Advances and Applications in Statistics

Published Online: November 2014
Available online at http://pphmj.com/journals/adas.htm
Volume 41, Number 2, 2014, Pages 137-165

LONGITUDINAL DATA ANALYSIS STRATEGIES - AN APPLICATION OF MIXED MODEL TO POST-TRANSPLANT SERUM CREATININE DATA

Zailong Wang

Integrated Information Science
Novartis Pharmaceuticals Corporation
One Health Plaza, 135/108A
East Hanover, NJ 07936, U. S. A.

Abstract

In transplant clinical studies, serum creatinine data are collected repeatedly on same individuals over time to monitor individual renal function changes and to determine treatment effects on renal function as well as on other efficacy and safety endpoints. Hence longitudinal analysis methods are appropriate to model such data. In this paper, we systematically present the strategies of using mixed-effect regression model to analyze serum creatinine data from a Novartis transplant clinical trial. These strategies include fixed-effect covariate selection approach; covariance model selection; and model based treatment effect tests. Furthermore, various types of residuals from mixed-effect regression model and influence diagnostics are discussed in detail for model diagnosis. Related background of longitudinal analysis and clinical trial is also presented.

1. Introduction

A measurement of the serum creatinine (SCr) concentration level is used

Received: June 6, 2014; Accepted: July 28, 2014
2010 Mathematics Subject Classification: 62-07; 62P10.
Keywords and phrases: mixed-effect regression, covariance structure analysis, residual analysis, serum creatinine data.

to evaluate kidney function. Despite numerous known limitations, including error-prone variability in assay techniques, interference by endogenous chromogens and by clinical or environmental conditions independent of renal function, etc. [13, 23], SCr remains a generally acceptable clinical assessment. The clinical utility of the SCr measurements centers on its relation to the glomerular filtration rate (GFR) which is the filtering capacity of the kidneys. Several formulas have been developed to estimate GFR from SCr measurements. Among them, the Cockcroft-Gault method [3] and the modification of diet in renal disease (MDRD) method [14] are most commonly used.

In transplant clinical studies, post-transplant SCr data are collected repeatedly on the same individuals over time (longitudinally) to monitor individual renal function and to determine treatment effects on renal toxicity. Hence longitudinal analysis methods are appropriate to model the data, to identify predictors of renal function, and to test the corresponding hypotheses such as no difference among treatment groups. There are variants of models with a variety of names dealing with many of peculiarities of longitudinal data, for example, variance component models [5], random-effect models [10, 12], random coefficient models [4, 18] and mixed models [17]. These models in general can be used for analysis of longitudinal data or clustered data in which the data within subjects/clusters are dependent. This paper will investigate the proper use of longitudinal analysis models to assess the relationship between SCr measurements and covariates. In particular, we will present the mixed-effect regression model which generally contains some fixed effects as well as the random effects.

There are several features of above mentioned models making them useful for longitudinal data analysis. First subjects with incomplete data across time are not excluded from the analysis so the analysis is more powerful when data are missing at random. Second, time is treated as continuous so that subjects do not have to be measured at the same time points. Third, covariates within the models can be time-varying or invariant (e.g., patient characteristics at baseline). Finally, these models can be used to estimate individual change for each subject, in addition to average change in

a population. This could be particularly useful in longitudinal studies where a proportion of subjects may exhibit change across time that deviates from the population average trend.

The paper is organized as follows. In Section 2, we introduce the mixed-effect regression model which will be used to analyze our data. Section 3 gives some background information for the SCr data set from an actual clinical study. The detailed data analysis is presented in Section 4 followed by residual analysis in Section 5. Section 6 describes the hypotheses testing and Section 7 states the conclusion and some discussions.

2. Mixed-effect Regression Model

In this section, we describe the basic longitudinal model, i.e., mixed-effect regression model which we will use to analyze post-transplant SCr data for transplant clinical studies. Let $Y_i = (Y_{i1}, ..., Y_{in_i})^T$ define the response (SCr measurements in our case) vector for subject i, $i = 1, ..., m$ over time. Then the mixed-effect regression model can be written as:

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, \quad i = 1, ..., m; \tag{2.1}$$

where

$$X_i = (x_{i1}^T, ..., x_{in_i}^T)^T$$

is an $n_i \times p$ design matrix for the fixed effects (where $x_{ij} = (x_{ij1}, ..., x_{ijp})^T$, $j = 1, ..., n_i$ is the covariate vector for the ith subject at the jth time); β is a $p \times 1$ vector of unknown fixed parameters; Z_i is a $n_i \times q$ design matrix for the random effects; $\gamma_i \sim N(0, G)$ is a $q \times 1$ vector of unknown random coefficients; and $\varepsilon_i \sim N(0, \sigma^2 I_{n_i})$ is a $n_i \times 1$ residual vector independent of γ_i. Then the conditional variance-covariance matrix for response is of the form:

$$\Sigma_{y_i} = Z_i G Z_i^T + \sigma^2 I_{n_i}. \tag{2.2}$$

So the observations Y_i and the random coefficients γ_i have a joint multivariate normal distribution:

$$\begin{pmatrix} Y_i \\ \gamma_i \end{pmatrix} \sim N\left(\begin{bmatrix} X_i\beta \\ 0 \end{bmatrix}, \begin{bmatrix} Z_iGZ_i^T + \sigma^2 I_{n_i} & Z_iG \\ GZ_i^T & G \end{bmatrix}\right). \tag{2.3}$$

In general, the formulation for mixed model can be written as:

$$y = X\beta + Z\gamma + \varepsilon, \tag{2.4}$$

with an assumption that

$$\begin{pmatrix} \gamma \\ \varepsilon \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\right). \tag{2.5}$$

Therefore, the variance of y is $V = ZGZ^T + R$. The structure of matrixes G and R is described in Subsection 4.2 although $R = \oplus_{i=1}^{m} R_i$ is a block diagonal matrix with a block R_i for subject i, where $\oplus$ denotes the direct sum [27].

For estimating G and R, the maximum likelihood (ML) or residual/ restricted maximum likelihood (REML) method will be utilized [5, 12, 20]. A favorable theoretical property of ML and REML is that they accommodate data with missing at random [16, 26].

After G and R are estimated, β and γ can be obtained by solving the equations:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}. \tag{2.6}$$

The ML method is used to select a parsimonious model by removing unnecessary terms one at a time; and the corresponding p-values are applied to testing fixed-effect parameters (Subsection 4.1). The REML method is then used instead of ML to select the covariance structure and to estimate covariance parameters (Subsection 4.2). The reason is that REML provides unbiased estimates of covariance parameters while ML estimates are biased.

For further theoretical background and applications regarding mixed effect model and longitudinal data analysis, see e.g., the reference books by Fitzmaurice et al. [8], Diggle et al. [6], Verbeke and Molenberghs [31], Pinheiro and Bates [25], etc.

3. Clinical Study and Data Description

Cyclosporine (Neoral®, CsA) allows for long-term survival for patients receiving solid organ transplants but cannot completely prevent rejection. Higher doses of cyclosporine used to improve efficacy can lead to increased side effects such as nephrotoxicity [1]. Everolimus (Certican®, RAD), a derivative of rapamycin, has a different mode of action than cyclosporine [7]. It does not appear to have the nephrotoxic side effects of cyclosporine [11]. In preclinical models of allotransplantation, the combination of everolimus and cyclosporine was more effective than either drug alone. However, the nephrotoxicity of the combination had been observed and hence required further assessments.

In a randomized, multicenter, open-label efficacy and safety study of Everolimus plus reduced cyclosporine versus mycophenolic acid plus cyclosporine in kidney transplant recipients (RAD001A2309, ClinicalTrials.gov Identifier: NCT00251004), twenty-four month post-transplant renal function data measured by SCr levels are obtained. Our purpose is to investigate the relationship between SCr measurements and three treatment groups described below:

A: RAD 1.5mg + Simulect® + Reduced-dose neoral® + Steroids;

B: RAD 3.0mg + Simulect® + Reduced-dose neoral® + Steroids;

C: Myfortic® 1.44g + Simulect® + Standard-dose neoral® + Steroids.

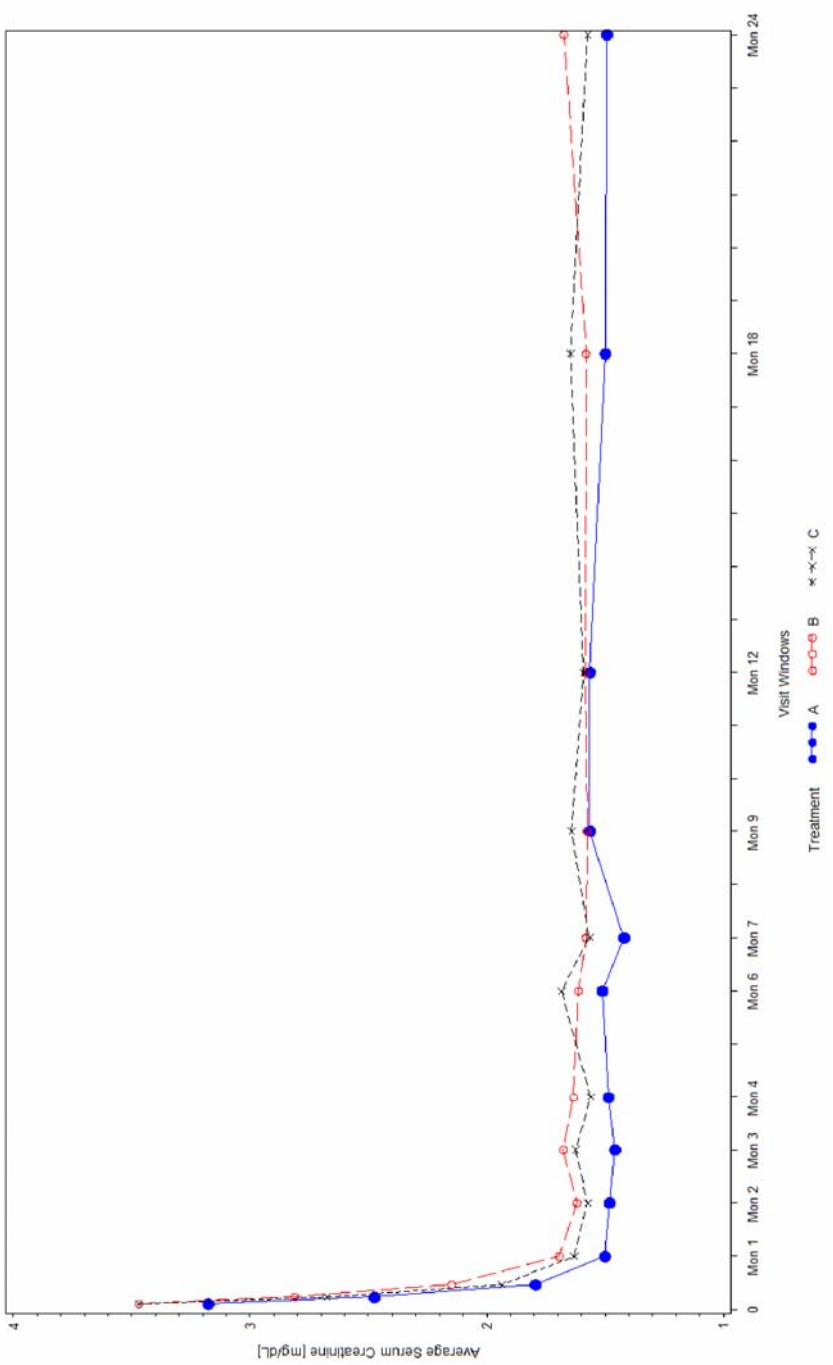

Figure 1. Average serum creatinine trough level across time by treatment [mg/dL].

The other covariates include: time; sex; age; height, weight, BMI (body mass index); diabetes (Yes/No); race; donor gender; donor age; and graft type (living donor vs. cadaveric donor) which are assumed to have possible effects on SCr levels. There are total of 792 patients (263 in Group A, 264 in Group B and 265 in Group C) who were randomized, had both SCr measurements and possible effects data. There are 9851 observations in total used in the analysis.

Figure 1 gives the average SCr level for each treatment group over each visit window. It indicates that SCr level decreases over time for each treatment group; and RAD 1.5mg group (treatment *A*) has consistently lower SCr level than other two groups.

4. Modeling the Creatinine Data

In this section, we fit the mixed regression model to the data. SAS (Statistical Analysis Software) 9.2 for windows is used for all analysis and referred to [15]. First we select the significant fixed effects including interactions which should be included in the model with the ML method. Then the REML method will be used to construct covariance structure and estimate parameters (see Subsection 4.2).

4.1. Fixed-effect selection

In order to identify the significant covariates (predictors) for the final model, we first fit a full model which includes all possible fixed-effects with a simpler variance-covariance structure. The fixed effects include time, treatments (*A*, *B*, *C*), weight, height, recipient and donor age (in years at baseline), recipient and donor sex, race, BMI, diabetes at baseline and graft type. For time variable, we consider day and logday (logarithm of the actual day of visit) in the full model since Figure 1 does not show the linear relationship between time and SCr level. Both day and logday are centralized at their mean values. The random effects include intercept, day and logday (both are centralized) since the objective of this analysis is to estimate the treatment effect on SCr level over time. The covariance structure for random

effects is assumed to be variance components (VC):

$$G = \begin{pmatrix} \sigma_I^2 & 0 & 0 \\ 0 & \sigma_D^2 & 0 \\ 0 & 0 & \sigma_{LD}^2 \end{pmatrix} \tag{4.1}$$

and the residual error is $R_i = \sigma^2 I_{n_i}$. Then we go through backward steps to obtain the appropriate predictors by deleting one fixed effect (with the largest type III test *p*-value, i.e., the least significant term) at each step followed by forward steps. Finally, the interaction items are checked and selected by applying similar backward/forward steps.

Backward elimination step. We begin from the full model with all possible fixed-effect terms in MODEL statement in PROC MIXED procedure. Then the term: donor gender (corresponding to the largest *p*-value: $p = 0.9256$) is eliminated and the model is refitted. Next, the term: weight (which has the largest *p*-value: $p = 0.9071$) is deleted, and so on. This backward step procedure continues until all remained effects in the current model have significant effects (i.e., type III *p*-value < Bonferroni-type adjustment significance level: 0.05/the number of fixed effects in the model) except for treatment which is always kept in the model for our purpose. Finally, donor gender, weight, height, and diabetes are removed. The last model from backward elimination step has 9 predictors. Except for Treatment $(p = 0.1669)$, the type III test *p*-values for other factors are less than adjusted significant level 0.00625 (= 0.05/8 for 8 tests in the model). Therefore, we claim that time, age, gender, race, BMI, donor age and graft type have significant effect on SCr measurements.

Forward addition step. Due to the possibility of complicated correlations among covariates, the backward step may delete the significant predictors before reaching the last model. Hence the forward selection step is necessary for capturing up significant terms that would have dropped during the backward elimination step.

During the first run of forward step, we refit four models each with one previously dropped term added. The results of the first run of forward step are shown in Table 1. It can be concluded from Table 1 that no additional term should be added back because all *p*-values for added terms are greater than Bonferroni-type adjusted significant level 0.05/9 = 0.0056.

Table 1. Type III test *p*-values for fixed effects (the first run of forward step)

Effects	Model 1	Model 2	Model 3	Model 4
Treatment	0.1508	0.1536	0.1848	0.1667
Day	<.0001	<.0001	<.0001	<.0001
Logday	<.0001	<.0001	<.0001	<.0001
Age	<.0001	<.0001	<.0001	<.0001
Donor age	<.0001	<.0001	<.0001	<.0001
Gender	<.0001	<.0001	<.0001	<.0001
BMI	<.0001	0.6120	<.0001	<.0001
Race	0.0004	0.0003	0.0001	0.0001
Graft type	<.0001	<.0001	<.0001	<.0001
Added term	0.0852	0.0956	0.1145	0.9572
Name of added term	Height	Weight	Diabetes	Donor gender

The forward addition step should continue until, among all models in one run, there are no models with all *p*-values less than the adjusted significant level.

Remarks. (1) If in one run of the forward step there are more than one model satisfying all *p*-values less than adjusted significance level, the AIC criterion could be used to decide the "best" model.

(2) It is possible that when adding a new term (var.b) to one model A and it is significant, an old term (var.a) which is originally significant becomes insignificant. Then a model B should be fitted with the new term (var.b)

instead of the old one (var.a) and AIC criterion should be used to choose the "better" model between A and B.

(3) Subject matter importance is always considered first such as treatment effect in clinical trials. Subject matter effects should always stay in the model even if they are insignificant.

Table 1 shows that when weight is added to the model, BMI becomes insignificant $(p = 0.6120)$. Following Remark (2) above, the new model with weight but without BMI is refitted $(\text{AIC} = 27915.7)$ and compared with original model (with BMI but without weight, $\text{AIC} = 27918.2$). Since new model has smaller AIC, the new model is selected for further analysis.

Table 2. Type III test *p*-values for fixed effects (the first run of interaction selection)

Effect	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Treatment	0.1591	0.1526	0.1935	0.3781	0.0809	0.2210	0.1005	0.6074
Day	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Logday	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Age	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Donor age	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Gender	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
BMI	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Race	0.0003	0.0003	0.0002	0.0004	0.0003	0.0004	0.0007	0.0003
Graft type	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Weight	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Interaction term	0.7018	0.7878	0.2747	0.1496	0.1073	0.8419	0.0048	0.8173
Name of added interaction term	Treatment *Day	Treatment *Logday	Treatment *Age	Treatment *Donor age	Treatment *Gender	Treatment *Race	Treatment *Graft type	Treatment *Weight

Interaction effect selection. In order to further investigate treatment effect, we consider possible treatment interaction with all other selected fixed effects in the last model from backward and forward steps. Analogous to

forward addition step, we add each interaction to the last model in the first run of interaction effect selection. The results are shown in Table 2. There is only one term: interaction between treatment and graft type with $p = 0.0048$ which is less than adjusted significant level 0.0056. Hence we add this interaction term to the model for covariance model selection.

4.2. Covariance model selection

Although the objective of this analysis usually is to compare treatment regression curves over time, modeling an appropriate covariance structure is essential so that the inferences about means are valid. In the above process, we assume independent errors within a subject (patient). That means all observations within a subject are equally correlated $R_i = \sigma^2 I_{n_i}$. However, it is more reasonable to assume that two measures taken at adjacent times are more highly correlated than two measurements taken several time points apart. In this section, we explore the appropriate covariance structure for our SCr data. From this point on, we use REML method in order to obtain unbiased estimates of covariance parameters.

For random effect covariance matrix *G* corresponding to intercept and time, except for the VC structure used above in (4.1), we also consider unstructured (UN) structure, that is,

$$G = \begin{pmatrix} \sigma_I^2 & \sigma_{ID} & \sigma_{ILD} \\ \sigma_{ID} & \sigma_D^2 & \sigma_{DLD} \\ \sigma_{ILD} & \sigma_{DLD} & \sigma_{LD}^2 \end{pmatrix}, \tag{4.2}$$

where σ_{ID}, σ_{ILD} and σ_{DLD} are the covariances between intercept and day, intercept and logday and day and logday, respectively.

For residual error covariance matrix *R*, the following structures are considered in addition to the homogeneous (HOMO) one, i.e., $R_i = \sigma^2 I_{n_i}$. Suppose $n_i = 3$ for the simplicity of the formula.

First-order autoregressive (AR(1)). It has homogenous variances. The correlation between any two adjacent measurements is equal to ρ with $|\rho| < 1$ for stationary although exact time periods are not equal between any adjacent observations. The correlation between any two elements separated by a third is ρ^2, and so on.

$$R_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}. \tag{4.3}$$

First-order autoregressive moving average (ARMA(1, 1)). It has homogenous variances. The correlation between two adjacent elements is $\theta\rho$, between elements separated by a third is $\theta\rho^2$, and so on. ρ and θ are the autoregressive and moving average parameters, respectively, with $|\rho| < 1$ and $|\theta| < 1$ for stationary.

$$R_i = \sigma^2 \begin{pmatrix} 1 & \theta\rho & \theta\rho^2 \\ \theta\rho & 1 & \theta\rho \\ \theta\rho^2 & \theta\rho & 1 \end{pmatrix}. \tag{4.4}$$

Toeplitz with two bands structure (TOEP(2)). Homogenous variances. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements which are not adjacent is 0.

$$R_i = \sigma^2 \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.5}$$

Spatial power structure (SP(POW)). Let d_{ij} be the Euclidean distance between *i*th and *j*th observational times (i.e., logday in our model), the covariance between *i*th and *j*th elements is $\sigma^2\rho^{dij}$.

$$R_i = \sigma^2 \begin{pmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} \\ \rho^{d_{12}} & 1 & \rho^{d_{23}} \\ \rho^{d_{13}} & \rho^{d_{23}} & 1 \end{pmatrix}. \quad (4.6)$$

Table 3 shows the AIC and BIC (a smaller value suggests a better fit) fit statistics and their ranks for 10 covariance structure models described above. The last row in Table 3 presents the number of variance-covariance parameters. From Table 3, one could easily identify that model 6 (rank 10 for both AIC and BIC) is the "best" choice when considering both AIC and BIC statistics jointly. Models 4 and 10 have almost the same AIC/BIC and are the second to the best choice.

Table 3. AIC and BIC fit statistics for covariance structure models

Fixed	**HOMO**		**AR(1)**		**ARMA(1, 1)**		**TOEP(2)**		**SP(POW)**	
Random	VC	UN	VC	UN	VC	UN	VC	UN	VC	UN
Model	1	2	3	4	5	6	7	8	9	10
AIC (rank)	27959.4 (1)	26569.6 (6)	27692.2 (4)	26387.9 (9)	27667.2 (5)	26362.0 (10)	27853.2 (2)	26481.7 (7)	27692.4 (3)	26388.0 (8)
BIC (rank)	27978.1 (1)	26602.3 (6)	27715.6 (4)	26425.3 (9)	27695.3 (5)	26404.0 (10)	27876.6 (2)	26519.1 (7)	27715.8 (3)	26425.4 (8)
Parm.#	4	7	5	8	6	9	5	8	5	8

By checking outputs from Model 6, we find that fixed effect race is no longer significant $(p = 0.1732)$. Same phenomenon occurs in Models 4 and 10. Therefore, we remove Race from three models and fit the data again. Outputs show that updated Models 4 and 10 still have almost the same AIC values (26386.4 and 26386.6 for Models 4 and 10, respectively) but updated Model 6 has larger AIC value (27675.5). Since times in our data are not equally spaced and subjects have the different observation times, we finally choose Model 10 which does not require equally spaced times. Hence, in the final model, fixed effects include treatment, actual day, logday, age, gender, weight, donor age, graft type, and interaction term: treatment*graft type. The covariance structure is spatial power (SP(POW)) for residual error and

unstructured (UN) structure for random-effects (i.e., intercept, day, logday). The model is as follows:

$$\begin{aligned}\text{SCr} &= 0.00205 * (\text{Day} - 181) - 0.46 * (\text{Logday} - 4.166) \\ &\quad - 0.01 * \text{Age} + 0.007 * \text{Weight} + 0.33(\text{Male}) + 0.013 * \text{Donor age} \\ &\quad + \begin{cases} 0.9444 + 0.217(\text{Cadeveric graft}); & \text{for treatment } A; \\ 1.0536 + 0.331(\text{Cadeveric graft}); & \text{for treatment } B; \\ 1.1487 - 0.008(\text{Cadeveric graft}); & \text{for treatment } C. \end{cases} \end{aligned} \tag{4.7}$$

Tables 4.1 and 4.2 show the covariance parameter estimates and type III *p*-values for fixed effects from final model (4.7), respectively. All covariance parameters are significant $(p < 0.0001)$ except for the variance of day which is almost 0 (0.000009175).

Table 4.1. Final model fitted results – covariance parameter estimates

Parameter	**Estimate**	**Std. err.**	***z*-value**	***p*-value**	**95% confidence interval**
Intercept σ_I^2	0.972	0.566	17.18	<.0001	(0.870, 1.093)
σ_{ID}	0.0024	0.0002	13.89	<.0001	(0.0021, 0.0028)
Day σ_D^2	9.175E-6	0.0000	.	.	.
σ_{ILD}	–0.530	0.037	–14.43	<.0001	(–0.602, –0.458)
σ_{DLD}	–0.0022	0.0001	–16.05	<.0001	(–0.0025, –0.0020)
Logday σ_{LD}^2	0.594	0.033	17.95	<.0001	(0.534, 0.664)
SP(POW) ρ	0.526	0.019	27.27	<.0001	(0.488, 0.563)
Residual σ^2	0.551	0.010	57.67	<.0001	(0.533, 0.570)

Table 4.2. Final model fitted results – type III test *p*-values for fixed effects

Fixed effect	Treatment	Day	Logday	Age	Donor age	Gender	Weight	Graft type	Treatment* Graft type
***p*-value**	0.0371	<0.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0010	0.0282

5. Residual Analysis

Residuals are commonly used to evaluate the validity of the assumptions of statistics models. There are three types of residuals for mixed-effect regression model that accommodate the extra source of variability [21]:

- *Marginal residual*: $\hat{\xi} = y - X\hat{\beta}$, that predicts the marginal errors, $\xi = y - X\beta = Z\gamma + \varepsilon$.

- *Conditional residual*: $\hat{\varepsilon} = y - X\hat{\beta} - Z\hat{\gamma}$, that predicts the conditional errors, $\varepsilon = y - X\beta - Z\gamma$.

- *Empirical best linear unbiased predictor* (*EBLUP*): $Z\hat{\gamma}$, that predicts the random effects, $Z\gamma$.

A review of residual analysis for linear mixed models could be found in [21] and the references cited therein. Here we discuss all these residuals for our final fitted model (4.7).

5.1. Marginal residuals

Marginal residuals $\hat{\xi} = (\hat{\xi}_1^T, ..., \hat{\xi}_m^T)^T$ could be employed to check the linearity of y with respect to fixed effects as the usual residuals in standard linear model. A random behavior around 0 is expected when the linear relationship holds. For our final fitted model (4.7), marginal residuals are plotted in Figure 2 showing residuals versus logday, age, donor age and weight, respectively. These plots support the regression model for SCr since there is no nonlinear trend identified from these plots.

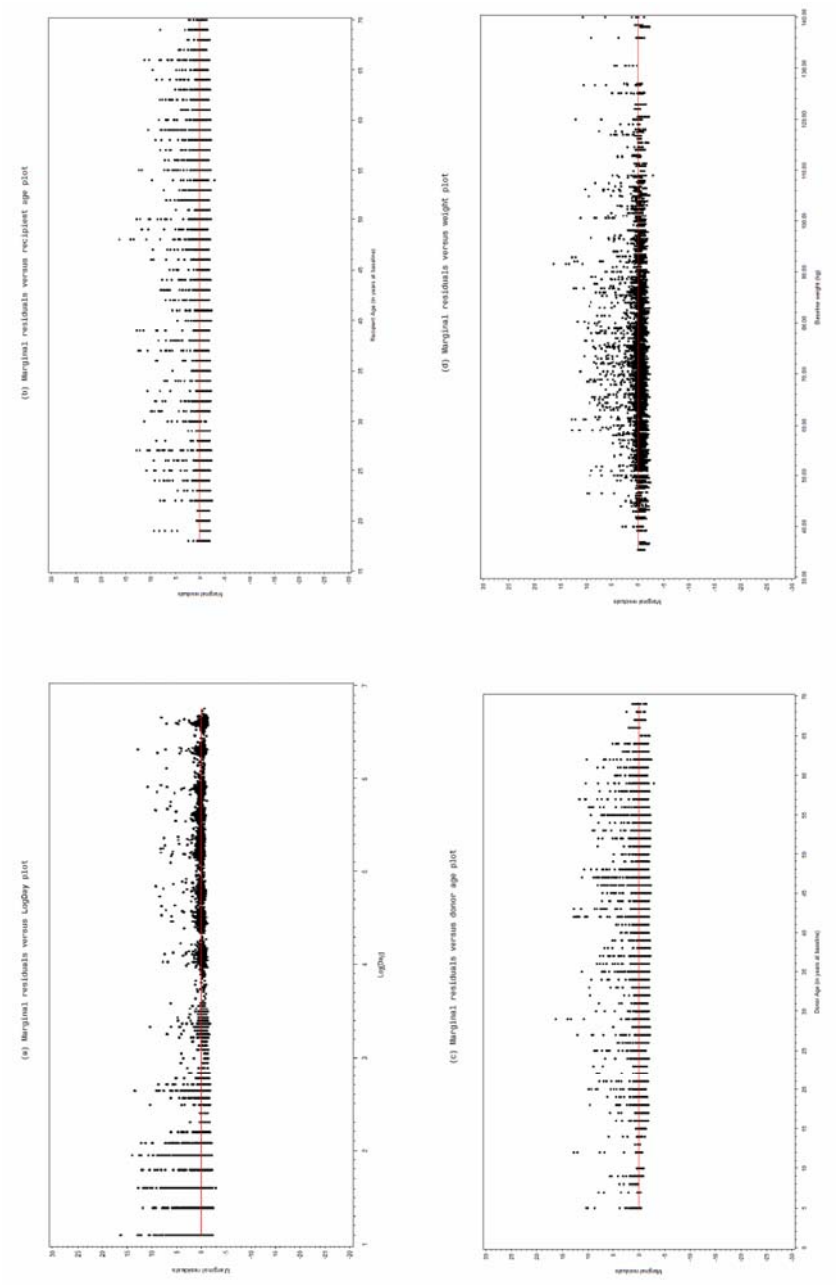

Figure 2. Marginal residuals versus logday, recipient age, donor age and weight plots.

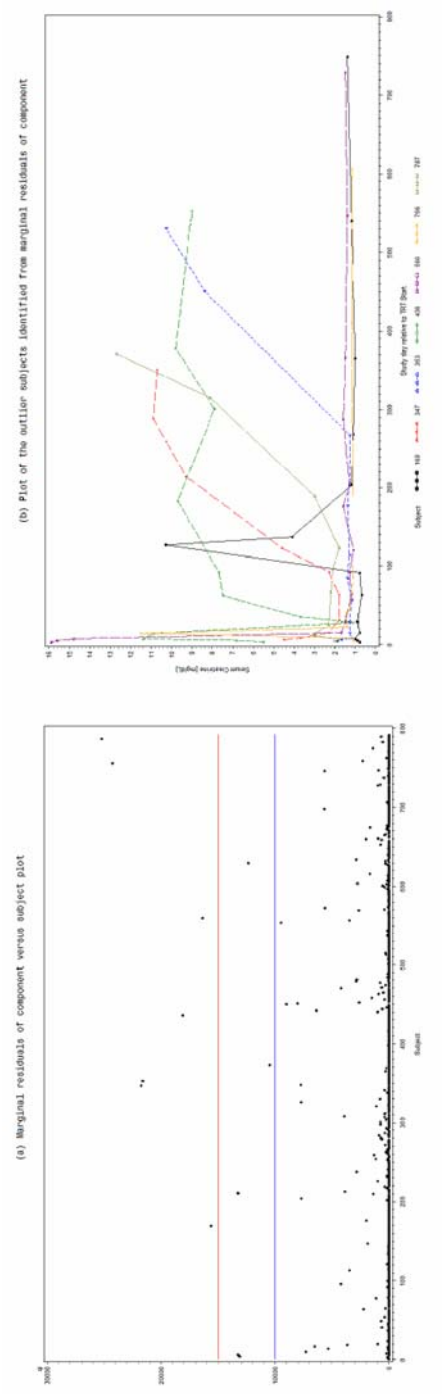

Figure 3. Marginal residuals of component q_i versus subject plot (a) and outlier subjects identified from marginal residuals of component q_i(b).

Marginal residuals may also be used to check the validity of the within-subject covariance structure, i.e., $V_i = Z_i G Z_i^T + R_i$. Let $Q_i = \hat{V}_i^{-1/2}\hat{\xi}_i$, Verbeke and Lesaffre [30] suggest using the interpretable component $q_i = \| I_{n_i} - Q_i Q_i^T \|^2$ as a diagnostic for the within-subject covariance matrix, where $\| A \|$ denotes the Forbenius norm of matrix A. q_i is expected to be close to zero and a plot of such values versus the subject indices is useful in identifying outlier subjects for which the assumed covariance structure does not fit well. Figure 3(a) of the final model residuals for the structure of the covariance matrix versus subject shows several outlier subjects suggesting that the fitted covariance matrix is not adequate for these subjects. The source SCr data for the outlier subjects are presented in Figure 3(b). It is easy to see that data from these subjects deviate from overall mean data trend in Figure 1.

Remarks. In practice, one should further refine the final model by checking and removing obvious outlier subjects or observations both clinically and statistically. Those clinically possible observations should not be considered as outliers even if they are identified from above residual analysis.

5.2. Conditional residuals

The final model conditional error is $\varepsilon_{ij} \sim N(0, R_i)$, where $R_i = SP(POW)$. Hence the Pearson-type standardized conditional residuals are $\hat{\varepsilon}_{ij}/\hat{\sigma}_j \sim N(0, 1)$ approximately. Plots of the elements of $\hat{\varepsilon}/\hat{\sigma}$ versus subject ID (Figure 4(a)) or predicted SCr: $\hat{y} = X\hat{\beta} + Z\hat{\gamma}$ (Figure 4(b)) and histogram and Q-Q plots of $\hat{\varepsilon}/\hat{\sigma}$ (Figures 4(c) and 4(d)) may be utilized to check homoscedasticity and normality of the conditional error ε [22, 32].

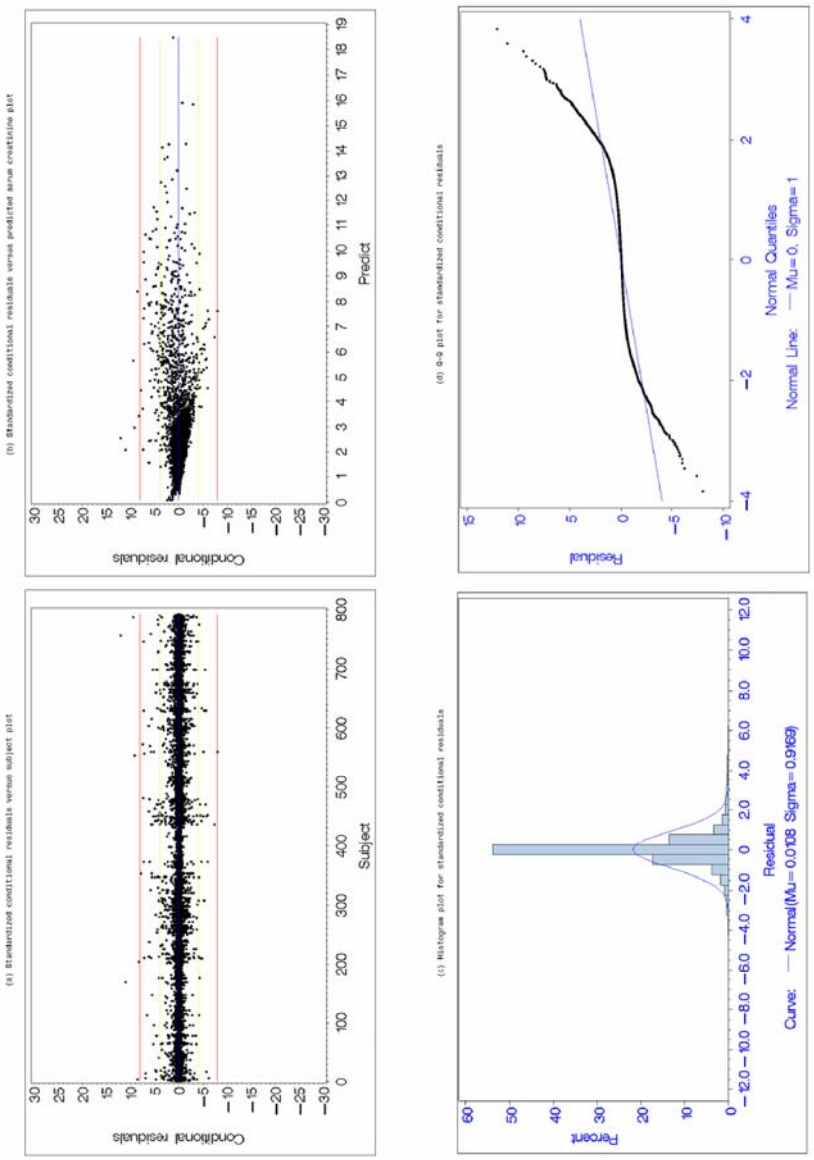

Figure 4. Standardized conditional residual plots.

Moreover, one may check studentized conditional residuals which are simply defined as: $\hat{\varepsilon}_{ij} / \sqrt{\hat{V}ar(\varepsilon_{ij})} \sim N(0, 1)$ approximately [9]. Plots of studentized conditional residuals are shown in Figure 5. Similar to marginal

residuals, these figures for conditional residuals also indicate several outlier subjects and hence there exist deviations from normal assumptions for those outliers.

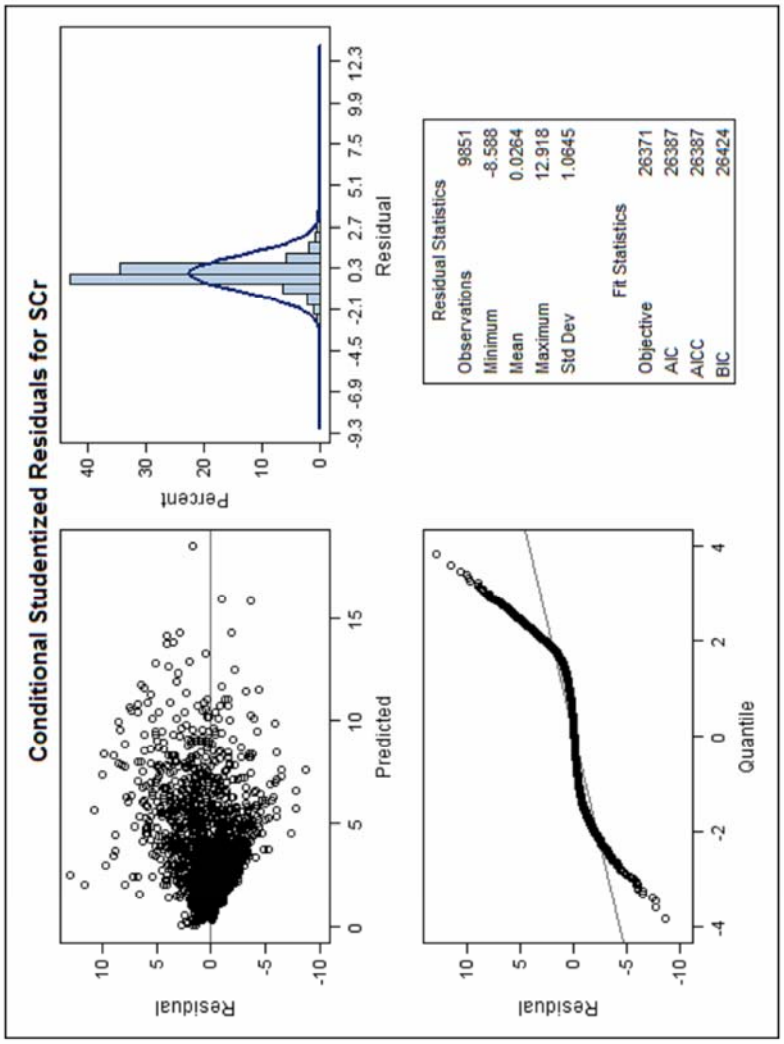

Figure 5. Studentized conditional residual plots.

5.3. EBLUP

Applying estimated covariance parameters to predict γ, we obtained empirical best linear unbiased predictor $\hat{\gamma}$. Then $Z_i\hat{\gamma}_i$ reflects the difference

between the predicted responses for the *i*th subject and the population average; therefore it could be used to identify outlying subjects as suggested in [19, 30], etc. Plot (not shown) of $Z_i\hat{\gamma}_i$ versus subject for model (4.7) indicates that outlying subjects could be identified.

One can also check the normality of $\hat{\gamma}$ by plotting the standardized $\hat{\gamma}_i$ versus subject or by *Q*-*Q* plot of standardized $\hat{\gamma}_i$ based on standard normal distribution as we did above for conditional residuals. These plots (not shown) indicate that $\hat{\gamma}_i$ follows a normal distribution approximately.

5.4. Influence diagnostics

In order to check whether any subjects have large influence on the final model, a non-iterative influence analysis is performed to the model. Figure 6 presented plots for restricted likelihood distance (RLD), Cook's distance and covariance ratio (CR), and root mean squared error (RMSE) when delete one of subjects. The formulas for these parameters could be found in the SAS support website for residuals and influence diagnostics:

http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect024.htm

Together with the influence diagnostic table (not shown), we could identify the following outlier subjects from these plots: subject #436 (maximum RLD = 3.126, Cook = 0.055, minimum CR = 0.777, minimum RMSE=0.734); #4 (RLD = 2.04, CR = 0.815, RMSE = 0.7356); #6 (CR = 0.821, RMSE = 0.7359); #10 (maximum Cook = 0.091); #19 (Cook = 0.058); and #326 (Cook = 0.052). The overall effects of outlier subjects are minor based on total 792 subjects and relatively smaller values for these extreme observations from influence analysis. Of course, deleting these outlier subjects (as a set of values from all visits) will increase the estimated precision of the fixed effects estimates although just minor.

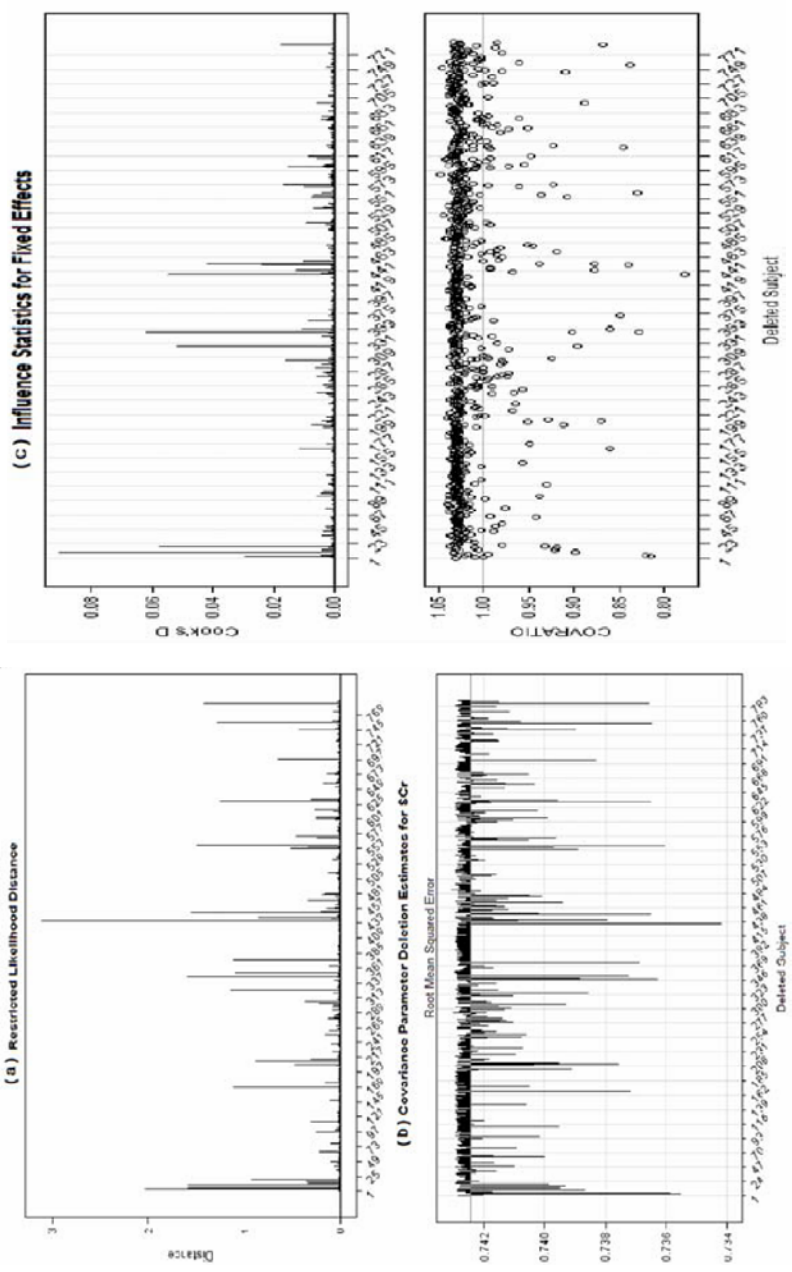

Figure 6. Influence diagnostics plots.

6. Model Based Testing

The main purpose for this analysis is to characterize as well as to test

treatment effect on renal function measurement: SCr concentration level. From the analysis above, one can conclude that not only does treatment have a significant effect on SCr, so do the other covariates such as day, age and weight. Obviously, the general *Z*-test or *T*-test for treatment effect will not remove other covariate effects and so the test results are not reliable. The model based testing is then a better choice since it accounts for and eliminates the other covariate effects involved in the model. Of course, the model itself should be trustworthy first (see Section 5).

With a mixed-effect regression model, one can test the null hypothesis $L'\phi = 0$, where $L' = (K'\ M')$ and $\phi' = (\beta'\ \gamma')$. The test statistic is

$$F = \frac{\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}^T L(L'\hat{C}L)^{-1}L'\begin{bmatrix}\hat{\beta}\\ \hat{\gamma}\end{bmatrix}}{rank(L'\hat{C}L)}$$

with an *F*-distribution under the null hypothesis, where

$$\hat{C} = \begin{pmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1}\end{pmatrix}^{-}$$

is the approximate variance-covariance matrix of $(\hat{\beta} - \beta, \hat{\gamma} - \gamma)$. For fixed effect (the γ portion of L is assumed to contain all 0s), the denominator degree of freedom for a test is calculated as the minimum contribution to the rank of $(X\ Z)$ of random effects containing the fixed effect of interest. For random effect, because any linear combination of γ is estimable, the χ^2 statistic associated with the likelihood ratio test based on REML shall be applied.

Table 5 shows testing results for treatment effects on SCr based on the model (4.7), where label "$A = B$" refers to the null hypothesis that treatments *A* and *B* have the same effect on SCr, etc. We can see from Table 5 that treatments *A* and *B* have significantly different effects on SCr $(p = 0.0104)$, but treatment *C* has no significant difference from treatment *A* $(p = 0.1554)$ and treatment *B* $(p = 0.2493)$.

Table 5. Model based hypothesis testing for treatment difference

Test	$A = B$	$A = C$	$B = C$	$A = B = C$
p-value	0.0104	0.1544	0.2493	0.0371

Table 6 shows least square mean of SCr for each treatment group over time. The least square mean differences are $A - B = -0.17$; $A - C = -0.09$ and $B - C = 0.07$. Together with Table 5, we conclude that treatment A has the best renal function (smaller SCr means better renal function); it is significantly better than treatment B and numerically better than treatment C, the control group.

Table 6. Least square mean of SCr for each treatment over time [mg/dL]

Day	3	15	30	60	90	120	180	240	360	480	600	720
RAD 1.5mg	2.86	2.15	1.86	1.60	1.48	1.41	1.34	1.33	1.39	1.51	1.65	1.81
RAD 3.0mg	3.03	2.31	2.02	1.77	1.64	1.57	1.51	1.50	1.56	1.67	1.82	1.98
Myfortic	2.95	2.24	1.95	1.69	1.57	1.50	1.43	1.42	1.48	1.60	1.74	1.91

Notice that there is an interaction between treatment and graft type (living donor versus deceased donor), treatment effect on SCr shall be different for different type of donor. Table 7 shows least square mean difference of SCr by graft type. The table indicates that treatment A is significantly better than control group for living donor type $(p = 0.0188)$ but no significant difference between A and C for deceased donor type $(p = 0.8275)$. Treatment A is significantly better than treatment B for deceased donor type and numerically better than treatment B for living donor type.

Table 7. Least square mean difference of SCr and test by graft type [mg/dL]

Test	$A - B$	$A - C$	$B - C$
Living donor	–0.1092 (p = 0.2108)	–0.2403 (p = 0.0188)	–0.0951 (p = 0.2733)
Deceased donor	–0.2232 (p = 0.0204)	0.0208 (p = 0.8275)	0.2439 (p = 0.0109)

7. Conclusion and Discussion

In transplant clinical studies, longitudinal data are frequently collected for monitoring patient safety and treatment effects. However, there is rarely statistical analysis applying longitudinal data analysis strategies in clinical trials [24]. Another explanation to this phenomenon could be the lack of a standard methodology for analysis, diagnosis, and interpretation. The paper is trying to provide a systematic approach of longitudinal data analysis during clinical trials. We present a brief review of longitudinal data analysis strategies and applied them to post-transplant serum creatinine data. The proposed strategies use the iteration of fixed effects selection and covariance analysis followed by residual analysis, testing, prediction and interpretation.

Figure 7 shows the average predicted serum creatinine values across time for each treatment from the final model. Comparing with Figure 1, together with residual analysis outputs in Section 5, one could see that the fitted model is good for our data in spite of several outliers. When all outlier subjects identified in Section 5 are removed and data are refitted with the same model structure, there is a little change to outputs. The least square mean difference $A - B$ changes from –0.17 to -0.16 $(p = 0.0008)$; $A - C$ from –0.09 to –0.06 $(p = 0.1721)$; and $B - C$ from 0.07 to 0.09 $(p = 0.0447)$.

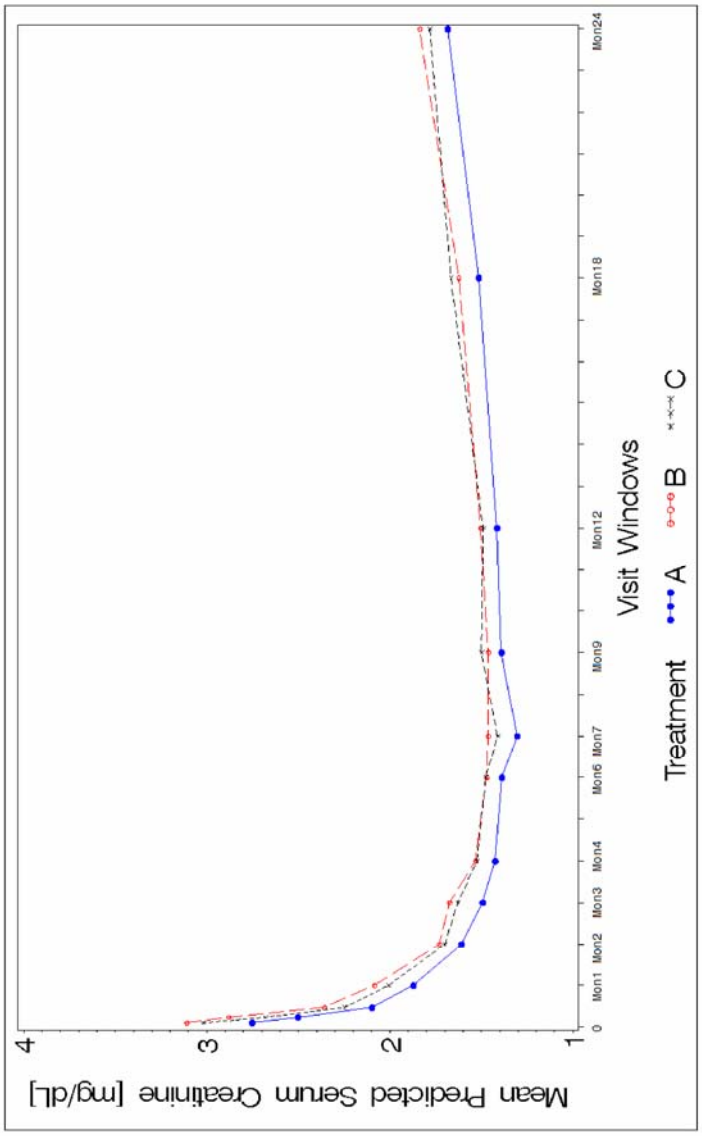

Figure 7. Average predicted SCr trough levels across time by treatment [mg/dL] - final model.

Hence we could make the following conclusions based on our fitted model and testing results:

(1) Treatment group, day (since start of study medication), logday, age, weight (at baseline), donor age and graft type have significant effects on serum creatinine measurements.

(2) Graft type has an interaction effect with treatment: treatment *A* is significantly better than treatment *B* for deceased donor $(p = 0.02)$ and treatment *C* for living donor $(p = 0.02)$. It is also numerically better than treatments *B* for living donor.

(3) Overall, treatment *A* is significantly better than *B* (decreasing SCr by 0.17 [mg/dL]) and numerically better than *C* (decreasing SCr by 0.09 [mg/dL]). Treatment *B* is worse than *C*.

(4) Outliers could be identified from the model but they have minor effect to estimates and no effect on the conclusion.

The outputs confirm that the use of RAD 1.5mg with reduced-dose Neoral is associated with improved renal function for kidney transplant recipients but RAD 3.0mg is not. Therefore, RAD 1.5mg plus reduced (or minimized) cyclosporine regime is widely investigated in organ transplant population to improve renal function. The same analysis strategy was applied to other endpoints in study RAD A2309 [29, 28, 2] and confirmed that renal function (GFR) in RAD 1.5mg + reduced cyclosporine group is better than active control group. Hence, among all others, this exemplary longitudinal data analysis strategy could be applied to any clinical observations measured by time.

Another topic that the author is working on is trying to incorporate a Bayesian approach into longitudinal data analysis. The basic idea is first to obtain information/model from similar previous studies. The model should then be modified based on current study data.

References

[1] W. M. Bennett, The nephrotoxicity of immunosuppressive drugs, Clinical Nephrology 43 (suppl. 1) (1995), S3-S7.

[2] D. Cibrik, H. T. Silva et al., Randomized trial of everolimus-facilitated calcineurin inhibitor minimization over 24 months in renal transplantation, Transplantation 95(7) (2013), 933-942.

[3] D. W. Cockcroft and M. H. Gault, Prediction of creatinine clearance from creatinine, Nephron. 16 (1976), 31-41.

[4] J. Deleeuw and I. Kreft, Random coefficient models for multilevel analysis, J. Edu. Stat. 11 (1986), 57-85.

[5] A. P. Dempster, D. B. Rubin and R. K. Tsutakawa, Estimation in covariance components models, JASA 76 (1981), 341-353.

[6] P. Diggle, P. Heagerty, K. Y. Liang and S. Zeger, Analysis of Longitudinal Data, 2nd ed., Oxford University Press, Oxford, 2002.

[7] H. J. Eisen, E. M. Tuzcu et al., Everolimus for the prevention of allograft rejection and vasculopathy in cardiac-transplant recipients, The New England Journal of Medicine 349(9) (2003), 847-858.

[8] G. Fitzmaurice, N. Laird and J. Ware, Applied Longitudinal Analysis, John Wiley and Sons, Inc., Hoboken, 2004.

[9] D. M. Hawkins and R. A. J. Wixley, A note of the transformation of chi-squared variables to normality, The American Statistician 40 (1986), 296-298.

[10] D. Hedeker and R. J. Mermelstein, Application of random-effects regression models in relapse research, Addiction 91 (Supplement) (1996), S211-S229.

[11] B. D. Kahan, J. Y. Chang and S. N. Sehgal, Preclinical evaluation of a new potent immunosuppressive agent, rapamycin, Transplantation 52 (1991), 185-191.

[12] N. M. Laird and J. H. Ware, Random-effects models for longitudinal data, Biometrics 38 (1982), 963-974.

[13] A. S. Levey, R. L. Berg, J. J. Gassman, P. M. Hall and W. G. Walker, Creatinine filtration, secretion and excretion during progressive renal disease, Modification of Diet in Renal Disease (MDRD) Study Group, Kidney Int Suppl. 27 (1989), S73-S80.

[14] A. S. Levey, J. P. Bosch, J. B. Lewis, T. Greene, N. Rogers and D. Roth, A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation, Modification of Diet in Renal Disease Study Group, Ann. Intern. Med. 130(6) (1999), 461-470.

[15] R. C. Littell, G. A. Milliken, W. W. Stroup, R. D. Wolfinger and O. Schabenberger, SAS for Mixed Models, 2nd ed., SAS Institute Inc., Cary, NC, 2006.

[16] R. J. A. Little, Modeling the drop-out mechanism in repeated-measures studies, JASA 90 (1995), 1112-1121.

[17] N. T. Longford, A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, Biometrika 74 (1987), 817-827.

[18] N. T. Longford, Random Coefficient Models, Oxford University Press, New York, 1993.

[19] N. T. Longford, Simulation-based diagnostics in random-coefficient models, J. Roy. Statist. Soc. A 164 (2001), 259-273.

[20] C. N. Morris, Parametric empirical Bayes inference: theory and applications, JASA 78 (1983), 47-55.

[21] J. S. Nobre and J. M. Singer, Residual analysis for linear mixed models, Biom. J. 49 (2007), 863-875.

[22] S. D. Oman, Checking the assumptions in mixed-model analysis of variance: a residual analysis approach, Comput. Statist. Data Anal. 20 (1995), 309-330.

[23] R. D. Perrone, N. E. Madias and A. S. Levey, Serum creatinine as an index of renal function: new insights into old concepts, Clin. Chem. 38 (1992), 1933-1953.

[24] S. Piantadosi, Clinical Trials: A Methodologic Perspective, 2nd ed., John Wiley and Sons, Inc., Hoboken, 2005.

[25] J. Pinheiro and D. Bates, Mixed-effect Models in S and S-PLUS, Springer, New York, 2000.

[26] D. B. Rubin, Inference and missing data, Biometrika 63(3) (1976), 581-592.

[27] S. R. Searle, Matrix Algebra Useful for Statistics, Wiley, New York, 1982.

[28] F. S. Shihab, D. Cibrik et al., Association of clinical events with everolimus exposure in kidney transplant patients receiving reduced cyclosporine, Clin. Transplant. 27 (2012), 217-226.

[29] H. T. Silva, D. Cibrik et al., Everolimus plus reduced-exposure CsA versus mycophenolic acid plus standard-exposure CsA in renal-transplant recipients, American Journal of Transplantation 10 (2010), 1401-1413.

[30] G. Verbeke and E. Lesaffre, A linear mixed-effect model with heterogeneity in the random-effects population, JASA 91 (1996), 217-221.

[31] G. Verbeke and G. Molenberghs, Linear Mixed Models for Longitudinal Data, Springer-Verlag, New York, 2000.

[32] R. E. Weiss and C. G. Lazaro, Residual plots for repeated measures, Statistics in Medicine 11 (1992), 115-124.