Journal of Modern Mathematics and Statistics 3 (3): 60-68, 2009
ISSN: 1994-5388

Constitution of Linear Mixed Models (LMMs) in the Analysis of Correlated Data: Random Intercept Model (RIM) for Repeated Measurements Data

Neslihan Iyit and Asir Genc
Department of Statistics, Faculty of Science, Selcuk University, 42031 Konya, Turkey

Abstract: In this study, a Random Intercept Model (RIM) as a special case of Linear Mixed Models (LMMs) implying a Compound Symmetry (CS) variance-covariance structure assumption that each pair of repeated measurements has the same correlation, variance and covariance terms is constituted to a repeated measurements data set obtained from a clinical trial. The superiority of Random Intercept Model (RIM) bringing about the advantage of modeling heterogeneity between subjects than General Linear Model (GLM) for repeated measurements data implying a Variance Components (VC) variance-covariance structure assumption that each pair of repeated measurements are uncorrelated and have constant variance is emphasized.

Key words: Repeated measurement, linear mixed model, random intercept model, compound symmetry variance-covariance structure, generalized least squares method

INTRODUCTION

A repeated measurements design is one, in which at least one of the factors consists of repeated measurements on the same subjects or experimental units, under different conditions. A factor consisting of repeated measurements on the same subjects or experimental units, under different conditions is commonly called a within-subjects factor. A between subjects factor is one, in which each level of the factor contains different subjects or experimental units (Hamer and Simpson, 1989; Iyit *et al*., 2006; Iyit and Genc, 2007).

A model, where we have both fixed effects and random effects is called mixed model because of the mixture of different types of these effects on response variable. Fixed effects are the effects attributable to a finite set of levels of a factor on response variable representing all possible levels of the variable, in which inferences are to be made. On the other hand, random effects are the ones attributable to an infinite set of levels of a factor on response variable of which, only a random sample of potential levels of the factor is taken to draw inferences for the complete population of levels (Davis, 2002; Iyit and Genc, 2005a).

In mixed models fixed effects are used to explain the expected value of the observations and random effects to explain the variance-covariance structure of the dependent variable. If the relationship between the observations of the response variable and these effects is linear, then the model is called Linear Mixed Model (LMM) (Hamer and Simpson, 1989; Davis, 2002; Iyit *et al*., 2006).

The Linear Mixed Model (LMM) approach to repeated measurements allows explicit modeling and analysis of the variation between-subjects and within-subjects factors. Linear Mixed Models (LMMs) provide a tool for analyzing repeated measurements data by taking into consideration these 2 types of variability as well as the linear relationship between the explanatory variables and the response variable (McCulloch and Searle, 2001; Fitzmaurice *et al*., 2004; Iyit and Genc, 2007).

The theoretical base of Linear Mixed Models (LMMs) is well-established and the methodology has applications in many areas not only involving repeated measurements. McLean *et al*. (1991) provide a general introduction to Linear Mixed Models (LMMs) and Ware (1985) gives an overview of their application to the analysis of repeated measurements. Raudenbush and Bryk (2002) investigate Random Intercept Model (RIM) as a special case of Linear Mixed Models (LMMs) in details. Bryk and Driscoll (1988), use this model type to examine how characteristics of school organization are related to teachers' sense of efficacy in their research in the field of education.

Iyit and Genc (2005a, b, 2007) and Iyit *et al*. (2006) compare General Linear Model (GLM) implying Variance Components (VC) variance-covariance structure with Random Intercept Model (RIM) implying Compound Symmetry (CS) variance-covariance structure and Random Intercept and Slope Model (RISM) implying

Corresponding Author: Neslihan Iyit, Department of Statistics, Faculty of Science, Selcuk University, 42031 Konya, Turkey

flexible various variance-covariance structure forms of the vector of response variables representing repeated measurements data as a special case of Linear Mixed Models (LMMs). They also show, the superiority of Random Intercept and Slope Model (RISM) allows to modeling possible heterogeneity in intercepts and in slopes of the individual's own regression line for repeated measurements data.

In this study, Random Intercept Model (RIM) as a special case of Linear Mixed Models (LMMs) implying a Compound Symmetry (CS) variance-covariance structure assumption that each pair of repeated measurements has the same correlation, variance and covariance terms is constituted to a repeated measurements data set obtained from a clinical trial. The superiority of Random Intercept Model (RIM) bringing about the advantage of modeling heterogeneity between subjects than General Linear Model (GLM) for repeated measurements data implying a Variance Components (VC) variance-covariance structure assumption that each pair of repeated measurements are uncorrelated and have constant variance is emphasized.

This study is organized as follows; in study, a general introduction to Random Intercept Model (RIM) as a special case of Linear Mixed Models (LMMs) for repeated measurements is given. In study, an application of a statistical analysis of repeated measurements data with Random Intercept Model (RIM) is considered. Results and discussion also conclusion parts obtained from the analysis with this procedure are given in study.

RANDOM INTERCEPT MODEL (RIM) AS A SPECIAL CASE OF LINEAR MIXED MODELS (LMMs) FOR REPEATED MEASUREMENTS

Let, $\underline{Y}_i$; $i = 1,2,\ldots,N$ be the $n_i \times 1$ dimensional vector of response variables representing all repeated measurements taken on ith subject:

$$\underline{Y}_i = \left(Y_{i1}, Y_{i2}, \ldots, Y_{in_i}\right)'; \quad i = 1,2,\ldots,N \tag{1}$$

Laird and Ware (1982) consider the Linear Mixed Model (LMM) equation for repeated measurements data in vector-matrix notation as follows:

$$\underline{Y}_i = X_i\underline{\beta} + Z_i\underline{u}_i + \underline{\varepsilon}_i;\ 1 \le i \le N \tag{2}$$

Where:

X_i = The $n_i \times (p + 1)$ dimensional design matrix for fixed effects parameter vector belonging to ith subject

$\underline{\beta}$ = The $(p + 1) \times 1$ dimensional parameter vector for fixed effects

Z_i = The $n_i \times q_i$ dimensional design matrix for random effects vector belonging to ith subject

$\underline{u}_i$ = The $q_i \times 1$ dimensional unobserved random effects vector belonging to ith subject

$\underline{\varepsilon}_i$ = The $n_i \times 1$ dimensional vector of random error terms belonging to ith subject (Fitzmaurice *et al*., 2004; Iyit, 2008)

From Eq. 2, it can be easily seen that GLM is a special case of LMM when $Z_i = 0$ and $R_i = C\,ov\,(\underline{\varepsilon}_i) = \sigma^2_e I_{ni}$ where, $\underline{Y}_i$; the $n_i \times 1$ dimensional vector of response variables representing all repeated measurements taken on ith subject has Variance Components (VC) variance-covariance structure as given Eq. 3:

$$C\,ov\left(\underline{Y}_i\right) = V_i = R_i;\ i = 1,2,\ldots,N = \begin{bmatrix} \sigma^2_e & 0 & \cdots & 0 \\ 0 & \sigma^2_e & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_e \end{bmatrix} \tag{3}$$

If $\underline{u}_i$; the unobserved random effects vector belonging to ith subject, which takes place in Eq. 2 consists of only one random variable u_i representing the random effect of ith subject on response variable, then the special case of the LMM equation is called as Random Intercept Model (RIM). In this manner, including the ith subject random effect in the LMM by RIM brings about the advantage of modeling heterogeneity between subjects (Iyit and Genc, 2005a, b, 2007; Iyit, 2008).

By benefiting from the basic assumptions on RIM given by Eq. 2 as follows:

$$\begin{aligned} &u_i \sim N\left(0, \sigma^2_u\right),\ \underline{\varepsilon}_i \sim N\left(\underline{0}, R_i = \sigma^2_e I_{n_i}\right) \\ &C\,ov\left(u_i, \underline{\varepsilon}_{i'}'\right) = \underline{0}_{1\times n_i};\ i, i' = 1,2,\ldots,N \end{aligned} \tag{4}$$

Where:

Y_{ik} = The repeated measurement taken in kth time on ith subject

$$C\,ov\left(Y_{ik}, Y_{ik'}\right) = \sigma^2_u \quad k \ne k';\ var\left(Y_{ik}\right) = \sigma^2_u + \sigma^2_e \tag{5}$$

$\underline{Y}_i$; the $n_i \times 1$ dimensional vector of response variables representing all repeated measurements taken on ith subject given by Eq. 1 has Compound Symmetry (CS) variance-covariance structure as given in Eq. 6:

$$C\,ov\left(\underline{Y}_i\right) = V_i = Z_i D Z_i' + R_i;\ i = 1,2,\ldots,N = \begin{bmatrix} \sigma^2_u + \sigma^2_e & \sigma^2_u & \cdots & \sigma^2_u \\ \sigma^2_u & \sigma^2_u + \sigma^2_e & \cdots & \sigma^2_u \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_u & \sigma^2_u & \cdots & \sigma^2_u + \sigma^2_e \end{bmatrix} \tag{6}$$

As shown by Eq. 6, total variation in the structure of the response variable consists of σ^2_u; variation belonging to subject random effect and σ^2_e; variation belonging to random error term called as variance components (McCulloch and Searle, 2001; Fitzmaurice *et al*., 2004; Iyit, 2008).

Hence, $\underline{Y}_i$; the vector of response variables representing all repeated measurements taken on ith subject given by Eq. 1 has the following multivariate normal distribution:

$$\underline{Y}_i \sim N\left(X_i\underline{\beta}, V_i\right);\quad i = 1,2,\ldots,N \tag{7}$$

Estimation methods for the fixed effects parameter vector $\underline{\beta}$ given by Eq. 2 and the covariance parameters $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ in the CS variance-covariance structure of the response variable V_i given by Eq. 6 in RIM are Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) methods based on Newton-Raphson (NR) iterative method and Generalized Least Squares (GLS) method (McCulloch and Searle, 2001; Iyit, 2008).

In ML method, log-likelihood function of the fixed effects parameter vector $\underline{\beta}$ and the variance-covariance matrix of the response variable V_i having elements $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ known as variance components is as follows:

$$\begin{aligned} l = \log L\left(\underline{\beta}, V_i/\underline{Y}_i\right) = &-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|V_i| \\ &-\frac{1}{2}\left(\underline{Y}_i - X_i\underline{\beta}\right)' V_i^{-1}\left(\underline{Y}_i - X_i\underline{\beta}\right) \end{aligned} \tag{8}$$

By taking first derivatives of the log-likelihood function given by Eq. 8 with respect to $\underline{\beta}$ and $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ on the boundary of the parameter space as given in Eq. 9:

$$\begin{aligned} \Omega = \Big\{\left(\underline{\beta}, \underline{\sigma}^2\right): &-\infty < \beta_j < \infty, j = 0,1,2,\ldots,p; \\ &\sigma^2_e > 0 \text{ and } \sigma^2_u \geq 0\Big\} \end{aligned} \tag{9}$$

ML equations that maximize the log-likelihood function given by Eq. 8 on the parameter space given by McCulloch and Searle (2001) and Iyit (2008) Eq. 9 is as follows:

$$\left(X_i'\hat{V}_i^{-1}X_i\right)\hat{\underline{\beta}} = X_i'\hat{V}_i^{-1}\underline{Y}_i;\ i = 1,2,\ldots,N \tag{10}$$

$$\begin{aligned} \text{tr}\left(\hat{V}_i^{-1}Z_iZ_i'\right) = &\left(\underline{Y}_i - X_i\hat{\underline{\beta}}\right)'\hat{V}_i^{-1}Z_iZ_i'\hat{V}_i^{-1} \\ &\left(\underline{Y}_i - X_i\hat{\underline{\beta}}\right);\ i = 1,2,\ldots,N \end{aligned} \tag{11}$$

From the first ML equation given by Eq. 10, Maximum Likelihood Estimator (MLE) of the fixed effects parameter vector $\hat{\underline{\beta}}_{ML}$ can be easily derived by using MLE of the variance-covariance matrix of the response variable $\hat{V}$ derived from Eq. 11 as follows:

$$\text{MLE}\left(\underline{\beta}\right) = \hat{\underline{\beta}}_{ML} = \left(X_i'\hat{V}_i^{-1}X_i\right)^{-} X_i'\hat{V}_i^{-1}\underline{Y}_i;\quad i = 1,2,\ldots,N \tag{12}$$

Where:

$\left(X_i'\hat{V}_i^{-1}X_i\right)^{-}$ = The generalized inverse of the singular matrix $X_i'\hat{V}_i^{-1}X_i$

On the other hand, the second ML equation given by Eq. 11 is nonlinear with respect to the variance components $\hat{\underline{\sigma}}^2 = \left(\hat{\sigma}^2_u, \hat{\sigma}^2_e\right)'$ occurred in $\hat{V}_i^{-1}$. Thus, explicit analytical expressions for the solutions of Eq. 11 can be obtained by Newton-Raphson (NR) iterative method based on the derivatives of the log-likelihood function given by Eq. 8 (McCulloch and Searle, 2001; Iyit and Genc, 2007; Iyit, 2008).

In REML method, restricted log-likelihood function of the variance-covariance matrix of the response variable V_I having elements $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ k known as variance components not involving the fixed effects parameter vector $\underline{\beta}$ is as follows:

$$\begin{aligned} l_R = \log L_R\left(K_i'V_iK_i/K_i'\underline{Y}_i\right) = &-\frac{(N-r)}{2}\log(2\pi) \\ &-\frac{1}{2}\log\left|K_i'V_iK_i\right| \\ &-\frac{1}{2}\underline{Y}_i'K_i\left(K_i'V_iK_i\right)^{-1}K_i'\underline{Y}_i \end{aligned} \tag{13}$$

Where,

$$K_i' = \left[\underline{k}_1', \underline{k}_2', \ldots, \underline{k}_{N-r}'\right]$$

matrix satisfies the condition $K_i'X_i = 0$ and rank $(X_{ini\times(p+1)}) = r$.

By taking first derivatives of the restricted log-likelihood function given by Eq. 13 with respect to $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ on the boundary of the parameter space as given in Eq. 14:

$$\Omega = \left\{\underline{\sigma}^2 :\ \sigma^2_e > 0 \text{ and } \sigma^2_u \geq 0\right\} \tag{14}$$

REML equations that maximize the restricted log-likelihood function given by Eq. 13 on the parameter space given by Eq. 14 are as follows (McCulloch and Searle, 2001; Iyit, 2008):

$$\text{tr}(\hat{P}_iZ_iZ_i') = \underline{Y}_i'\hat{P}_iZ_iZ_i'\hat{P}_i\underline{Y}_i;\ i = 1,2,\ldots,N \tag{15}$$

Where,

$$\hat{P}_i = \hat{V}_i^{-1} - \hat{V}_i^{-1}X_i(X_i'\hat{V}_i^{-1}X_i)^{-}X_i'\hat{V}_i^{-1}$$

Also, REML equations given by Eq. 15 are nonlinear with respect to the variance components $\hat{\sigma}^2(\hat{\sigma}_u^2, \hat{\sigma}_u^2)'$ occurred in $\hat{V}_i^{-1}$. Thus, explicit analytical expressions for the solutions of Eq. 15 can be obtained by Newton-Raphson (NR) iterative method based on the derivatives of the restricted log-likelihood function given by Eq. 13 (McCulloch and Searle, 2001; Iyit and Genc, 2007; Iyit, 2008).

In GLS method, Generalized Least Squares Estimator (GLSE) of the fixed effects parameter vector $\hat{\underline{\beta}}_{GLS}$, which minimizes sum of squares error function Q ($\underline{\beta}$) by ignoring the random effect terms related to $Z_i\underline{u}$:

$$\begin{aligned} Q(\underline{\beta}) = \underline{\varepsilon}_i'\underline{\varepsilon}_i &= \left(\underline{Y}_i - X_i\underline{\beta} - Z_i\underline{u}_i\right)' V_i^{-1}\left(\underline{Y}_i - X_i\underline{\beta} - Z_i\underline{u}_i\right) \\ &= \underline{Y}_i' V_i^{-1}\underline{Y}_i - 2\underline{\beta}' X_i' V_i^{-1}\underline{Y}_i + \underline{\beta}' X_i' V_i^{-1} X_i\underline{\beta}; \\ & \qquad i = 1,2,\ldots,N \end{aligned} \tag{16}$$

can be derived as follows (McCulloch and Searle, 2001; Fitzmaurice *et al*., 2004; Iyit, 2008):

$$GLSE(\underline{\beta}) = \hat{\underline{\beta}}_{GLS} = \left(X_i' V_i^{-1} X_i\right)^{-} X_i' V_i^{-1}\underline{Y}_i; \quad i = 1,2,\ldots,N \tag{17}$$

Henderson *et al*. (1959) developed a set of equations known as Henderson Mixed Model (HMM) equations:

$$\begin{aligned} X_i' R_i^{-1} X_i\hat{\underline{\beta}} + X_i' R_i^{-1} Z_i\tilde{\underline{u}}_i &= X_i' R_i^{-1}\underline{Y}_i \\ Z_i' R_i^{-1} X_i\hat{\underline{\beta}} + \left(Z_i' R_i^{-1} Z_i + D^{-1}\right)\tilde{\underline{u}}_i &= Z_i' R_i^{-1}\underline{Y}_i; \\ & \qquad i = 1,2,\ldots,N \end{aligned} \tag{18}$$

that simultaneously yield Best Linear Unbiased Estimator (BLUE) of the fixed effects parameter vector BLUE (β) and Best Linear Unbiased Predictor (BLUP) of the unobserved random effects vector belonging to ith subject BLUP ($\underline{u}_i$) as Eq. 19:

$$\begin{aligned} \hat{\underline{\beta}} &= GLSE(\underline{\beta}) = BLUE(\underline{\beta}) = \left(X_i' V_i^{-1} X_i\right)^{-} X_i' V_i^{-1}\underline{Y}_i \\ \tilde{\underline{u}}_i &= BLUP(\underline{u}_i) = \left(Z_i' R_i^{-1} Z_i + D^{-1}\right)^{-1} Z_i' R_i^{-1}\left(\underline{Y}_i - X_i\hat{\underline{\beta}}\right); \\ & \qquad i = 1,2,\ldots,N \end{aligned} \tag{19}$$

The approximate variance-covariance matrix of $\left(\hat{\underline{\beta}} - \underline{\beta}, \tilde{\underline{u}}_i - \underline{u}_i\right)$ derived from HMM equations given by Eq. 18 as BLUE ($\underline{\beta}$) and BLUP ($\underline{u}_i$) in Eq. 19 is as follows (McLean *et al*., 1991; Littell *et al*., 2005; Iyit, 2008):

$$\hat{K} = Cov\left(\hat{\underline{\beta}} - \underline{\beta}, \tilde{\underline{u}}_i - \underline{u}_i\right) = \begin{bmatrix} X_i\hat{R}_i^{-1}?_i & X_i'\hat{R}_i^{-1}Z_i \\ Z_i'\hat{R}_i^{-1}X_i & Z_i'\hat{R}_i^{-1}Z_i + \hat{D}^{-1} \end{bmatrix}; \quad i = 1,2,\ldots,N \tag{20}$$

Statistical inferences concerning the fixed effects parameter vector and the random effects vector in the RIM given by Eq. 2 can be obtained by testing the null hypothesis of the estimable linear combinations of $\underline{\beta}$ and $\underline{u}_i$ of the following form;

$$H_0 : L\begin{bmatrix} \underline{\beta} \\ \underline{u}_i \end{bmatrix} = \underline{0}; \quad i = 1,2,\ldots,N \tag{21}$$

Where:

L = The $q \times \left(p + 1 + \sum_{i=1}^{N} q_i\right)$ dimensional contrast coefficients matrix fits the null hypothesis in case of rank (L) = q

When, L consists of a single row vector $\underline{l}'$, a general t-statistic can be constructed as follows:

$$t = \frac{\underline{l}'\begin{bmatrix} \hat{\underline{\beta}} \\ \tilde{\underline{u}}_i \end{bmatrix}}{\sqrt{\underline{l}'\hat{K}\underline{l}}}; i = 1,2,\ldots,N \tag{22}$$

approximately, t-distributed as $t_{1-\alpha/2,\hat{v}}$ where, $\hat{v}$ is the estimated degrees of freedom depending on Satterth Waite (SW) method (Satterthwaite, 1941):

$$\hat{v} \approx \frac{2\left(E\left(\underline{l}'\hat{K}\underline{l}\right)\right)^2}{\hat{var}\left(\underline{l}'\hat{K}\underline{l}\right)} \tag{23}$$

Here, from the associated confidence interval for the fixed effects parameter vector $\underline{\beta}$ and the random effects vector u_i in the RIM given by Eq. 2 can be considered as follows (McLean *et al*., 1991; Littell *et al*., 2005; Iyit, 2008):

$$\underline{l}'\begin{bmatrix} \hat{\underline{\beta}} \\ \tilde{\underline{u}}_i \end{bmatrix} \pm t_{1-\alpha/2,\hat{v}}\sqrt{\underline{l}'\hat{K}\underline{l}}; i = 1,2,\ldots,N \tag{24}$$

Statistical inferences concerning the covariance parameters $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$ in the CS variance-covariance structure of the response variable V_i given by Eq. 6 can be obtained by testing the null hypothesis of the following form where, $\sigma^2_0 = \sigma^2_e$ and $\sigma^2_1 = \sigma^2_u$:

$$\begin{aligned} H_0 &: \sigma_i^2 = 0 \\ H_1 &: \sigma_i^2 > 0; i = 0,1 \end{aligned} \tag{25}$$

by Wald Z test statistic, which is computed as the parameter estimate divided by its estimated asymptotic standard error as follows:

$$\text{Wald}(Z) = \frac{\hat{\sigma}_i^2}{SE(\hat{\sigma}_i^2)}; i = 1,0 \quad (26)$$

asymptotically χ^2-distributed as $\chi^2_{(v)}$ with $v = 2Z^2$ degrees of freedom.

Here from the associated confidence interval for the covariance parameters $\underline{\sigma}^2 = (\sigma^2_u, \sigma^2_e)'$; where, $\sigma^2_0 = \sigma^2_e$ and $\sigma^2_1 = \sigma^2_u$ can be considered as follows McLean *et al.* (1991), Littell *et al.* (2005) and Iyit (2008):

$$\frac{v\hat{\sigma}_i^2}{\chi^2_{v,1-\alpha/2}} \le \sigma_i^2 \le \frac{v\hat{\sigma}_i^2}{\chi^2_{v,\alpha/2}}; i = 1,0 \quad (27)$$

In model selection stage to determine, whether GLM or RIM best fits to the repeated measurements data for testing the following null hypothesis versus the alternative hypothesis.

H_0: RIM no better fits to the repeated measurements data than GLM.

H_1: RIM better fits to the repeated measurements data than GLM.

Likelihood Ratio (LR) test statistics obtained by taking twice the difference in the respective Maximized Log-likelihoods (ML):

$$G^2 = -2(\log L_{GLM} - \log L_{RIM}) \quad (28a)$$

or in the respective restricted maximized log-likelihoods (REML):

$$G^2 = -2(\log RL_{GLM} - \log RL_{RIM}) \quad (28b)$$

compared with a χ^2_d; chi-square distribution having degrees of freedom equal to the d; difference between the number of parameters in RIM and GLM or the information criterions given by Table 1 can be used as goodness-of-fit measures to test whether one model is significantly better than the other (Landau and Everitt, 2004; Fitzmaurice *et al.*, 2004).

In Table 1, logL represents the value of Maximized Log-likelihood (ML) or maximized restricted log-likelihood (REML) function, p is the total number of parameters in each model and N is the number of subjects included in the data set. The larger the information criterion, the better the model happens to be (Fitzmaurice *et al.*, 2004).

Table 1: Goodness-of-fit measures used to compare RIM and GLM fits to the repeated measurements data

Information criterions	References	Formulation
Akaike's Information Criterion (AIC)	Akaike (1974)	AIC = log L - p
Schwarz's Bayesian Information Criterion (BIC)	Schwarz (1978)	BIC = log L - p/2log (N)

APPLICATION

The data used in this study are taken from a retrospective clinical trial of 237 patients having high blood total cholesterol level (240 mg dL^{-1} and above), 103 in the ezetimibe drugs treatment group and 134 in the stain drugs treatment group to lower cholesterol. The main response variable used in this trial is cholesterol measurements recorded prior to the treatment (baseline), at 1, 2, 4 and 6 months follow-up the treatment. These kind of data are repeated measurements data with time as the single within-subject factor, which explains the within-subject variability and treatment group as the between-subject factor, which explains the between-subject variability coded as follows: ezetimibe treatment = 1, stain treatment = 2 (Iyit, 2008).

Let $\underline{Y}_i$; i = 1, 2, ..., 237 be the vector of response variables representing cholesterol repeated measurements taken at 1, 2, 4 and 6 months follow-up the treatment on ith subject:

$$\underline{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})'; \quad i = 1,2,\ldots,237 \quad (29)$$

the RIM equation as a special case of the LMM for cholesterol repeated measurements data in vector-matrix notation is as follows (Iyit, 2008):

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} = \begin{bmatrix} 1 & \text{treatment}_j & \text{baseline} & \text{time} & \text{treatment}_j \times \text{time} \\ 1 & \text{treatment}_j & \text{baseline} & \text{time} & \text{treatment}_j \times \text{time} \\ 1 & \text{treatment}_j & \text{baseline} & \text{time} & \text{treatment}_j \times \text{time} \\ 1 & \text{treatment}_j & \text{baseline} & \text{time} & \text{treatment}_j \times \text{time} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} u_i + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}; \begin{matrix} i = 1,2,\ldots,237 \\ j = 1,2 \end{matrix} \quad (30a)$$

$$(\underline{Y}_i)_{4\times1} = (X_i)_{4\times5}(\underline{\beta})_{5\times1} + (Z_i)_{4\times1}(\underline{u}_i)_{1\times1} + (\underline{\varepsilon}_i)_{4\times1}; \quad i = 1,2,\ldots,237 \quad (30b)$$

under the basic assumptions on RIM given:

$$u \sim N(0, \sigma_u^2), \underline{\varepsilon}_i \sim N(\underline{0}, R_i = \sigma_e^2 I_4)$$
$$\text{Cov}(u_i, \underline{\varepsilon}_{i'}') = \underline{0}_{1\times4}; \quad i,i' = 1,2,\ldots,237 \quad (31)$$

In the RIM, each patient's trend over time is assumed parallel to the treatment group's average trend, only the intercepts of each patient's regression line differ by including the random subject effect term u_i into the model is graphically illustrated in Fig. 1 (Landau and Everitt, 2004).

In the first constitution stage of the RIM, as a special case of the LMM for cholesterol repeated measurements data, Maximum Likelihood (ML) method based on Newton-Raphson (NR) iterative method and Generalized Least Squares (GLS) method are used to determine statistically significant fixed effects. From Table 2, it can be easily shown that treatment group, baseline cholesterol measurement, time and treatment × time interaction factor are found statistically significant fixed effects for RIM belonging to cholesterol repeated measurements data at $\alpha = 0.05$ significance level (Iyit, 2008).

In the second constitution stage of the RIM, as a special case of the LMM for cholesterol repeated measurements data, it must be decided to include the random subject effect term u_i into the model with the aim of modeling heterogeneity between subjects and also explaining total variation in the structure of the response variable given by Eq. 6 due to σ^2_u; variation belonging to ith subject random effect.

For this aim goodness-of-fit measures given by Table 3 will be helpful to compare GLM and RIM belonging to cholesterol repeated measurements data having the same statistically significant fixed effects given by Table 2.

From Table 3, the Likelihood Ratio (LR) test statistics values obtained by taking twice the difference in the respective Maximized Log-likelihoods (ML) given by Eq. 28a:

$$G^2 = 7913\text{-}7582.5 = 330.5 \tag{32a}$$

or in the respective Restricted Maximized Log-likelihoods (REML) given by Eq. 28b:

$$G^2 = 7935.9\text{-}7604.7 = 331.2 \tag{32b}$$

are compared with a $\chi^2_{1.0.95} = 3.84$ chi-square test statistic having one degrees of freedom equal to the difference between the number of parameters in the RIM and GLM. From Eq. 32a and b, both values for $G^2 > \chi^2_{1.0.95}$ are obtained then it can be decided that RIM clearly provides a better fit to the cholesterol repeated measurements data than GLM.

On the other hand, the higher values of AIC and BIC information criterions for RIM then GLM given by Table 3 indicate that RIM better fits to the cholesterol repeated measurements data than GLM.

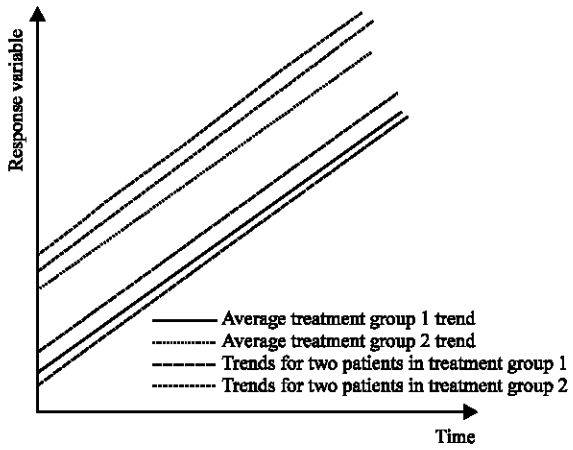

Fig. 1: Graphical illustration of RIM by time versus response variable (Landau and Everitt, 2004)

In this manner, it is approved that the random subject effect term u_i can be included into the model belonging to cholesterol repeated measurements data for modeling heterogeneity between subjects and explaining total variation in the structure of the response variable.

After determining statistically significant fixed effects and subject random effect in the RIM as a special case of the LMM for cholesterol repeated measurements data, predictions of randomly selected 32nd and 237th subject random effects from 237 subjects included into the study are given by Table 4 by using Henderson Mixed Model (HMM) equations.

For the observed vector of response variables representing cholesterol repeated measurements taken at 1, 2, 4 and 6 months follow-up the treatment on randomly selected 32nd subject included into the study:

$$\underline{y}_{32} = (270, 243, 234, 200)' \tag{33}$$

by using ML and GLS parameter estimates of fixed effects given by Table 2 and HMM equations predictions of subject random effects given by Table 4 for RIM belonging to cholesterol repeated measurements data:

$$\begin{bmatrix} \hat{y}_{32,1_{ML}} \\ \hat{y}_{32,2_{ML}} \\ \hat{y}_{32,3_{ML}} \\ \hat{y}_{32,4_{ML}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 14.873 & -2.25 & 1 \\ 1 & 1 & 14.873 & -1.25 & 1 \\ 1 & 1 & 14.873 & 0.75 & 1 \\ 1 & 1 & 14.873 & 2.75 & 1 \end{bmatrix} \begin{bmatrix} 239.02 \\ 8.8686 \\ 0.7644 \\ -7.6995 \\ 1.2763 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} (-19.3044) \tag{34a}$$

Table 2: Parameter estimates of fixed effects for RIM belonging to cholesterol repeated measurements data

Parameters	Parameter estimation method for fixed effects	Estimate	SE	df	t-test	Sig.	95% confidence interval	
							Lower bound	Upper bound
Intercept	ML	239.0200	1.10230	237	216.84	<0.0001	236.8500	241.1900
	GLS	239.0200	1.11170	233	215.01	<0.0001	236.8300	241.2100
Treatment	ML	8.8686	1.67200	237	5.30	<0.0001	5.5748	12.1625
	GLS	8.8686	1.68630	233	5.26	<0.0001	5.5463	12.1909
Baseline	ML	0.7644	0.04366	237	17.51	<0.0001	0.6784	0.8504
	GLS	0.7644	0.04403	233	17.36	<0.0001	0.6776	0.8511
Time	ML	-7.6995	0.23850	711	-32.29	<0.0001	-8.1677	-7.2313
	GLS	-7.6995	0.23910	707	-32.20	<0.0001	-8.1690	-7.2300
Treatment×time	ML	1.2763	0.36170	711	3.53	0.0004	0.5661	1.9864
	GLS	1.2763	0.36270	707	3.52	0.0005	0.5641	1.9884

Table 3: Goodness-of-fit measures used to compare GLM and RIM belonging to cholesterol repeated measurements data

Model	Parameter estimation method for fixed effects	Estimation method for covariance parameters	Covariance structure	Goodness-of-fit measures		
				-2 (Restricted) log-likelihood	AIC	BIC
GLM	ML	ML		7913.0	-3957.5	-3959.9
	GLS	REML	VC	7935.9	-3969.0	-3971.4
RIM	ML	ML		7582.5	-3793.3	-3796.7
	GLS	REML	CS	7604.7	-3804.4	-3807.8

Table 4: Predictions of subject random effects for RIM belonging to cholesterol repeated measurements data

Subject	Subject random effect	Parameter estimation method for fixed effects	Henderson Mixed Model (HMM) equations predictions for subject random effects	SE	df	t	Sig.
32	u_{32}	ML	-19.3044	4.9513	946	-3.90	0.0001
		GLS	-19.3500	4.9729	938	-3.89	0.0001
237	u_{237}	ML	16.4200	4.9363	947	3.33	0.0009
		GLS	16.4587	4.9577	939	3.32	0.0009

$$\begin{bmatrix} \hat{y}_{32,1_{GLS}} \\ \hat{y}_{32,2_{GLS}} \\ \hat{y}_{32,3_{GLS}} \\ \hat{y}_{32,4_{GLS}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 14.873 & -2.25 & 1 \\ 1 & 1 & 14.873 & -1.25 & 1 \\ 1 & 1 & 14.873 & 0.75 & 1 \\ 1 & 1 & 14.873 & 2.75 & 1 \end{bmatrix} \begin{bmatrix} 239.02 \\ 8.8686 \\ 0.7644 \\ -7.6995 \\ 1.2763 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} (-19.3500) \quad (34b)$$

the ML and GLS estimated vector of response variables for randomly selected 32nd subject is obtained as follows:

$$\underline{\hat{y}}_{32_{ML}} = \underline{\hat{y}}_{32_{GLS}} = (259, 251, 235, 220)' \quad (35)$$

Covariance parameter estimates of the random effects $\underline{\hat{\sigma}}^2 = (\hat{\sigma}_u^2, \hat{\sigma}_t^2)'$ in the Compound Symmetry (CS) variance-covariance structure of the response variable V_i shown by Eq. 6 for RIM belonging to cholesterol repeated measurements data are given by Table 5.

By using Table 5, ML estimate of the Compound Symmetry (CS) variance-covariance matrix V_i given by Eq. 6 for the vector of response variables representing cholesterol repeated measurements taken at 1, 2, 4 and 6 months follow-up the treatment on ith subject is as follows:

$$\hat{V}_{i_{ML}} = \begin{bmatrix} 246.94 & 134.62 & 134.62 & 134.62 \\ 134.62 & 246.94 & 134.62 & 134.62 \\ 134.62 & 134.62 & 246.94 & 134.62 \\ 134.62 & 134.62 & 134.62 & 246.94 \end{bmatrix} \quad i = 1, 2, \ldots, 237 \quad (36a)$$

REML estimate of the CS variance-covariance matrix V_i is also as follows:

$$\hat{V}_{i_{REML}} = \begin{bmatrix} 250.22 & 137.26 & 137.26 & 137.26 \\ 137.26 & 250.22 & 137.26 & 137.26 \\ 137.26 & 137.26 & 250.22 & 137.26 \\ 137.26 & 137.26 & 137.26 & 250.22 \end{bmatrix} \quad i = 1, 2, \ldots, 237 \quad (36b)$$

The ML estimate of intra-class correlation coefficient (ρ), which explains the total variation in the structure of the response variable given by Eq. 6 due to σ^2_u; variation belonging to subject random effect

$$\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2) \quad (37)$$

is $\hat{\rho}_{ML}$ = 134.62/(134.62 + 112.32) = 0.5452. The REML estimate of intra-class correlation coefficient is

Table 5: Covariance parameter estimates of random effects for RIM belonging to cholesterol repeated measurements data

Covariance parameters	Estimation method for covariance parameters	Estimate	SE	Wald Z	Sig.	95% confidence interval	
						Lower bound	Upper bound
σ^2_ε	ML	112.32	5.9571	18.85	<0.0001	101.50	124.98
	REML	112.96	6.0078	18.80	<0.0001	102.05	125.72
σ^2_u,	ML	134.62	15.0202	8.96	<0.0001	109.43	169.69
	REML	137.26	15.4061	8.91	<0.0001	111.44	173.26

$$\hat{\rho}_{REML} = 137.26/(137.26+112.96) = 0.5486$$

From here, it can be concluded that subject random effect explains the total variation in the structure of the response variable 54.5 and 54.9% for ML and REML, respectively.

By using the ML and REML estimates of the intra-class correlation coefficient given above, ML estimate of the correlation matrix R_i for the vector of response variables representing cholesterol repeated measurements taken at 1, 2, 4 and 6 months follow-up the treatment on ith subject is as follows:

$$\hat{R}_{i_{ML}} = \begin{bmatrix} 1 & 0.5452 & 0.5452 & 0.5452 \\ 0.5452 & 1 & 0.5452 & 0.5452 \\ 0.5452 & 0.5452 & 1 & 0.5452 \\ 0.5452 & 0.5452 & 0.5452 & 1 \end{bmatrix} \quad (38a)$$

$$i = 1, 2, \ldots, 237$$

REML estimate of the c orrelation matrix R_i is also as follows:

$$\hat{R}_{i_{REML}} = \begin{bmatrix} 1 & 0.5486 & 0.5486 & 0.5486 \\ 0.5486 & 1 & 0.5486 & 0.5486 \\ 0.5486 & 0.5486 & 1 & 0.5486 \\ 0.5486 & 0.5486 & 0.5486 & 1 \end{bmatrix} \quad (38b)$$

$$i = 1, 2, \ldots, 237$$

RESULTS AND DISCUSSION

In this study, RIM as a special case of LMMs implying CS variance-covariance structure assumption that each pair of repeated measurements has the same correlation, variance and covariance terms is constituted to cholesterol repeated measurements data set recorded prior to the treatment (baseline), at 1, 2, 4 and 6 months follow-up the treatment with time as the single within subject factor and treatment group as the between-subject factor. The superiority of RIM bringing about the advantage of modeling heterogeneity between subjects in the intercepts of each subject's own regression line by including random subject effect term u_i into the model than GLM for repeated measurements data implying VC variance-covariance structure assumption that each pair of repeated measurements are uncorrelated and have constant variance is emphasized.

It is observed that generally covariances between observations made closer together in time are likely to be higher than those made at greater time points. So, GLM including only modeling the expected value of the observations as a linear function of explanatory variables based on the assumption that the observations from different subjects are statistically independent, uncorrelated and that the variance-covariance structure is the same for each subject is not found appropriate for the statistical analysis of such a repeated measurements design. On the other hand, RIM implying CS variance-covariance structure given by Eq. 6 for the covariance matrix of the repeated measurements that variances at different time points are equal and covariances between each pair of time points are equal is in fact a very restrictive and unrealistic assumption.

CONCLUSION

From this study, it can be clearly seen that all repeated measurements of the response variable are independent and uncorrelated with the same variance-covariance structure of each subject for RIM belonging to cholesterol repeated measurements data. On the other hand, making such a restrictive assumption is not meaningful because it is expected that the closer repeated measurement pairs have higher correlations then the others.

In the light of this study, it would be interesting to model the covariance structure of the vector of response variables representing all repeated measurements taken on ith subject in various flexible forms such as homogeneous and heterogeneous variance-covariance structures other than CS pattern by Random Intercept and Slope Model (RISM) in a further study.

Also, in the further study, a comparative evaluation of RIM with RISM bringing about the advantage of modeling heterogeneity between subjects in both intercepts and slopes of each subject's own regression line by including random subject effect terms into the model as a special case of LMMs will be taken into consideration following a repeated measurements design.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control, 19 (6): 716-723. DOI: 10.1109/TAC. 1974.1100705. http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1100705.

Bryk, A.S. and M.E. Driscoll, 1988. The high school as community: Contextual influences and consequences for students and teachers. Technical Research Report. Madison, WI: University of Wisconsin-Madison, National Center on Effective Secondary Schools. http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED302539.

Davis, C.S., 2002. Statistical Methods for the Analysis of Repeated Measurements. 1st Edn. Springer-Verlag Inc., New York, pp: 125-167. ISBN: 0-387-95370-1.

Fitzmaurice, G.M., N.M. Laird and J.H. Ware, 2004. Applied Longitudinal Analysis. 1st Edn. John Wiley and Sons Inc., New York, pp: 187-234. ISBN: 0-471-21487-6.

Hamer, R.M. and M.P. Simpson, 1989. An introduction to the analysis of repeated measures for continuous response data using PROC GLM and PROC MIXED. Sugi Beginning Tutorials. http://www2.sas.com/proceedings/sugi23/Begtutor/p58.pdf.

Henderson, C.R., O. Kempthorne, S.R. Searle and C.N. Von Krosigk, 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics, 15: 192-218. DOI: 10.2307/2527669. http://www.jstor.org/pss/2527669.

Iyit, N. and A. Genc, 2005a. An application related to linear mixed models from hierarchical modelling techniques in correlated response data analysis. Proceedings of the 4th Statistics Congress, 8-12 May Antalya, Turkey pp: 270-271.

Iyit, N. and A. Genc, 2005b. An application related to linear mixed effects models for repeated measurements. Proceedings of the International Conference Ordered Statistical Data: Approximations, Bounds and Characterizations, 15-18 June. Izmir, Turkey, pp: 60.

Iyit, N., A. Genc and F. Arslan, 2006. Analysis of repeated measures for continuous response data using General Linear Model (GLM) and mixed models. Proceedings of the International Conference on Modeling and Simulation 28-30 August, Konya, Turkey: pp: 937-942.

Iyit, N. and A. Genc, 2007. A comparative study of fixed effects models and random intercept/slope models as a special case of linear mixed models for repeated measurements. Selcuk J. Appl. Math., 8 (1): 57-75. http://www.sumam.selcuk.edu.tr/sjam078107.html.

Iyit, N., 2008. Constitution of linear mixed models in the analysis of correlated data. Philosophy of Doctora (Ph.D.) Thesis, Selcuk University, Graduate School of Natural and Applied Sciences, Department of Mathematics, Konya, Turkey. Turkish Government Higher Education Board National Thesis Center Thesis Number: 178233. http://tez2.yok.gov.tr/.

Laird, N.M. and J.H. Ware, 1982. Random-effects models for longitudinal data. Biometrics, 38: 963-974. DOI: 10.2307/2529876. http://www.jstor.org/pss/2529876.

Landau, S. and B.S. Everitt, 2004. A Handbook of Statistical Analyses using SPSS. 1st Edn. Chapman and Hall/CRC Press, Boca Raton, pp: 196-226. ISBN: 1-58488-369-3.

Littell, R.C., G.A. Milliken, W.W. Stroup and R.D. Wolfinger, 2005. SAS System for Mixed Models. 1st Edn. SAS Institute Inc., Cary, NC, USA, pp: 733-754. ISBN: 1-59047-500-3.

McCulloch, C.E. and S.R. Searle, 2001. Generalized, Linear and Mixed Models. 1st Edn. John Wiley and Sons Inc., New York, pp: 156-184. ISBN: 0-471-19364-X.

McLean, R.A., W.L. Sanders and W.W. Stroup, 1991. A unified approach to mixed linear models. Am. Statist., 45: 54-64. DOI: 10.2307/2685241. http://www.jstor.org/pss/2685241.

Raudenbush, S.W. and A.S. Bryk, 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Edn. Advanced Quantitative Techniques in the Social Sciences Series, Sage Publications, London, pp: 99-116. ISBN: 0-7619-1904-X.

Satterthwaite, F.E., 1941. Synthesis of variance. Psychometrika, 6: 309-316. DOI: 10.1007/BF02288586. http://www.springerlink.com/content/6h77h7288554h864/.

Schwarz, G., 1978. Estimating the dimension of a model. The Ann. Stat., 6 (2): 461-464. DOI: 10.2307/2958 889. http://www.jstor.org/pss/2958889.

Ware, J.H., 1985. Linear models for the analysis of longitudinal studies. Am. Statist., 39: 95-101. DOI: 10.2307/2682803. http://www.jstor.org/pss/268 2803.