CORRELATION COEFFICIENTS BETWEEN LONGITUDINALLY MEASURED MARKERS IN PEDIATRIC LIVER TRANSPLANT CANDIDATES WITH BILIARY ATRESIA

by

Shushen Xu

B. Med. in Pediatrics, China Medical University, China, 1989

D. Med. in Microbiology and Immunology, Chinese Academy of Preventive Medicine, China, 2000

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 was presented

by

Shushen Xu

It was defended on

July 29, 2015

and approved by

John Wilson, Ph.D, Assistant Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Jonathan Yabes, Ph.D, Assistant Professor, Division of General Internal Medicine, School of Medicine, University of Pittsburgh

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

Copyright © by Shushen Xu

2015

Chung-Chou H. Chang, Ph.D.

CORRELATION COEFFICIENTS BETWEEN LONGITUDINALLY MEASURED MARKERS IN PEDIATRIC LIVER TRANSPLANT CANDIDATES WITH BILIARY ATRESIA

Shushen Xu, M.S.

University of Pittsburgh, 2015

ABSTRACT

Biliary atresia (BA) is the most common pediatric liver disease leading to liver transplantation during childhood, with very poor prognosis if untreated. In this study, we aimed to apply a linear mixed effect (LME) model to estimate the correlation coefficients among longitudinally measured total serum bilirubin, international normalized ratio (INR) for prothrombin time, and serum albumin, the three important prognosis predictors of pretransplant mortality. The dataset was obtained from the Standard Transplant Analysis and Research (STAR) of the United Network of Organ Sharing (UNOS). The primary analysis cohort consists of 1,700 pediatric liver transplant candidates who started their liver transplant waiting list between February 27, 2002 and June 24, 2010 with at least one follow-up measurement and had primary diagnosis of biliary atresia at the time of listing. In applying the LME model, we estimated the longitudinally measured markers via two different correlation structures: autoregressive of order one (AR1) and compound symmetry (CS) in rearranged data by a 7-day equally spaced repeated measures interval. Under the AR(1) structure, the estimated total correlation coefficients between total bilirubin and INR, total bilirubin and albumin, and INR and albumin were 0.4151, -0.2404, and -0.206, respectively, whereas the partial correlation coefficients (within-subject correlation) were 0.0656, 0.0916, and -0.0451, respectively. Under the CS structure, the estimated total correlation coefficients were 0.4307, -0.2432, and -0.1912,

respectively and the partial correlation coefficients were 0.1742, -0.0678, and -0.0509, respectively for the above analysis. AR(1) structure had a better fit based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Several sensitivity analyses were conducted to understand the stability of the estimated overall correlation. The magnitudes of the estimates obtained from different sensitivity methods do not differ substantially.

Public health significance: For two repeatedly measured markers, the total correlation, the between-subject correlation with time-averaged values, and partial correlation for within-subject measurements will provide a more complete picture of the correlations for these markers. Correlation by stacking all measurements of a subject together or between-subject correlation with time-averaged values is a measurement ignoring time effects and could either over or under estimate the total correlation coefficients. The estimated correlations between any two markers measured repeatedly for patients awaiting liver transplantation will give physicians a tool to analyze the relationship between two markers for patients during the waitlist period and may further help physicians understand disease progression and refine treatment strategy for candidates prior to receiving a transplant.

Keywords: biliary atresia, liver transplant, longitudinal measures, partial correlation coefficient, total correlation coefficient.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my advisor Professor Chung-Chou H. Chang for providing me this opportunity and training. I also wish to thank Professor John Wilson and Professor Jonathan Yabes for being on my thesis committee. Finally, I cannot achieve this without the love and support from my families and friends.

1.0 INTRODUCTION

Biliary Atresia (BA) is the most common liver disease leading to liver transplantation (LT) among pediatric patients with end-stage liver disease, occurring approximately 1 in 8,000 (Asian countries) to 1 in 18,000 (European countries) live births, characterized by complete fibrotic obliteration of the lumen of all or part of the extrahepatic biliary tree within 3 months of life. The prognosis for untreated BA remains extremely poor, with a patient's median survival ranging from 8 to 16 months. [1-4]

Markers including total serum bilirubin, international normalized ratio (INR) for prothrombin time, and serum albumin are closely monitored in untreated BA patients for the purpose of prognosis prediction [5], but so far little is known about their correlation in repeated observations. Analyzed on most recent follow-up data from 48 patients awaiting LT (median waiting time was 17.08 months), Lee et al. (2005) have found that INR was significantly correlated with bilirubin in Pearson correlation of 0.63 ($P < 0.001$). [6]

For a longitudinal study with repeated measured markers, the correlation coefficient between two of the markers was first investigated by Bland and Altman (1995) as a weighted correlation coefficient, using the number of repeated measures as weights. The within-subject correlation (partial correlation) was measured by the method of analysis of covariance, treating subject as a categorical factor using dummy variables. [7, 8] This subject has been studied by several authors thereafter. Assuming that the repeated measures are correlated under either AR(1)

or CS structure, Roy (2006) proposed a linear mixed effects model to calculate the correlation coefficient between the two variables with or without the subject effect. [9] In this study, we used the model proposed by Roy to estimate the total and partial correlation coefficient between any of the two above-mentioned longitudinally measured markers for patients awaiting liver transplantation.

2.0 METHODS

2.1 DATA

The data used in this study was extracted from the Standard Transplant Analysis and Research (STAR) of the United Network of Organ Sharing (UNOS) which includes all liver transplant recipients in the United States who were on the transplant waiting list between February 27, 2002 and June 24, 2010. Patients whose primary diagnosis was biliary atresia and who were 17 years or younger were selected (N=2,039). During the data checking and cleaning phase, patients with no baseline information (N=7) or with data input errors (N=4) were excluded from the study. Finally, a cohort of 2,028 pediatric liver transplant candidates with primary diagnosis of biliary atresia were included in the analysis. In this study, we planned to investigate the correlation among three markers measured repeatedly during the time of waiting for liver transplantation: total serum bilirubin level, INR, and serum albumin level. After consulting with transplant and medical clinicians, the boundaries (max, min) for total serum bilirubin, INR, and albumin were set at (60.7, 0.1 mg/dl), (10, 0.5), and (5.42, 1.10 g/dl) respectively. There were a few data with these markers out of the reasonable ranges. We were unable to determine whether those values were actual measured values or due to input errors. To deal with those out-of-boundary values, we replaced the values by their reasonable maximum and minimum boundary values for subsequent analyses. Among 2,028 pediatric candidates in our analysis cohort, 328 did not have

follow-up measurements of these three markers. This resulted in 1,700 patients with at least one follow-up measurements.

2.2 STATISTICAL MODELS

We assumed that there are two variables measured repeatedly and the total correlation coefficient ($\rho_{y_1y_2}$) between these two variables will be estimated from a LME model. We applied the method to the liver transplant data and fit three separate models to estimate correlation coefficients for bilirubin and INR, bilirubin and albumin, and INR and albumin. In this section, we use bilirubin and INR as an example to explain our method.

For the i-th patient (i = 1, . . . , N), let $y_{ij} = (y_{1ij}, y_{2ij})'$ be a 2 × 1 vector of the j-th repeated observations (j = 1, . . . , n_i) of variables y_1 and y_2, to represent bilirubin and INR respectively, where N is the total number of patients. We assume that both variables are measured at the same time points, and let p be the maximum number of repeated measurements, i.e., p = max(n_1, n_2, …, n_N). For the i-th patient, we assume that the 2 × 1 vector y_{ij} follows a bivariate normal distribution with the form,

$$\begin{bmatrix} y_{1ij} \\ y_{2ij} \end{bmatrix} \sim BVN\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma\right),$$

where μ_1 and μ_2 are the means of y_{1ij} and y_{2ij} respectively, and Σ is the variance-covariance matrix. If we further assume that $\sigma^2_{y_1}$, $\sigma^2_{y_2}$ are the variances of y_{1ij} and y_{2ij} and $\sigma_{y_1y_2}$, $\tilde{\rho}_{y_1y_2}$ are the

covariance and the correlation between $\sigma^2_{y_1}$, $\sigma^2_{y_2}$, then the variance-covariance matrix Σ can be rewritten into the following form,

$$\Sigma = \begin{bmatrix} \sigma^2_{y_1} & \tilde{\rho}_{y_1 y_2}\sigma_{y_1}\sigma_{y_2} \\ \tilde{\rho}_{y_1 y_2}\sigma_{y_1}\sigma_{y_2} & \sigma^2_{y_2} \end{bmatrix}.$$

Note that the correlation coefficient for variable y_j measured at different time points is denoted as ρ_{y_j} (j = 1, 2). The correlation coefficient between variables y_1 and y_2 at a single time point, also called the partial correlation coefficient, can be calculated by $\tilde{\rho}_{y_1 y_2} = \frac{\sigma_{y_1 y_2}}{\sigma_{y_1}\sigma_{y_2}}$. We also define the correlation between the two outcome variables measured at different time points as $\mathbf{Corr}(y_{1ij}, y_{1ij*}) = \delta\rho_{y_1 y_2}$, where the correction factor δ is usually less than 1, indicating that correlations between the two variables measured at different time points are lower in magnitude than those measured at the same time point.

Based on the specification described above, the LME model for the $2n_i$ dimensional outcome vector $y_i = (y_{i1}, y_{i2}, \ldots, y_{ip})' = (y_{1i1}, y_{2i1}, y_{1i2}, y_{2i2}, \ldots, y_{1ip}, y_{2ip})'$, $1 \le 2n_i \le 2p$ can be written in the following form, [10]

$$y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i , \qquad (1)$$

where X_i and Z_i are the fixed and the random design matrices respectively; β is a vector of fixed effects; γ_i is a vector of random effects; and ε_i is a vector of random errors. We assume that $\gamma_i \sim N(0, G)$, $\varepsilon_i \sim N(0, R_i)$, and γ_i and ε_i are independent., i.e., $Cov(\gamma_i, \varepsilon_i) = 0$.

From (1), we have $E(y_i) = X_i\beta$ and $Var(y_i) = Z_iGZ_i' + R_i$. The variance-covariance matrix between the two outcome variables at a given time point is defined as a 2 × 2 matrix Σ =

$\begin{bmatrix} \sigma^2_{y_1} & \sigma_{y_1 y_2} \\ \sigma_{y_1 y_2} & \sigma^2_{y_2} \end{bmatrix}$ and is assumed to be the same for all time points. If we would like to specify a correlation structure for the repeated measures, matrix R_i can be written as the form $R_i = V_i \otimes \Sigma$, where $\otimes$ is the Kronecker product; V_i is the $p \times p$ matrix representing the correlation of repeated measures on a given outcome variable and is assumed to be the same for both outcome variables. When a compound symmetric (CS) correlation structure or an autoregressive of order one (AR[1]) correlation structure is specified, matrix V_i will have the form

$$V_i = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \rho & \vdots \\ \vdots & \rho & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix} \quad \text{or} \quad V_i = \begin{bmatrix} 1 & \rho & \rho^2 & & \rho^{p-1} \\ \rho & 1 & \rho & \cdots & \rho^{p-2} \\ \rho^2 & \rho & 1 & & \rho^{p-3} \\ & \vdots & & \ddots & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \cdots & 1 \end{bmatrix}$$

respectively. The CS correlation structure assumes equal correlation among all the repeated measurements, whereas the AR(1) correlation structure assumes a time decay between successive observations and is usually a more realistic assumption for longitudinal observations. Note that it is required that $-\frac{1}{p-1} < \rho < 1$ for V_i to be positive definite. [9, 11-12]

2.3 DATA ANALYSIS

We first examined the between-subject correlations without considering the time effect. We constructed scatter plots and calculated Pearson's correlation coefficient using the time-averaged values of the markers within a subject. We will refer to this type of analysis as *between subject*

correlation ignoring time-effect ($\bar{\rho}_{y_1 y_2}$). We also calculated correlation coefficients between two markers ignoring repeated measures by stacking all pair of measurements of the two variables together. We will refer to this analysis as *naïve correlation* ($\grave{\rho}_{y_1 y_2}$).

The longitudinal correlation coefficients between two markers were analyzed using the LME model described in Section 2.2. Equally spaced time interval is required for using this methodology. Because the median time between measurements (9 days) is close to weekly measurements, we fixed our time interval into 7 days. Using LME model, we analyzed the repeated measurements that only occurred on multiples of 7 days (0, 7, 14,...) from the time of listing. Missing data that occurred on these days were imputed using last observation carried forward (LOCF) from the closest available prior visit. In an attempt to apply the model to the liver transplant data ignoring the actual measurement times, we also analyze the data by treating the order of the measurement time (called the *visit* time) as an equally spaced assessment.

We created three different versions of datasets for the analysis in order to balance among maximum information, minimum noise, and model convergence. **Dataset 1** was the original cohort described in Section 2.1. **Dataset 2** was generated from Dataset 1 by including only those patients who had at least one follow-up lab measurements. **Dataset 3** was obtained from Dataset 2 by including only those patients aged 11 years old or younger at the time of first being on the liver transplant waiting list.

All data managements and analyses were implemented in SAS 9.4. The SAS code was given in the Appendix. In the SAS code, variable `Patient` represents the subject pseudo ID and variable `Follow` represents the number of measurements for each subject. A vector variable `mvar` was used to define the 3 longitudinal markers; `mvar` = 1 for variable `bilirubin` or

`n_bilirubin` (rescaled bilirubin, `n_bilirubin` = 1/10 of `bilirubin`), `mvar` = 2 for variable `INR` and `mvar` = 3 for variable `albumin`. Models were fit with the intercept and `mvar` were the fixed effects and `mvar` was the random effect.

3.0 RESULTS

3.1 DESCRIPTIVE ANALYSIS

Table 1 gives the summary statistics for the 3 analysis datasets. Similar distributions were observed for each variable across the datasets. Approximately 70% of the transplant candidates were under one year old. For Datasets 2 and 3, half of the candidates had at least 5 measurements for each marker while for Dataset 1 half of the candidates had at least 4 measurements. The maximum follow-up day was 2,869 days and the median follow-up time for Dataset 1 was 60 days and for Datasets 2 and 3 were about 83 days. The median time between measurements was 9 days for all three datasets. Similar distributions of albumin, bilirubin, and INR were found across all three datasets.

Table 1. Characteristics of patients in three datasets

Characteristics	Dataset 1	Dataset 2	Dataset 3
Number of patients	2028	1700	1675
Age of being on the waiting list, median, mean ± SD (years)	0, 1.17 ± 2.74	0, 1.10 ±2.68	0, 0.90 ± 2.12
< 1	1381 (68.10)	1181 (69.47)	1181 (70.51)
1	321 (15.83)	260 (15.29)	260 (15.52)
2	71 (3.50)	61 (3.59)	61 (3.64)
3-4	63 (3.11)	50 (2.94)	50 (2.99)
5-9	127 (6.26)	96 (5.65)	96 (5.73)
10-11	34 (1.68)	27 (1.59)	27 (1.61)
12-17	31 (1.53)	25 (1.47)	
Number of measurements per individual, median, mean ± SD Range	4, 6.09 ± 5.99 1 – 73	5, 7.07 ± 6.07 2 - 73	5, 7.10 ± 6.08 2 - 73
Days of follow up, median, mean ± SD (days) range	60, 168.07 ± 313.80 0 – 2869	83, 200.50 ± 333.13 1 - 2869	84, 201.55 ± 334.57 1 - 2869
Time between measures, median, mean ± SD (days) Range	9, 33.03 ± 75.23 1 – 1312	9, 33.03 ± 75.23 1 - 1312	9, 33.04 ± 75.34 1 - 1312
Laboratory values, median, mean ± SD			
Albumin (g/dl)	3.00, 3.00 ± 0.71	3.00, 3.00 ± 0.71	3.00, 2.99 ± 0.71
Total Bilirubin (mg/dl)	12.60, 14.54 ± 11.43	12.70, 14.65 ± 11.48	12.70, 14.65 ± 11.48
INR	1.34, 1.59 ± 0.81	1.36, 1.59 ± 0.81	1.35, 1.59 ± 0.82

Dataset 1: the original dataset; Dataset 2: generated from Dataset 1 by including only those patients who had at least one follow-up lab measurements; Dataset 3: obtained from Dataset 2 by including only those patients aged 11 years old or younger at the time of first being on the liver transplant waiting list.

3.2 STATISTICAL RESULTS

Table 2 summarizes the estimated naïve correlation and between-subject correlation, ignoring time effect for bilirubin and INR, bilirubin and albumin, and INR and albumin for three datasets. The estimated correlations are similar across all three datasets: moderate positive correlation between bilirubin and INR, and relatively weak negative correlation between bilirubin and albumin and INR and albumin. As an illustration, scatter plots in Figure 1 give visual

representations of the between-subject correlations ignoring time effect based on Dataset 2. Compared the naïve correlation with between-subject correlations, between-subject correlations have slightly stronger association in absolute magnitude.

Table 2. Naïve and between-subject correlations between two repeatedly measured markers in three different datasets

		bilirubin(y_1) & INR(y_2)	bilirubin(y_1) & albumin(y_2)	INR(y_1) &albumin(y_2)
Correlation coefficient				
Dataset 1	$\grave{\rho}_{y_1y_2}$	0.3958	-0.1476	-0.1350
	$\bar{\rho}_{y_1y_2}$	0.4839	-0.2579	-0.1954
Dataset 2	$\grave{\rho}_{y_1y_2}$	0.3952	-0.1447	-0.1350
	$\bar{\rho}_{y_1y_2}$	0.4991	-0.2616	-0.2137
Dataset 3	$\grave{\rho}_{y_1y_2}$	0.3962	-0.1397	-0.1328
	$\bar{\rho}_{y_1y_2}$	0.5002	-0.2561	-0.2087

$\grave{\rho}_{y_1y_2}$, Pearson correlation coefficients for stacking all pair of measurements of the two variables together (naïve correlation);

$\bar{\rho}_{y_1y_2}$, Pearson correlation coefficients for between-subject correlation ignoring time effect.

In fitting LME models, we experienced convergence issues which might be caused by small sample size at later time points. We checked the model convergence for all three datasets with two different types of correlation structures (AR[1] and CS), and two different versions of bilirubin measurements (in the original scale and in one-tenth of the original level [rescaled]) and presented the findings in Table 3. The models converged for both correlation structures when Dataset 2 and rescaled bilirubin were used. For the subsequent analyses, we will present the results based on Dataset 2.

Table 4 shows the comparison of estimated parameters and goodness-of-fit statistics between the two correlation structures. The estimated total correlation coefficients ($\rho_{y_1y_2}$) were

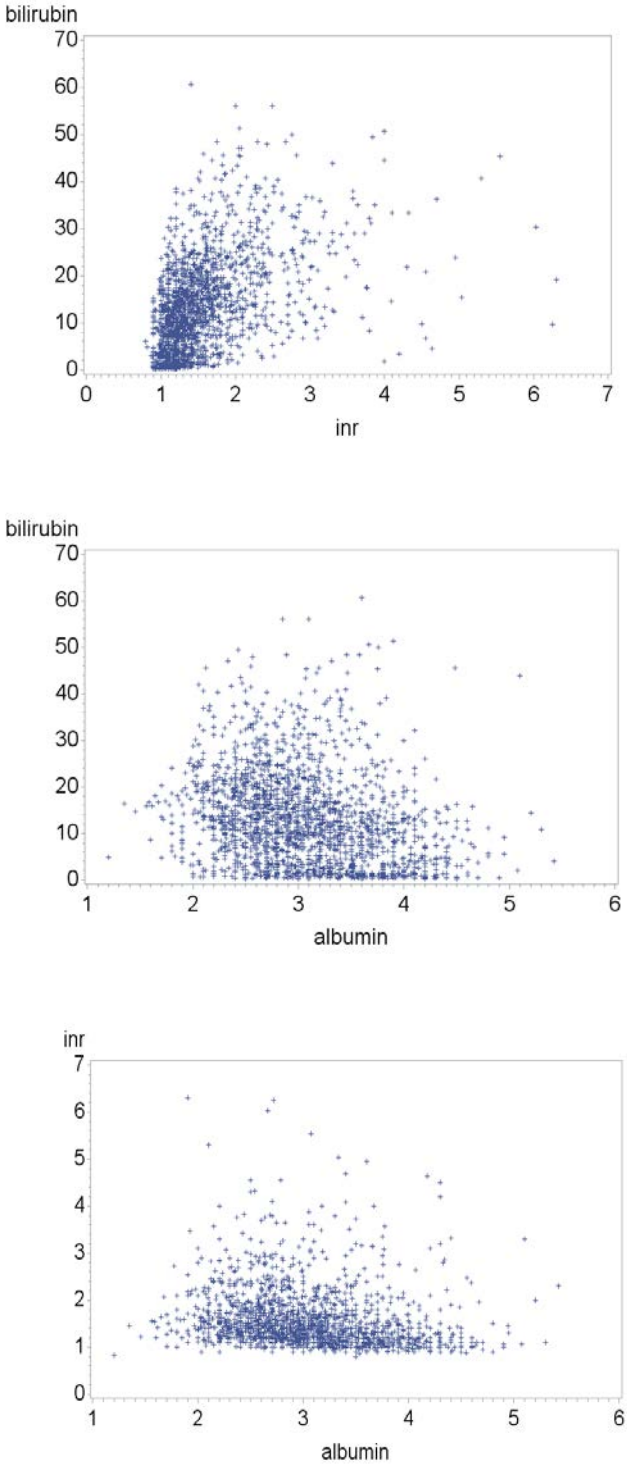

Figure 1. Scatter plots between two markers. Each point on the plot was the time-averaged marker value for each subject

Table 3. Summary of convergence in using different datasets and scales of bilirubin

	bilirubin & INR		bilirubin & albumin		INR & albumin	
Correlation structure	AR(1)	CS	AR(1)	CS	AR(1)	CS
Dataset 1						
bilirubin	x	√	x	√	√	x
rescaled bilirubin*	√	x	√	x	√	x
Dataset 2						
bilirubin	x	√	x	√	√	√
rescaled bilirubin*	√	√	√	√	√	√
Dataset 3						
bilirubin	x	√	x	√	√	x
rescaled bilirubin*	√	x	√	x	√	x

√: converge; x : not converge; rescaled bilirubin*: 1/10 of bilirubin

Table 4. Comparison of linear mixed effects analysis results for Dataset 2 using different correlation structures

	bilirubin(y_1) & INR(y_2)		bilirubin(y_1) & albumin(y_2)		INR(y_1) & albumin(y_2)	
Correlation Structure	AR(1)	CS	AR(1)	CS	AR(1)	CS
Parameter						
$\rho_{y_1y_2}$	0.4151	0.4307	-0.2404	-0.2432	-0.2060	-0.1912
$\tilde{\rho}_{y_1y_2}$	0.0656	0.1742	0.0916	-0.0678	-0.0451	-0.0509
ρ	0.8868	0.2530	0.9290	0.0489	0.8659	-0.0005
δ	0.9949	0.9436	1.0087	0.9576	0.9893	0.9467
Fit statistics						
AIC	-36378.4	48697.4	-54770.2	69576.1	-37110	50710.1
BIC	-36329.5	48746.4	-54721.3	69625.1	-37061	50759.1

$\rho_{y_1y_2}$, total correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ, correlation coefficients for repeated measures; δ, correction factor.

similar based on the models with AR(1) and CS correlation structures for the repeated measures. These estimated results also indicated a moderate positive correlation between bilirubin and INR, and relatively weak negative correlations between bilirubin and albumin, and INR and albumin. The partial correlation coefficients ($\tilde{\rho}_{y_1y_2}$) at a single time point were very weak (ranged from 0.0451 to 0.1742 in absolute value), and the correlation coefficients among the repeated

measurements (ρ) were very strong (ranged from 0.8659 to 0.9290) based on the AR(1) structure, and very weak (ranged from 0.0005 to 0.2530 in absolute value) based on the CS structure. Note that the overall model fit was better for using AR(1) than CS because of a smaller value in Akaike information criterion (AIC) and Bayesian information criterion (BIC). Except in analyses of bilirubin and albumin (δ=1.0087), the estimates of δ were less than 1, which indicates that the correlation between two markers measured at different time points were mostly lower than the correlation between those markers measured at the same time point.

If analysis used visits rather than actual measurement times, we found slightly weaker total correlation coefficients between the markers and slightly weaker correlations among repeated measures, while we found similar partial correlation coefficients between the markers. This analysis was based on Dataset 2 and assumed having AR(1) correlation structure (Table 5). If we compared the same data with different correlation structures, model with CS had a slightly stronger total correlation in absolute magnitude, and much weaker correlation among repeated measures (Table 6).

Table 5. Comparison of linear mixed effects analysis results for Dataset 2 using actual measurement times versus visits as the time factors in the models

	bilirubin(y_1) & INR(y_2)		bilirubin(y_1) & albumin(y_2)		INR(y_1) & albumin(y_2)	
Repeated Measures	actual time	visit	actual time	visit	actual time	visit
Parameter						
$\rho_{y_1y_2}$	0.4151	0.3120	-0.2404	-0.1344	-0.2060	-0.1451
$\tilde{\rho}_{y_1y_2}$	0.0656	0.0669	0.0916	0.0926	-0.0451	-0.0551
ρ	0.8868	0.7842	0.9290	0.8413	0.8659	0.6813

$\rho_{y_1y_2}$, overall correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ , correlation coefficients for repeated measures.

Table 6. Comparison of linear mixed effects analysis results for Dataset 2 between different correlation structures using visits as the time factors in the models

	bilirubin(y_1) & albumin(y_2)		INR(y_1) & albumin(y_2)	
Correlation Structure	AR(1)	CS	AR(1)	CS
Parameter				
$\rho_{y_1y_2}$	-0.1344	-0.1747	-0.1451	-0.1515
$\tilde{\rho}_{y_1y_2}$	0.0926	0.0990	-0.0551	-0.0372
ρ	0.8413	-0.0139	0.6813	-0.0139
Fit statistics				
AIC	30028.7	43125.9	33608.8	41460.6
BIC	30077.7	43174.9	33657.8	41509.6

The analysis of bilirubin & INR did not converge for both correlation structures in dataset 2.

$\rho_{y_1y_2}$, total correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ, correlation coefficients for repeated measures.

For comparison purpose, we also refit the LME models with AR(1) correlation structure and actual measurement times as time interval for Dataset 1 and Dataset 3. The estimated total correlation, partial correlation, and correlation repeated measures were very similar across three datasets (Table 7).

As we described before, sample size became smaller when follow-up time got longer. To check the stability of the estimates, we investigated the effect on estimated total correlation coefficients for data truncated at different follow-up times: 3 months, 6 months, 1 year, 2 years, and data with no truncation. Figure 2 depicts these estimates using either data with equally spaced time-interval (Panel A) or data with different visits (Panel B). In the latter case, the estimates did not change much. In the former case, the absolute correlation estimates tended to be stronger when follow-up time gets longer.

Because of a wide range of age in our study subjects, we refit the LME models by including age as a fixed effect and re-estimated the correlations between markers. Age was

Table 7. Comparison of linear mixed effects analysis results among 3 different datasets

	bilirubin(y_1) & INR(y_2)			bilirubin(y_1) & albumin(y_2)			INR(y_1) & albumin(y_2)		
Dataset	1	2	3	1	2	3	1	2	3
Parameter									
$\rho_{y_1y_2}$	0.4079	0.4151	0.4159	-0.2370	-0.2404	-0.2367	-0.1924	-0.2060	-0.2019
$\tilde{\rho}_{y_1y_2}$	0.0657	0.0656	0.0656	0.0915	0.0916	0.0924	-0.0450	-0.0451	-0.0445
ρ	0.8866	0.8868	0.8863	0.9289	0.9290	0.9286	0.8659	0.8659	0.8653

Models used AR(1) correlation structure and actual measurement times as the time factors.
Dataset 1: the original dataset; Dataset 2: generated from Dataset 1 by including only those patients who had at least one follow-up lab measurements; Dataset 3: obtained from Dataset 2 by including only those patients aged 11 years old or younger at the time of first being on the liver transplant waiting list.

$\rho_{y_1y_2}$, total correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ, correlation coefficients for repeated measures

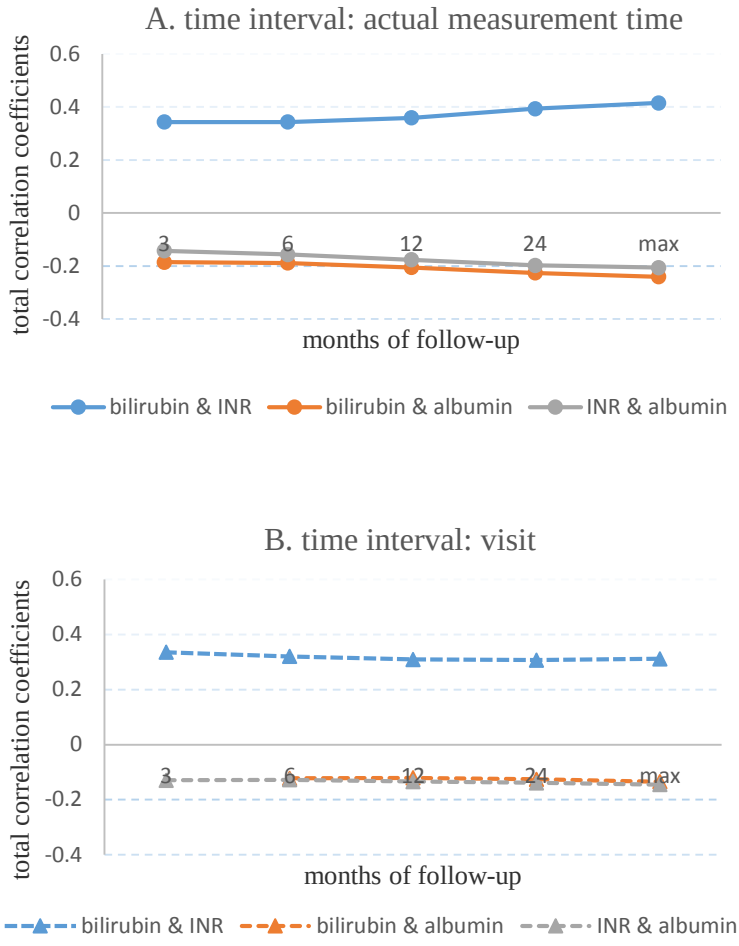

Figure 2. The estimated total correlation coefficients for data truncated at different follow-up time points: panel A, using actual measurement times; panel B, using visits as time factors

treated both as a continuous variable and a categorical variable in order to catch linear or possibly nonlinear effects of age on the markers. As shown in Tables 8, although the models including age had a little improvement according to the AIC and BIC values, the estimated correlation coefficients after adjusting for age were almost identical.

Table 8. Comparison of linear mixed effects analysis results between models with and without age as a fixed effect, with AR(1) correlation structure and Dataset 2

	bilirubin(y_1) & INR(y_2)			bilirubin(y_1) & albumin(y_2)			INR(y_1) & albumin(y_2)		
Fixed effect	no age	age, cat.*	age, con.**	no age	age, cat.*	age, con.**	no age	age, cat.*	age, con.**
Parameter									
$\rho_{y_1y_2}$	0.4151	0.4110	0.4131	-0.2404	-0.2403	-0.2396	-0.2060	-0.2099	-0.2081
$\tilde{\rho}_{y_1y_2}$	0.0656	0.0656	0.0657	0.0916	0.0913	0.0916	-0.0451	-0.0453	-0.0452
ρ	0.8868	0.8870	0.8868	0.9290	0.9290	0.9290	0.8659	0.8651	0.8653
Fit statistics									
AIC	-36378.4	-36382.7	-36376.9	-54770.2	-54787.1	-54768.6	-37110.0	-37136.1	-37123.4
BIC	-36329.5	-36301.1	-36322.5	-54721.3	-54705.6	-54714.2	-37061.0	-37054.5	-37069.0

age, cat*: categorical variable of age; age, con*: continuous variable of age.

$\rho_{y_1y_2}$, total correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ, correlation coefficients for repeated measures.

Sensitivity analysis was performed by comparing results in different lengths of follow-up days (7, 5, and 3 days) using Dataset 2. The results are summarized in Table 9. The estimated total correlation coefficients, partial correlation coefficients, and correlations among repeated measures are consistent among the comparisons and the models are robust.

Table 9. Comparison of linear mixed effects analysis results for Dataset 2 using different lengths of intervals for measurement times

	bilirubin(y_1) & INR(y_2)			bilirubin(y_1) & albumin(y_2)			INR(y_1) & albumin(y_2)		
Lengths (days)	7	5	3	7	5	3	7	5	3
Parameter									
$\rho_{y_1y_2}$	0.4151	0.4154	0.4113	-0.2404	-0.2449	-0.2427	-0.2060	-0.2131	-0.2092
$\tilde{\rho}_{y_1y_2}$	0.0656	0.0572	0.0465	0.0916	0.0885	0.1001	-0.0451	-0.0603	-0.0518
ρ	0.8868	0.9129	0.9522	0.9290	0.9490	0.9685	0.8659	0.8986	0.9421

$\rho_{y_1y_2}$, total correlation coefficient; $\tilde{\rho}_{y_1y_2}$, partial correlation coefficient; ρ, correlation coefficients for repeated measures

For model diagnostics, we used the marginal residuals to check the normality assumption of models with AR(1) correlation structure using Dataset 2 as suggested by Cheng et al. [13] Jackknifed studentized residual plots and normal plots were presented in Figures 3-5. Both the Kolmogorov-Smirnov normality tests ($P < 0.01$ in all 3 analyses: correlations between bilirubin and INR, bilirubin and albumin, and INR and albumin) and the quantile-quantile plots show some deviations from the normal distributions. No outlier was detected.

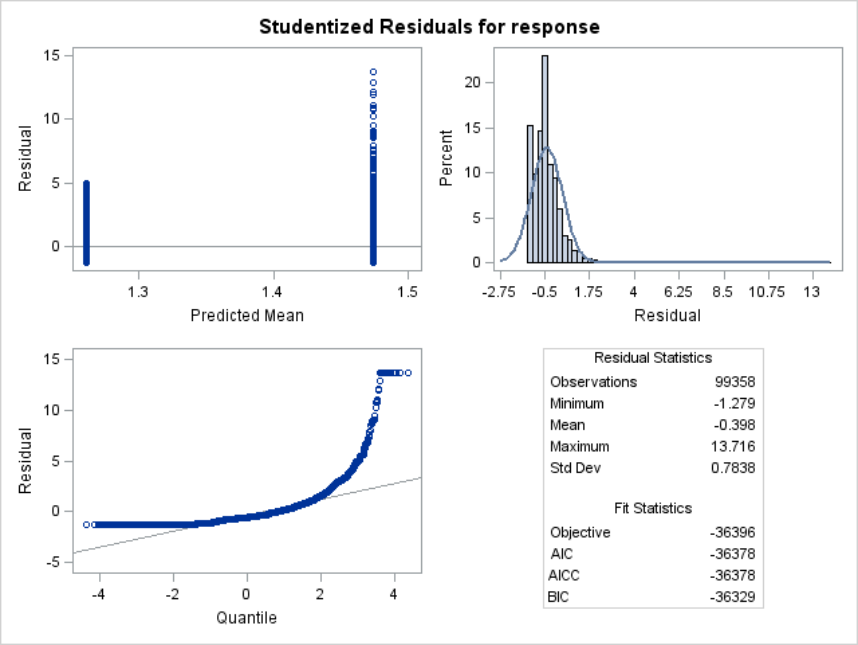

Figure 3. The Jackknifed studentized residual for the correlation analysis of bilirubin and INR

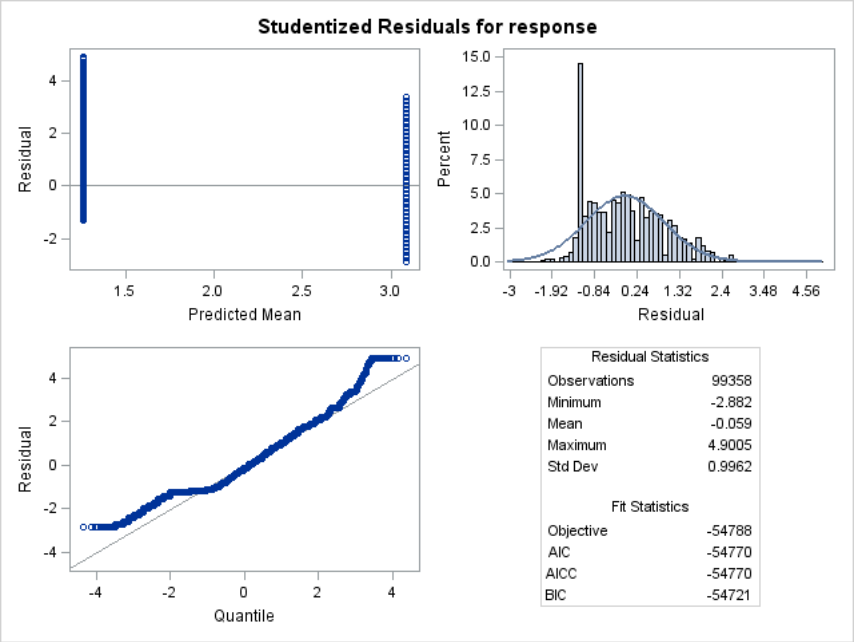

Figure 4. The Jackknifed studentized residual for the correlation analysis of bilirubin and albumin

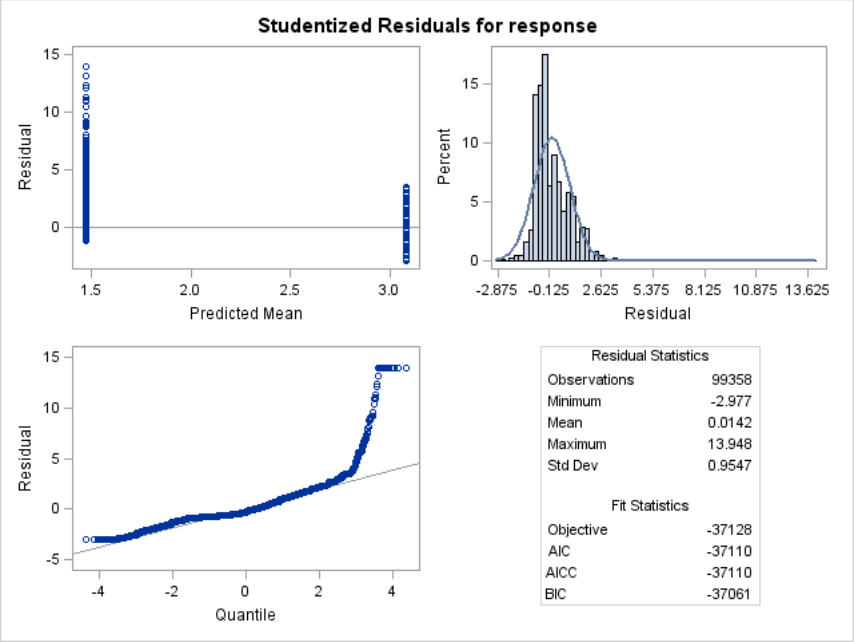

Figure 5. The Jackknifed studentized residual for the correlation analysis of INR and albumin

4.0 DISCUSSION

In our analysis, we mainly focused on a cohort consisting of 1,700 pediatric liver transplant candidates with primary diagnosis of biliary atresia and with at least one lab follow-up measure. We were interested in estimating correlation coefficients of the three prognosis markers total serum bilirubin, INR, and serum albumin that were measured repeatedly over the waitlist period.

For each patient, study baseline was defined as time of entry to the liver transplant waiting list. We used three different methods to estimate the correlation coefficients between any two of the three above-mentioned markers: between-subject correlation for markers averaged across all time points for each patient, correlation between two markers with all measurements stacked together and ignoring correlation due to repeated measurements, and total correlation using the linear mixed effects modeling adjusting for within-patient correlations. The sign of the correlation coefficients estimated from these three methods are the same and the magnitudes of the estimates obtained from these three methods do not differ substantially. The similarity of these three estimates could be due to relatively stable prognosis during the waitlist period and the high within-subject correlation for a given variable. The correlation between bilirubin and INR was moderately positive whereas the correlations between bilirubin and albumin and INR and albumin were weakly negative. Note that based on the values of AIC and BIC, linear mixed effects model with AR(1) correlation structure of repeated measurements had a better fit than that with CS correlation structure.

Several sensitivity analyses were conducted to understand the stability of the estimated total correlation. First, we selected various lengths of equally spaced repeated measurements for the linear mixed effects modeling: 3, 5, and 7 days. The estimated total correlations were not much different. Second, we included repeated measurements in various lengths of follow-up for the analysis: 3-month, 6-month, 1-year, 2-year, and the maximum (about 8 years) follow up. The longer the follow-up time is, the higher percent of missing measurements will be in the later follow-up time points. Despite having various lengths of follow-up and various percentage of missing measurements, we found that the estimated total correlations were not much different. Third, we excluded pediatric patients who were 12 years or older at the baseline. Since January 11, 2005, the UNOS changed eligible age for using the PELD organ allocation system from 17 years and younger to 11 years and younger. The third sensitivity analysis compared the estimation results among pediatric patients defined using these two different thresholds and the estimated correlations do not differ much. Finally, we used the measurements at different visits and considered visits as the time factor in the linear mixed effects model. For a patient, when disease progression was suspected to change during the waitlist period, lab tests were usually ordered by a doctor and marker values were then updated. Our results showed that the estimated total correlations were smaller for using visit than using equally-spaced time interval as the time factors in the model.

There are some limitations which we are aware of for estimating total correlation using the linear mixed effects modeling. First, because of higher percentage of missing data when follow-up time is longer in the analysis, linear mixed effects model may face the issue of nonconvergence. This missing data problem is common in any longitudinal study. We suggest conducting different sensitivity analyses including the between-subject time-averaging method

and the naïve stacking method to resolve the convergence issue and to understand the correlation patterns under different scenarios. Second, violation of homogeneous variance for a linear mixed effects model may also occur. To obtain reliable estimates, regression diagnostics and further adjustments to correct for violation will be needed.

Correlation coefficient is an easy-to-interpret quantity for nonstatistician clinical researchers. Correlation between two longitudinally measured markers can give us a more accurate estimation of the linear relationship between two variables if it is adjusted for the similarity from a within-subject association. Because of the use of a regression-based method in estimating correlation, several extensions including adjusting for other fixed effect confounding factors or adjusting for clustering effects in a multilevel setting can be further investigated in the future.

APPENDIX: SAS CODE FOR ANALYSIS OF CORRELATION BETWEEN BILIRUBIN AND INR UNDER AR(1) STRUCTURE IN DATASET 2

```
data data3;
input id calc_age  agec bilirunim inr albumin  init_age dayfrominitdate visit maxvisit;
cards;
7240     0    1     5.6    3.6   1.4      0       0      1      1
11630    1    2    12.9    3.9   1.4      1       0      1      4
11630    1    2     8.9    3.3   1.4      1      24      2      4
11630    1    2    10.6    3.9   1.5      1      45      3      4
11630    1    2    10.6    3.9   1.5      1     119      4      4
80280    0    1    12.4    2.6   1.5      0       0      1      2
80280    0    1    12.4    2.6   1.5      0       9      2      2
  .      .    .     .       .     .        .       .      .      .

;
Run;
data new_b; set data3; n_bilirubin=bilirubin/10;
if maxvisit=1 then delete;
```

```
data data3a_new_b; set new_b;
by id;
if first. id then do ; patient+1; end;

data data3b_new_b;
merge data3a_new_b data3a_new_b (firstobs=2 keep=patient dayfrominitdate rename = (patient
=_patient dayfrominitdate=_dayfrominitdate)); output;
if patient=_patient then do;
 do i=dayfrominitdate+1 to _dayfrominitdate-1; dayfrominitdate=i;output;
 end; end;
drop i _:;
run;

data data4_new_b; set data3b_new_b;
follow=round(dayfrominitdate/7,.001);
if int(follow)=follow;

data data5_new_b; set data4_new_b;
by patient;
response=n_bilirubin; mvar=1; output;
response=inr; mvar=2; output;
response=albumin; mvar=3; output;
run;
```

```
proc mixed data=data5_new_b method = ml covtest;
class patient mvar follow;
model response = mvar/ solution ddfm=residual;
random mvar/ type=un subject = patient v =2 vcorr =2;
repeated mvar follow/ type = un@ar(1) subject = patient r =2 rcorr =2;
where mvar=1 or mvar=2;
run;
```

`calc_age`, the age at the time of follow-up

`agec`, the categorical variable of `calc_age` with the `agec`=1, 2, 3, 4, 5, 6, 7, corresponding to `calc_age`= 0, 1, 2, [3, 4], [5, 9], [10, 11], [12, 17] years old respectively

`init_age`, the age of being listed as a liver transplant candidate

`dayfrominitdate`, from the date on the waiting list to the date of follow-up

`visit`, the order of measurement

`maxvisit`, the maximum number of visit

BIBLIOGRAPHY

1. Sokol RJ. Pathogenesis and Outcome of Biliary Atresia: Current Concepts. Journal of Pediatric Gastroenterology & Nutrition 2003; 37(1): 4-21

2. Bates MD, Bucuvalas JC, Alonso MH, Ryckman FC. Biliary atresia: pathogenesis and treatment. Semin Liver Dis 1998; 3: 281–293.

3. Goss JA, Shackleton CR, Swenson K, Satou NL, Nuesse BJ, Imagawa DK, Kinkhabwala MM, Seu P, Markowitz JS, Rudich SM, McDiarmid SV, Busuttil RW. Orthotopic liver transplantation for congenital biliary atresia: a 11-year, single center experience. Ann Surg 1996; 224:276–287.

4. Fouquet V, Alves A, Branchereau S, Grabar S, Debray D, Jacquemin E, Devictor D, Durand P, Baujard C, Fabre M, Pariente D, Chardot C, Dousset B, Massault PP, Bernard D, Houssin D, Bernard O, Gauthier F, Soubrane O. Long-term outcome of pediatric liver transplantation for biliary atresia: a 10-year follow-up in a single center. Liver Transpl 2005; 11(2):152–160.

5. Ebell MH. Predicting prognosis in patients with end-stage liver disease. Am Fam Physician 2006; 74(10):1762-1763.

6. Lee YM, Wai CT, Da Costa M, Lee KH, Sutedja D, Tan KC, Isaac J, Lim SG. Bilirubin appears to be the only independent variable affecting mortality on liver transplant waiting list if waiting time exceeds 1 year. Transplant Proc 2005; 37(10): 4365-4366

7. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 1 – Correlation within Subjects. British Medical Journal 1995; 310: 446.

8. Bland JM, Altman DG. Calculating correlation coefficients with repeated observations: Part 2 – Correlation between Subjects. British Medical Journal 1995; 310: 633.

9. Roy A. Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biom J 2006; 48(2): 286-301

10. Larid NM, Ware JH. Random effects models for longitudinal data. Biometrics 1982; 38(4): 963–974.

11. Verbeke G, Molenberghs G. Linear Mixed Models for Longitudinal Data. Springer-Verlag; New York: 2000.

12. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. SAS for Mixed Models. SAS Institute; Cary, NC: 2006.

13. Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: building a good enough mixed model. Stat Med. 2010; 29(4): 504-520.