	7.68	6	2	7.30	4.73	5.91
3	1	7.40	5.67	5.31	6	3	7.29	5.12	6.82
3	2	7.42	3.64	6.20	6	4	7.33	4.93	5.49
3	3	7.41	4.32	6.79	6	5	7.31	5.03	7.84
3	4	7.37	4.73	6.39	6	6	7.33	4.93	8.35
3	5	7.34	4.96	7.29	7	1	6.86	6.85	3.93
3	6	7.35	5.04	6.95	7	2	6.94	6.44	4.53
3	7	7.28	5.22	9.53	7	3	6.92	6.52	5.25
3	8	7.30	4.82	8.91	8	1	7.19	5.28	6.27
3	9	7.34	5.07	9.36	8	2	7.29	4.56	6.22
4	1	7.36	5.67	6.82	8	3	7.21	4.34	6.24
4	2	7.33	5.10	6.48	8	4	7.25	4.32	6.47
4	3	7.29	5.53	7.50	8	5	7.20	4.41	6.18
4	4	7.30	4.75	5.97	8	6	7.19	3.69	7.37
4	5	7.35	5.51	7.49	8	7	6.77	6.09	6.95
5	1	7.35	4.28	6.19	8	8	6.82	5.58	8.94
5	2	7.30	4.44	6.43					

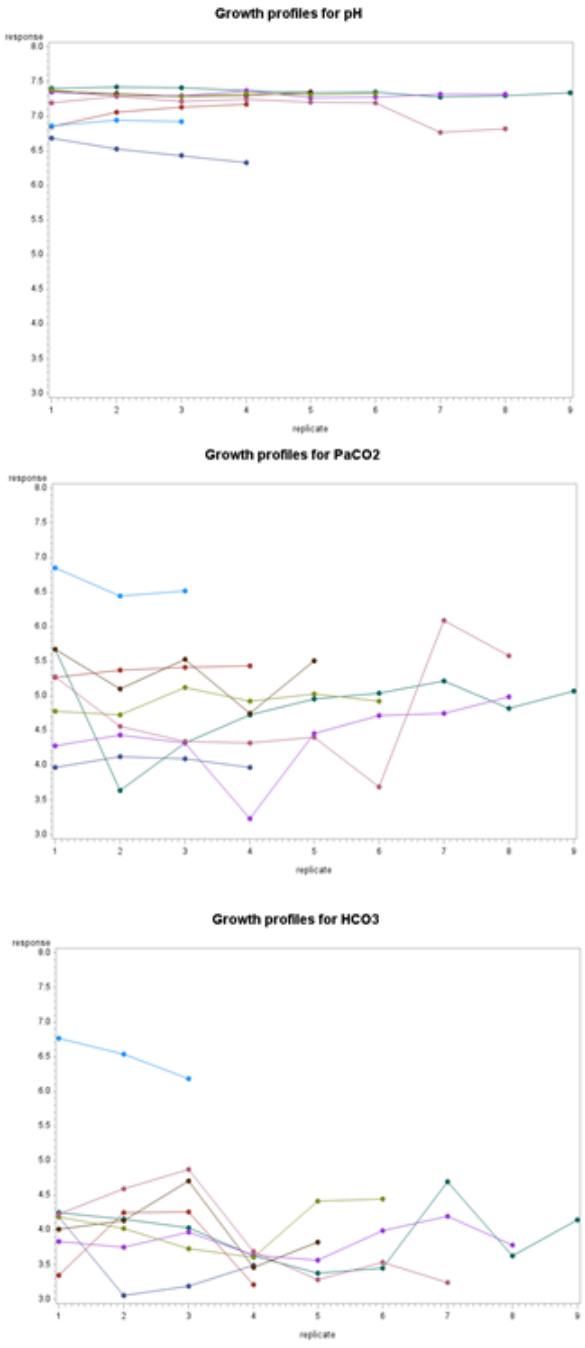

Figure 4.1: The profile plots of pH, PaCO2, and HCO3 for each individual

Table 4.2: Comparison of estimates for linked case

Parameters	Hamlett et al	Roy	Proposed
μ_1	7.1151	7.1177	7.1146
μ_2	5.0082	5.0333	5.0344
μ_3			4.1735
σ_1^2	0.0862	0.0873	0.0886
σ_2^2	0.6799	0.7697	0.7577
σ_3^2			0.8656
ϱ_1	0.8659	0.9453	0.9368
ϱ_2	0.6254	0.7385	0.7377
ϱ_3			0.8242
ϱ_{12}	-0.0100	0.0464	0.0092
ϱ_{13}			-0.0863
ϱ_{23}			0.6222
δ_{12}	-10.4724	2.3529	-0.1658
δ_{13}			1.4401
δ_{23}			0.9861

4.2 The Unlinked Case (Benzene Concentration in Indoor and Outdoor from Hamlett et al.)

To illustrate the theory with a real example, the benzene concentration dataset is selected from a study of Hamlett et al (2003). The dataset corresponds to ulniked repeated measurements of benzene concentration in indoor and ourdoor air measured on 25 Mexican families. The original dataset has only two variables (indoor vs. outdoor), but we add a third hypothetical variable, measurements in parking garage, so that we can see whether our model produces the same estimates as those in Hamlett et al (2003) after controlling for the third variable. The data with the three variables are given in Table 4.3 and the results are summarized in Table 4.4.

Roy (2006) did not cover the unlinked case in his study and we apply his method which implements the correlation structure on the repeated measurements for each patient that is present in the data. As a result, our proposed method preserves significant accuracy after controlling for third variable, benzene concentration in parking garage. Note that we cannot estimate δ terms in this unlinked case because there is no "time" effect in the unlinked case.

Table 4.3: Repeated measurements of benzene concentration (g/m3) in indoor (living room in this table), parking garage, and outdoor air taken at the homes of 35 Mexican families

Family	Benzene	Location	Family	Benzene	Location	Family	Benzene	Location	Family	Benzene	Location
1	12.11	LR	9	10.64	PG	19	8.61	Out	29	18.09	PG
1	12.1	Out	9	10.47	PG	19	4.09	PG	29	17.08	PG
1	14.08	PG	10	6.29	LR	19	8.56	PG	30	18.62	LR
2	7.25	LR	10	6.62	LR	21	3.82	Out	30	9.9	LR
2	13.63	Out	10	6.61	Out	22	6.2	LR	30	8.49	Out
2	9.78	PG	10	8.15	Out	22	7.03	LR	30	8.32	Out
3	6.43	Out	10	7.32	PG	22	6.86	Out	30	19.02	PG
4	27.12	LR	10	8.43	PG	22	9.03	PG	30	10.1	PG
4	41.08	Out	11	8.58	LR	22	9.6	PG	31	14.36	LR
4	30	PG	11	7.28	LR	23	4.26	Out	31	12.57	LR
5	5.93	LR	11	8.37	Out	24	3.84	LR	31	13.43	Out
5	18.05	Out	11	7.91	Out	24	3.41	LR	31	10.18	Out
5	7.14	PG	11	9.95	PG	24	5.49	PG	31	15.45	PG
6	6.92	LR	12	11.69	LR	24	4.35	PG	31	13.09	PG
6	7.56	LR	13	8.51	LR	25	3.12	Out	32	4.62	LR
6	7.19	Out	13	9.81	Out	25	2.76	Out	32	4.1	Out
6	7.64	Out	13	10.17	PG	26	7.25	LR	32	8.74	Out
6	7.1	PG	14	14.69	LR	26	8.91	LR	32	7.21	Out
6	9.38	PG	14	12.35	Out	26	9.35	Out	32	5.53	PG
7	10.08	LR	14	15.12	PG	26	8.27	Out	33	14.69	LR
7	9.35	LR	15	14.69	LR	26	7.66	PG	33	14.25	LR
7	8.58	Out	15	18.56	Out	26	9.22	PG	33	3.22	Out
7	8.68	Out	15	17.22	PG	27	4.77	LR	33	16.39	PG
7	11.76	PG	16	12.61	LR	27	4.79	LR	33	15.65	PG
7	10.53	PG	16	10.34	Out	27	6.7	PG	34	13.49	LR
8	9.33	LR	16	14.24	PG	27	7.12	PG	34	11.92	LR
8	8.98	LR	17	13.13	LR	28	11.94	LR	34	4.21	Out
8	7.6	Out	17	12.31	Out	28	10.74	LR	34	3.54	Out
8	7.59	Out	17	14.12	PG	28	8.62	Out	34	14.17	PG
8	9.37	PG	18	12.94	LR	28	8.19	Out	34	14.49	PG
8	10.81	PG	18	12.06	Out	28	12.48	PG	35	19.21	LR
9	9.26	LR	18	15.35	PG	28	11.22	PG	35	16.63	LR
9	9.77	LR	19	3.61	LR	29	15.4	LR	35	9.76	Out
9	9.33	Out	19	7.01	LR	29	15.21	LR	35	19.42	PG
9	9.28	Out	19	8.52	Out	29	13.19	Out	35	19.33	PG

Table 4.4: Comparison of estimates for unlinked case

Parameters	Hamlett et al	Roy	Proposed
μ_1	10.4563	10.4856	10.4770
μ_2	9.9691	9.9599	9.9722
μ_3			11.9356
σ_1^2	27.4675	23.3035	23.3855
σ_2^2	47.1361	44.4895	44.2744
σ_3^2			25.1548
ϱ_1	0.8916	0.8656	0.8677
ϱ_2	0.9730	0.9722	0.9719
ϱ_3			0.8582
ϱ_{12}	0.6143	0.6163	0.6601
ϱ_{13}			0.9958
ϱ_{23}			0.6924

4.3 HighDensity ICU Data

In this section, We assess correlation coefficients among three multiple repeatedly measured biomarkers (glucose level, APACHE III score, and the level of vasopressor) using the HighDensity ICU dataset from Clermont.

One of the main objectives in on-going study by Clermont is to identify the relationship between glucose level from patients in intensive care unit and their mortality. However, we must take into the consideration the association between glucose level, main predictor in his study, and other potential confounders such as APACHE III score and the level of vasopressor before we construct the appropriate model for mortality.

All of three biomarkers were measured as much as they needed in ICU and Clermont transformed glucose level into the new summary measure based on the calculation of area under the curve at each day during the first five days in ICU instead of using standard summary measures like average, maximum, or minimum. For demonstration, We use only 79,650 complete cases from 10,567 patients who received vosopressor treatment. The data from first 15 patients are shown in Table 4.5 and the scatter plots among three variables are shown in Figure 4.2.

For the purpose of comparison, the Pearson correlation coefficients between glucose and APACHE III score at each day are shown in Table 4.6. Although it is known that the blood glucose level increases with increasing APACHE score in critically ill patients (Krinsley, 2003), the Pearson correlation coefficients at each time point is very low with range of [0.06, 0.12] because we might use only the cases from the patients who suffered septic shock.

The results using linear mixed model in HighDensity ICU data are summarized in Table 4.7. Our proposed method has quite similar results for glucose and APACHE III score after controlling for the level of vasopressore but the correlations between glucose and APACHA III score are estimated as almost zero.

Table 4.5: Repeated measures of blood glucose, APACHE III score, and the level of vasopressor taken from first 20 patients with septic shock

	Glucose					Apache III					Vasopressor				
patient	Day1	Day2	Day3	Day4	Day5	Day1	Day2	Day3	Day4	Day5	Day1	Day2	Day3	Day4	Day5
6	161.0	127.1	106.1	115.8	129.8	120	97	81	75	84	102.3	99.8	99.8	99.8	99.8
12	197.0					97					3.2				
21	153.2					125					158.1				
27		122.5					91					10.0			
35	222.8					82					3.9				
39		117.6	128.1	138.6	149.1		108	102	75	49		18.5	70.9	63.0	63.0
48			170.1					63					3.7		
51	181.7					128					15.3				
55		91.1					44					3.5			
61	141.2	102.3				40	44				134.7	103.7			
66	141.6					92.0					2.8				
75		132.3	163.2				81.0	80.0				52.6	10.2		
81	182.2	267.2	228.3	112.5	247.6	170.0	155.0	95.0	98.0	117.0	373.8	476.8	274.3	151.8	58.3
89	295.8	223.6	246.2	301.2		112.0	94.0	122.0	162.0		95.5	30.6	171.8	320.0	
96	176.6	150.0	129.3	138.6	146.0	91.0	91.0	70.0	63.0	78.0	179.6	229.7	197.4	124.2	120.3

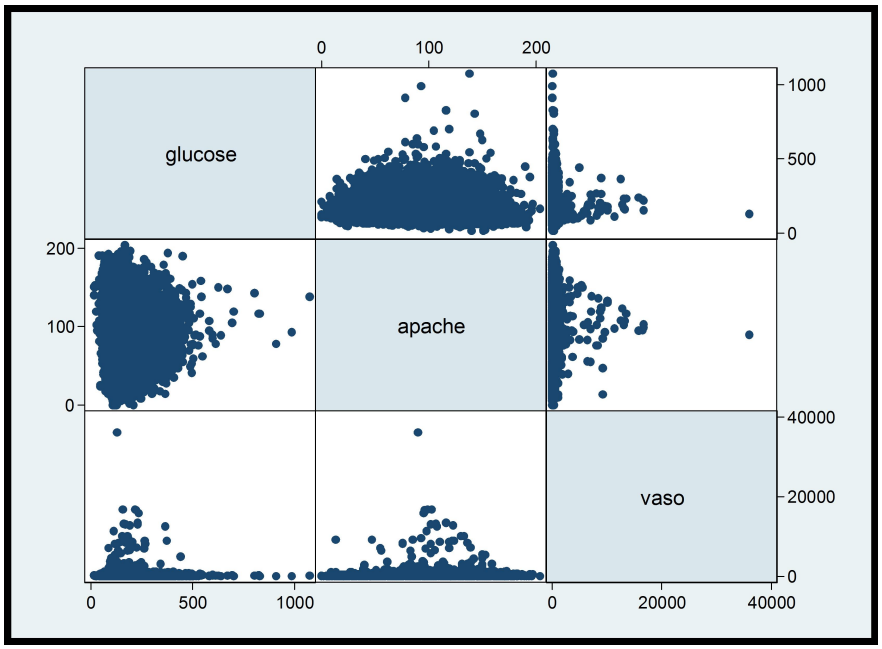

Figure 4.2: Scatter plot among glucose, APACHE III, and vasopressor

Table 4.6: Individual Pearson Correlation Coefficient Between Glucose and APACHA III Scores at Each Time Point

	glucose1	glucose2	glucose3	glucose4	glucose5	APACHE1	APACHE2	APACHE3	APACHE4	APACHE5
glucose1	1									
glucose2	0.6113	1								
glucose3	0.4064	0.6953	1							
glucose4	0.3811	0.5377	0.7734	1						
glucose5	0.3584	0.4689	0.6176	0.8044	1					
APACHE1	0.1046	0.1093	0.0441	0.0346	0.0315	1				
APACHE2	0.0933	0.1183	0.0606	0.0439	0.0424	0.9028	1			
APACHE3	0.0560	0.0860	0.0600	0.0539	0.0605	0.6891	0.8082	1		
APACHE4	0.0354	0.0673	0.0582	0.0733	0.0780	0.6003	0.6839	0.8606	1	
APACHE5	0.0218	0.0568	0.0516	0.0755	0.0858	0.5376	0.6146	0.7479	0.8810	1

Table 4.7: Comparison of estimates for HighDensity ICU Data

Parameters	Hamlett et al	Roy	Proposed
μ_1	157.60	157.65	157.63
μ_2	82.50	82.46	82.46
μ_3			103.14
σ_1^2	3,559.09	3,556.06	3,555.55
σ_2^2	1,082.48	1,081.36	1,081.05
σ_3^2			232,567
ϱ_1	0.5077	0.5068	0.5067
ϱ_2	0.7485	0.7480	0.7480
ϱ_3			0.7369
ϱ_{12}	0.0126	0.02444	0.02440
ϱ_{13}			0.02675
ϱ_{23}			0.09340

Chapter 5

DISCUSSION AND CONCLUSION

This study applies the linear mixed model framework to examine the partial correlation coecient between two repeatedly measured biomarkers while controlling for other multiple biomarkers that are also observed repeatedly. The estimated partial correlation coecients were investigated in both linked and unlinked cases under the assumption that the biomarkers follow a multivariate normal distribution. The results shown by the proposed method are generally consistent with the results of previous studies after controlling for the third variable. Therefore, it can be justified that our proposed methods will be useful to estimate partial correlation coecients when we have more than two repeatedly measured biomarkers. If continuous biomarkers do not follow a multivariate normal distribution, an appropriate transformation of the data must be done before applying the proposed method. Although the results are not presented in this thesis, we assessed the correlation coecient between pH and PaCO2 after controlling for two other variables created artificially: HCO3 and another variable, i.e., total four repeatedly measured markers, and our method gave very reasonable estimates. The estimation of partial correlation coecient after controlling for other repeatedly measured biomarkers can be easily achieved using the mixed procedure in SAS.

In this study, we assume that AR(1) or CS correlation matrix of the repeated measurements on each variable but the model with dierent correlation structure such as Toeplitz or ARMA(m,n) can be investigated in the future. This study could also have gone further by

developing the correlation coecient for multiple repeatedly measured categorical variables.

There are dierent types of biomarkers for glucose variability measures such as mean, standard deviation, maximum, special indexes like Mean Amplitude of Glycemic Excursions (MAGE), and so on (Eslami et al. 2011). In this thesis, we examined the relationship between APACHE III score and a new summary measure of glucose, the level of glucose calculated by the area under the blood glucose curve, after controlling for other potential con- founders. We found that the relationship between glucose level and APACHE III score with or without controlling for the level of vasopressor is very low among critically ill hospitalized adults. This finding suggests that increased glucose level of first five days after ICU admission is independently associated with increased severity of disease among patients with serious condition of septic shock. It is also possible not only that a new summary measure of glucose level based on the area under the curve does not explain the variability in blood glucose well in this cohort but also that a lot of missing values in our dataset could be a problematic.

APPENDIX: SAS CODE

```
data Linked3;
input patient repeat vartype response @@;
lines;
1 1 1 6.68    1 1 2 3.97    1 1 3 4.19
1 2 1 6.53    1 2 2 4.12    1 2 3 3.05
1 3 1 6.43    1 3 2 4.09    1 3 3 3.19
1 4 1 6.33    1 4 2 3.97    1 4 3 3.49
(omitted)
7 1 1 6.86    7 1 2 6.85    7 1 3 6.77
7 2 1 6.94    7 2 2 6.44    7 2 3 6.54
7 3 1 6.92    7 3 2 6.52    7 3 3 6.18
8 1 1 7.19    8 1 2 5.28    8 1 3 4.23
8 2 1 7.29    8 2 2 4.56    8 2 3 4.59
8 3 1 7.21    8 3 2 4.34    8 3 3 4.87
8 4 1 7.25    8 4 2 4.32    8 4 3 3.69
8 5 1 7.20    8 5 2 4.41    8 5 3 3.28
8 6 1 7.19    8 6 2 3.69    8 6 3 3.53
8 7 1 6.77    8 7 2 6.09    8 7 3 3.24
8 8 1 6.82    8 8 2 5.58    8 8 3 2.92
;

/* With all subjects assuming AR(1) structure */
proc mixed method = ml covtest data=Linked3;
class patient vartype repeat;
model response = vartype / solution ddfm = kr;
random vartype / type=un subject = patient v vcorr;
repeated vartype repeat/ type = un@ar(1) subject = patient r rcorr;
run;
```

BIBLIOGRAPHY

1. Bland, J.M and Altman D.G (1994). Correlation, Regression, and Repeated Data. British Medical Journal 308, 896.

2. Bland, J.M and Altman D.G (1995a). Calculating Correlation Coefficients with Repeated Observations: Part 1 - Correlation Within Subjects. British Medical Journal 310, 446.

3. Bland, J.M and Altman D.G (1995b). Calculating Correlation Coefficients with Repeated Observations: Part 2 - Correlation between Subjects. British Medical Journal 310, 633.

4. Lam, Miu, Webb, Catherine A. and O'Donnell, Denis E (1999). Correlation Between Two Variables in Repeated Measures. American Statistical Association, Proceedings of the Biometric Section 213-218.

5. Hamlett, A, Ryan, L, Serrano-Trespalacios, P and Wolfinger, R (2003). Mixed Models for Assessing Correlation in the Presence of Replication. Journal of the Air & Waste Management Association 53, 442-450.

6. Hamlett A, Ryan L, Wolfinger R (2004). On the use of Proc Mixed to estimate correlation in the presence of repeated measures. Proceedings of the Twenty-Ninth Annual SAS Users Group International Conference; SAS Institute Inc, 2004 Paper 19829.

7. Roy A (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biometrical Journal 2006;48:286301.

8. Krinsley JS (2003). Association between hyperglycemia and increased hospital mortality in a heterogeneous population of critically ill patients. Mayo Clin Proc. 2003 Dec;78(12):1471-8.

9. Eslami S, Taherzadeh Z, Schultz MJ, Abu-Hanna A. (2011). Glucose variability measures and their effect on mortality: a systematic review. Intensive Care Med. 2011 Apr;37(4):583-93.