A Score Test for Detecting Heterogeneous Within-Group Variances in Linear Mixed Models

Chul H. Ahn[1]
[1]Division of Biostatistics, Center for Devices and Radiological Health, US Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993

Abstract
One of the common assumptions in linear mixed models is that the within-group variances are homogeneous. This paper investigates the impact of heterogeneous within-group variances on the estimate of a regression coefficient especially when the within-group variance depends on a group-level covariate. We also develop a diagnostic test based on score statistic to detect heterogeneous within-group variances in linear mixed linear models. The method is applied to the data from an orthopedic device clinical trial.

Key Words: Diagnostic test, model violation, heteroscadasticity, score statistic

1. Introduction

This paper is concerned with the problem of testing constant variance assumption in linear mixed model. In linear mixed model, there may be two sources of heterogeneity – between-group variances and within-group variances. Regarding heterogeneity between group variances Heagerty and Kurland (2001) have shown in generalized linear mixed models, that there can be bias in the mean regression coefficient if the random effects variance depends on a between-subject covariate, and this dependence is not modeled. Regarding heterogeneity within-group variances Lin, Raz and Harlow (1997) extended linear mixed models to allow for heterogeneous within-cluster variances. They assumed that the within-group variances are related to a vector of subject-level covariates. These two papers show that when heteroscedasticity occurs in linear mixed models, the variances of errors and random effects may depend on subject-level covariates, but more generally, it could also depend on other relevant quantities such as time or spatial ordering.

In this paper we concentrate on the heterogeneous within-group variances which may be more applicable in clinical trials where the within-group variances are often quite different across investigational centers. Especially, in medical device trials, site-to-site variation can be larger and may lead to more serious problem compared to drug trials. The large variation between sites may be due to physician experience and training in using or implanting the device (sites with device inventor usually perform better). It may also be due to patient population, patient management, and reporting practices. Site effects are attributable to random variation. However, the site effects may be larger than those which can be attributed to random variation. One of the motivations with this paper

is that perhaps this extra variation between sites can be modeled through the variables defined for each site.

In Section 2, the linear mixed model with heterogeneous within-group variances will be formulated with several expanded models for within-group variances. In Section 3, the impact of heterogeneous within-group variances on the estimates of regression parameters will be presented using simulated data. It will be shown that the regression coefficient may not be correctly estimated if the within-group variance depends on a group-level covariate, and this dependence is not modeled properly. In Section 4, we propose a diagnostic test for detecting heterogeneous within-group variances based on score statistic, and it will be applied to a real example from a medical device trial.

2. The Linear Mixed Model with Heterogeneous Within-Group Variances

The linear mixed model can be represented as

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{U}\,\boldsymbol{b} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{Y}$ is an Nx1 response vector with elements Y_{ij} which denotes the *j*-th observation (the subscript *j* going from 1 to n_i) made within the *i*-th group (the subscript *i* going from 1 to *t*), $\boldsymbol{X}$ is an N x p known matrix, $\boldsymbol{\beta}$ is a p x 1 vector of unknown regression parameters assumed to be fixed, and $\boldsymbol{U}$ is an N x t design matrix for the random effects, $\boldsymbol{b}$ with independent elements b_i, which are assumed to be distributed, b_i ~ N(0, σ_b^2). The errors, ϵ is an N x 1 random vector with independent elements, ε_{ij}, which are assumed to be distributed , ε_{ij} ~ N(0, σ^2). The random effects and error are assumed to be independent. Note that we included only one random effect in model (1), but more than one random effect can be included. The $\boldsymbol{U} = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_t})$ where $\mathbf{u_i}$ is an N x 1 column vector with values one at the position corresponding to the group *i* and zero elsewhere. The model (1) can be generalized to random coefficient model if the column of the $\boldsymbol{U}$ is allowed to have independent variables.

One of the assumptions we are interested in the linear mixed model is that within-group variance or error variance is constant and that each realization of random variable has been drawn from the same normal population. As in any other modelling situation, we seek diagnostic methods concerning the plausibility of these assumptions. A standard method of deriving diagnostics is through the technique of model expansion, in which a model like (1) is embedded in a larger class, and a relatively simple procedure like a score test is then used to turn a test into a diagnostic. For model (1), useful expansion can be obtained by setting

$$var(\varepsilon_{ij}) = \sigma^2 \boldsymbol{h}(\boldsymbol{z_i}, \boldsymbol{\lambda}) \tag{2}$$

where $\boldsymbol{z_i}$ is a known *q* x 1 vector for the *i*-th group with elements, $\boldsymbol{z_i} = \{z_{ik}\}$, the subscript *k* going from 1 to *q*, and $\boldsymbol{\lambda}$ is a *q* x 1 vector of unknown parameters. We assume that function $\boldsymbol{h}(\cdot)$ is twice differential with respect to $\boldsymbol{\lambda}$ and there is a unique value $\boldsymbol{\lambda}_0$ of $\boldsymbol{\lambda}$ such that $\boldsymbol{h}(\boldsymbol{z_i}, \boldsymbol{\lambda_0})$=1 for all $\boldsymbol{z_i}$. Naturally, the null hypothesis for testing homogeneity will be $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$. In (2), the expanded model suggests that the within-group variance may be non-constant, but explainable through $\boldsymbol{z_i}$. This kind of model expansion technique was

used in Cook and Weisberg (1983) in examining the assumption of constant variance in the usual linear regression.

One of the useful functions for $\boldsymbol{h}$ would be exponential since it is not only continuously differentiable and recovers constant variance when $\boldsymbol{\lambda}_0 = 0$, but also due to its monotone feature we can find a direction such that the variance increases in that direction. Our first choice of $\boldsymbol{h}$ would be

$$\boldsymbol{h}(\boldsymbol{z_i}, \boldsymbol{\lambda}) = \exp(\boldsymbol{\lambda}^{\boldsymbol{T}} \boldsymbol{z_i}). \tag{3}$$

Here $\boldsymbol{z_i}$ can be considered as a vector of covariates for group *i*, and often it is appropriate to choose the variables from the column of $\boldsymbol{X}$. The model (3) says that the within-cluster variance depends on group covariates, $\boldsymbol{z_i}$ through a function of $\boldsymbol{h}$ with $\boldsymbol{\lambda}$, unknown parameter. We often observe that the sample variance of clinical outcome varies across sites. This model may also be used in repeated measurement or longitudinal data where it is common for the variance of the observations to increase as the magnitude of numbers increases. Sometimes heteroscedasticity can be detected when $\boldsymbol{z_i}$ are taken usual transformation such as square or log before it is multiplied to $\boldsymbol{\lambda}$**.** When the variance may depend on the expected response, we take function $\boldsymbol{h}$ to be of the form

$$\boldsymbol{h}(\boldsymbol{z_i}, \boldsymbol{\lambda}) = \boldsymbol{h}(E(\boldsymbol{y_i})). \tag{4}$$

3. Impact of Heteroscedasticity on Parameter Estimates

As Heagerty and Kurland (2001) and Lin, Raz and Harlow (1997) pointed out, there may be bias in the regression parameter estimates of linear mixed model if there exists heteroscedasticity in the data, and it is not correctly modeled. The heteroscedasticity in linear mixed model may arise in several forms, but in this paper, we will consider a case where within-group variance depends on group covariate such as one in (3). In (3) we are interested in finding out the impact of $\boldsymbol{\lambda} \neq 0$ on point estimates and confidence intervals of parameters. A simulation study was conducted in the balanced case to investigate such impact. The values of $\boldsymbol{\lambda}$ will be varied with the other parameter values fixed. The responses Y_{ij} (the clinical outcome from unit *j* in group *i*) were generated from the following model.

$$Y_{ij} = \beta_0 + \alpha \cdot T_{ij} + \beta_1 \cdot X_{ij} + b_i + \boldsymbol{\varepsilon_{ij}} \tag{5}$$

where

β_0 **:** intercept (overall mean)

α **:** treatment effect

T_{ij} **:** treatment assigned for Y_{ij} **(**1 for treatment 0, for control)

β_1 **:** regression coefficient for covariate X_{ij}

X_{ij} **:** covariate for unit j in group i (i going from 1 to t, and j from 1 to n_i)

b_i **:** group i (random effect) ~ N(0, σ_b^2**)**

ε_{ij} **:** error ~ N(0, $\sigma^2 e^{\lambda^T z_i}$**)**

The values of $\boldsymbol{\lambda}$ was varied from -2 to 2 (-2, -1, 0, 1, 2) with the other parameter values fixed: $\beta_0 = 0,\ \alpha = 1, \beta_1 = 1, \sigma_b^2 = 1,$ and $\sigma^2 = 1.$ The number of groups (t) was fixed as 20, and the number of units per group (n_i) as 10. The T_{ij}'s were randomly assigned with either 1 for treatment or 0 for control. The X_{ij}'s were generated from standard normal distribution. The two hundred data sets were generated for each set of parameter values, and the maximum likelihood estimates were calculated. The numerical algorithms such as Newton-Raphson and Levenberg-Marquardt methods were used to obtain maximum likelihood estimates of parameters. The SAS PROC IML (Version 9.3) was used as programming language. Here we report the estimates of treatment effect ($\hat{\alpha}$) and its standard error. Figure 1 shows the six box plots, each describing the distribution of 200 treatment effects ($\hat{\alpha}$'s). And, above each box plot, the minimum, maximum, and average of 200 $\hat{\alpha}$'s are displayed. The first box plot was drawn from 200 treatment effect estimates, each of which was obtained from the correctly specified within-group variance ($var(\varepsilon_{ij}) = \sigma^2 \exp(\boldsymbol{\lambda}^T \boldsymbol{z_i})$) with $\boldsymbol{\lambda}$ = -2. The label "Correct Lambda (-2)" implies that the model for the within-group variance is correct and the lambda ($\boldsymbol{\lambda}$) is -2. And, the second box plot with the label with "Incorr.Lambda (-2)" implies that the model for the within-group variance is incorrect (that is, $(var(\varepsilon_{ij}) = \sigma^2)$ in the sense that it ignores the heterogeneous within-group variances and the lambda ($\boldsymbol{\lambda}$) is -2. The rest of the box plots may be interpreted in a similar way. We may notice from this Figure that the two box plots are almost identical when $\boldsymbol{\lambda}$ = 0. When $\boldsymbol{\lambda} \neq 0$, however, the box plots are more spread when the within-group variance is incorrectly specified.

From this six box plots in Figure 1, we may conclude that the standard errors of treatment effect estimates may be larger when the within-group variance is not correctly specified, which is shown in Figure 2.

Figure 1. Distribution of treatment effect estimates ($\hat{\alpha}$) from 200 simulated datasets

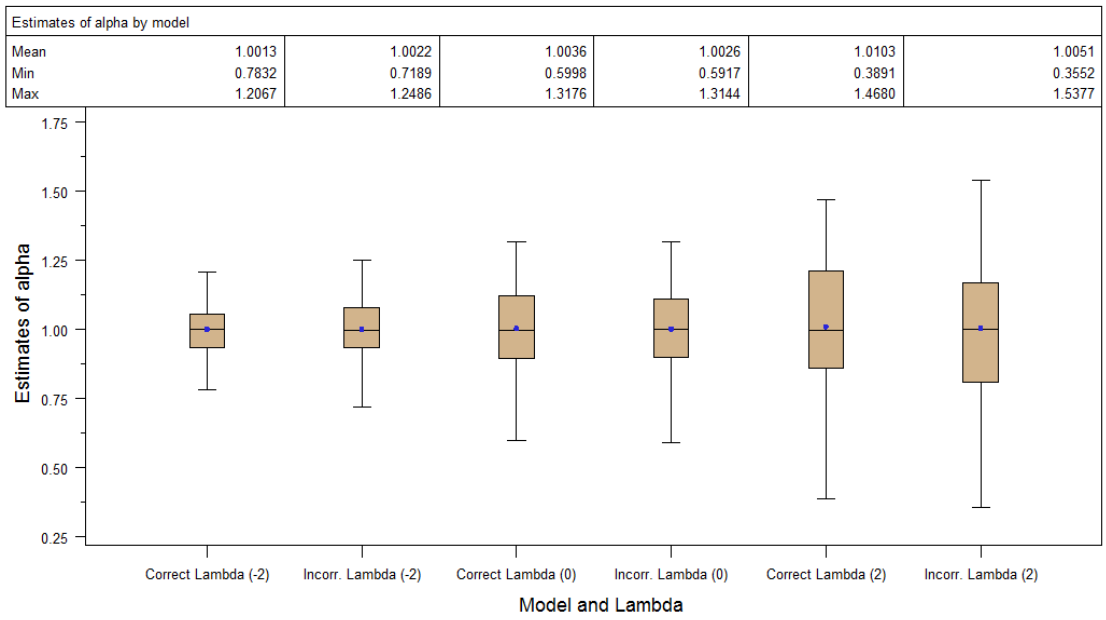
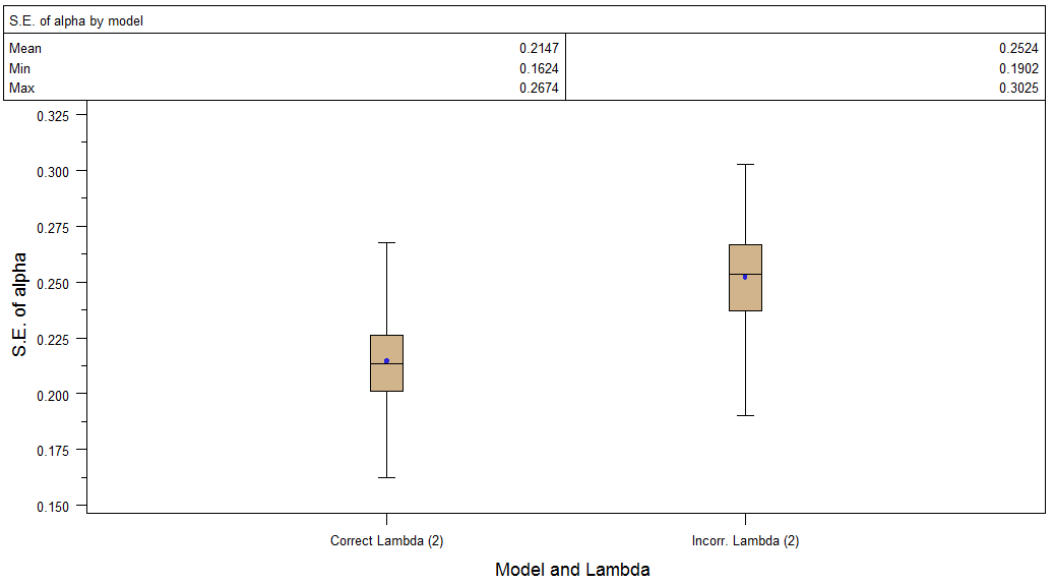


Figure 2 shows the two box plots describing the distribution of the standard errors of treatment effect estimates ($s.e.(\hat{\alpha})$) – the first box plot is when the within-group variance is correctly specified, and the second one, incorrectly specified. Both of them are when the Lambda ($\boldsymbol{\lambda}$) = 2. Above each box plot, the minimum, maximum, and average of 200 $s.e.(\hat{\alpha})$'s are shown. We can see by comparing the two box plots, the values of $s.e.(\hat{\alpha})$ are generally smaller when the variance structure is correctly specified. In fact, the average magnitude of $s.e.(\hat{\alpha})$ is 0.2147 when the within-group variance is correctly modeled, and 0.2524 when the within-group variance is incorrectly modeled. This finding confirms the result from the Figure 1.

Figure 2: Distribution of $s.e.(\hat{\alpha})$ from 200 simulated datasets



4. A Diagnostic Test for Detecting Heterogeneous Within-Group Variances

The linear mixed model in (1) may be rewritten for the response vector for group *i* as follows:

$$\boldsymbol{Y_i} = \boldsymbol{X_i\beta} + \boldsymbol{U_i}b_i + \boldsymbol{\varepsilon_i} \tag{6}$$

where $\boldsymbol{Y_i}$ denotes an n_i x 1 response vector with elements Y_{ij} (the clinical outcome from unit *j* in group *i*) with the subscript *i* going from 1 to *t*. And, the heterogeneous within-group variance will be modeled as in (2). That is,

$$var(\varepsilon_{ij}) = \sigma^2\boldsymbol{h}(\boldsymbol{z_i}, \boldsymbol{\lambda}) \tag{7}$$

Our interest is to develop a diagnostic test for detecting heterogeneous within-group variance by testing the null hypothesis, H_0: $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ versus the alternative hypothesis, H_1: $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0$. A likelihood ratio test (LRT) can be useful for testing this hypothesis, but it requires the maximum likelihood estimate of $\boldsymbol{\lambda}$**.** Moreover, the closed form solution for the m.l.e. of λ is not available. On the other hand, score statistic (a second order large-sample equivalent to LRT) does not require the m.l.e. of $\boldsymbol{\lambda}$ under the null.

Before we derive score test, we let $\boldsymbol{f(y_i, \eta)}$ to denote a probability density function of $\boldsymbol{Y_i}$ in (6) with $\boldsymbol{\eta}$, the unknown parameter vector, and assume that it satisfies the regularity conditions stated in Serfling (1980, P. 144). The unknown parameter vector, $\boldsymbol{\eta}$ will be decomposed into $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma_b^2}, \boldsymbol{\sigma^2})$ **.** For a given set of *t* independent response vectors on $\boldsymbol{Y_i}$, say, $\boldsymbol{y_1, y_2, \dots, y_t}$, denote $\boldsymbol{\ell(\theta, \lambda) = \ell_1(\theta, \lambda) + \ell_2(\theta, \lambda) + \cdots + \ell_t(\theta, \lambda)}$ a log-likelihood, where $\boldsymbol{\ell_i(\theta, \lambda) = \log f(y_i, \eta)}$. We are interested in testing

$$H_0: \boldsymbol{\lambda} = \boldsymbol{\lambda}_0 \text{ versus } H_1: \boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0. \tag{8}$$

The score vector, $\boldsymbol{d_\lambda}$ can be obtained by taking the first derivative of $\boldsymbol{\ell(\theta, \lambda)}$ with respect to $\boldsymbol{\lambda}$ and the information matrix, $\boldsymbol{J_{\lambda\lambda}}$ by taking the second derivative of $\boldsymbol{\ell(\theta, \lambda)}$ with respect to $\boldsymbol{\lambda}$. The $\boldsymbol{d_\lambda}$ and $\boldsymbol{J_{\lambda\lambda}}$ are evaluated at the restricted maximum likelihood estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$**.** The score statistic can then be defined as in Cox and Hinkley (1974) by

$$S = \hat{d}_\lambda^T (\hat{J}_{\lambda\lambda} - \hat{J}_{\lambda\theta} \hat{J}_{\theta\theta}^{-1} \hat{J}_{\theta\lambda})^{-1} \hat{d}_\lambda . \tag{9}$$

Under the null distribution the score statistic in (9) is asymptotically distributed as Chi-square with the degrees of freedom equal to the dimension of $\boldsymbol{\lambda}$.

Denoting the variance of ε_{ij} as $var(\varepsilon_{ij}) = \sigma^2 \boldsymbol{h(z_i, \lambda)}$, and the covariance matrix of $\boldsymbol{Y_i}$ as $\boldsymbol{Cov(Y_i) = \sigma^2 Q_i}$, the matrix $\boldsymbol{Q_i}$ can be written as $\boldsymbol{Q_i = \xi U_i U_i^T + h(\lambda^T z_i) I_{n_i}}$ with $\xi = \sigma_b^2/\sigma^2$ and $\boldsymbol{I_{n_i}}$ indicating the identity matrix with the dimension of n_i, and the log-likelihood of data can be written as

$$\ell_i(\boldsymbol{\theta^*, \lambda}) = -\frac{n_i}{2}\log 2\pi - \frac{1}{2}\log|\sigma^2 \mathbf{Q_i}| - \frac{1}{2}(\mathbf{y_i} - \mathbf{X_i^T}\boldsymbol{\beta})^T (\sigma^2 \mathbf{Q_i})^{-1} (\mathbf{y_i} - \mathbf{X_i^T}\boldsymbol{\beta})$$

where $\boldsymbol{\theta^* = (\beta^T, \sigma^2, \xi)^T}$**.**

In the balanced case, the score test statistic is simplified as follows:

$$S = \frac{1}{2}\mathbf{R^T} \cdot \overline{\mathbf{C}} (\overline{\mathbf{C}}^{\mathbf{T}} \overline{\mathbf{C}})^{-1} \overline{\mathbf{C}}^{\mathbf{T}} \cdot \mathbf{R}$$

where R is a t x 1 vector with elements $R_i = \phi \cdot \bar{e}_{i\cdot} / \hat{\sigma}^2$, $\boldsymbol{\phi = n/(1 + n\xi)}$,

$\phi = \dfrac{n}{1 + n\hat{\xi}}$ and $\bar{e}_{i\cdot} = \sum_{j=1}^{n} (Y_{ij} - x_{ij}^t \hat{\boldsymbol{\beta}}) / n$ with $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$, $\hat{\xi}$ maximum likelihood estimates.

$\mathbf{C}_{t \times q} = \{ [h'(\lambda_0^T z_i)]^{\mathrm{T}} \}$ with $h'(\lambda_0^T z_i)_{q \times 1} = \{ \partial h(\lambda^T z_i) / \partial \lambda_k \}_{\lambda = \lambda_0}$.

$\overline{\mathbf{C}}_{t \times q}$ is a matrix obtained from $\mathbf{C}$ by subtracting column averages.

Computationally, $S = \dfrac{1}{2}$ Regression Sum of Squares from the regression of $\mathbf{R}$ on $\mathbf{C}$ in the constructed model $\mathbf{R} = \boldsymbol{\gamma_0} + \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon_R}$.

5. Applications to Orthopedic Trial Data

The score test statistic in (9) was applied to the analysis of data from an orthopedic clinical trial, which was a prospective, multicenter, randomized controlled trial conducted in the Unites States. The data was a little bit modified to be used for this example. The new device was to treat chronic low back pain. The effectiveness was assessed using mobility score, which was measured 5, 10, 15, 20, 25 weeks after surgery to implant a device. A total of 215 subjects were enrolled in 16 investigational sites, and 108 subjects were randomized to treatment and 107 to control group. In addition to the treatment effect, it was of interest to find out whether the treatment effect varies from week to week, and the interaction term (Treatment by Weeks) was included in the model. The following linear mixed model was fit with baseline mobility score (B_Score) as covariate and treating investigational site as random effect.

$$Y_{ijk} = \beta_0 + \alpha \cdot T_i + \boldsymbol{\beta}_1 \cdot \mathbf{w_i} + \boldsymbol{\gamma} \cdot (T * \mathbf{w})_\mathbf{i} + \beta_2 \cdot x_i + b_k + \varepsilon_{ijk}$$

Y_{ijk} : clinical outcome from week j in subject i within site k

T_i : treatment (TRT) assigned for subject i

$\mathbf{w_i}$ = vector of weeks (5, 10, 15, 20, 25 weeks)

x_i : baseline mobility score (B_Score) for subject i

b_k : site k (random effect) ~ $N(0, \sigma_b^2)$

ε_{ijk} : error ~ $N(0, \sigma^2)$

The interaction of Treatment-by-Week (T *$\mathbf{w}$) turned out to be non-significant, and therefore the interaction term was removed. Table 1 below shows the output from SAS PROC Mixed after removing the interaction term. The P-value for the treatment effect (TRT) is 0.1206, which seems to indicate that there is no statistically significant treatment effect.

Table 1. SAS output from PROC Mixed

Effect	Week	Estimate	StdErr	DF	tValue	Probt
Intercept	_	18.3425	2.8678	15	6.40	<.0001
TRT	_	2.8601	1.8409	1053	1.55	0.1206
Week	5	-11.2988	1.2002	1053	-9.41	<.0001
Week	10	2.7238	1.2002	1053	2.27	0.0234
Week	15	2.6849	1.2002	1053	2.24	0.0255
Week	20	1.3424	1.2002	1053	1.12	0.2636
Week	25	0	.	.	.	.
B_Score	_	0.7789	0.06216	1053	12.53	<.0001

Now we consider the expanded model for the within-subject variance (error variance) using the form of $\sigma^2 \exp(\boldsymbol{\lambda}^T \boldsymbol{z_i})$. The possible candidates for $\boldsymbol{z_i}$ were age, BMI (body mass index), investigational site, and baseline mobility score. The score test statistic was

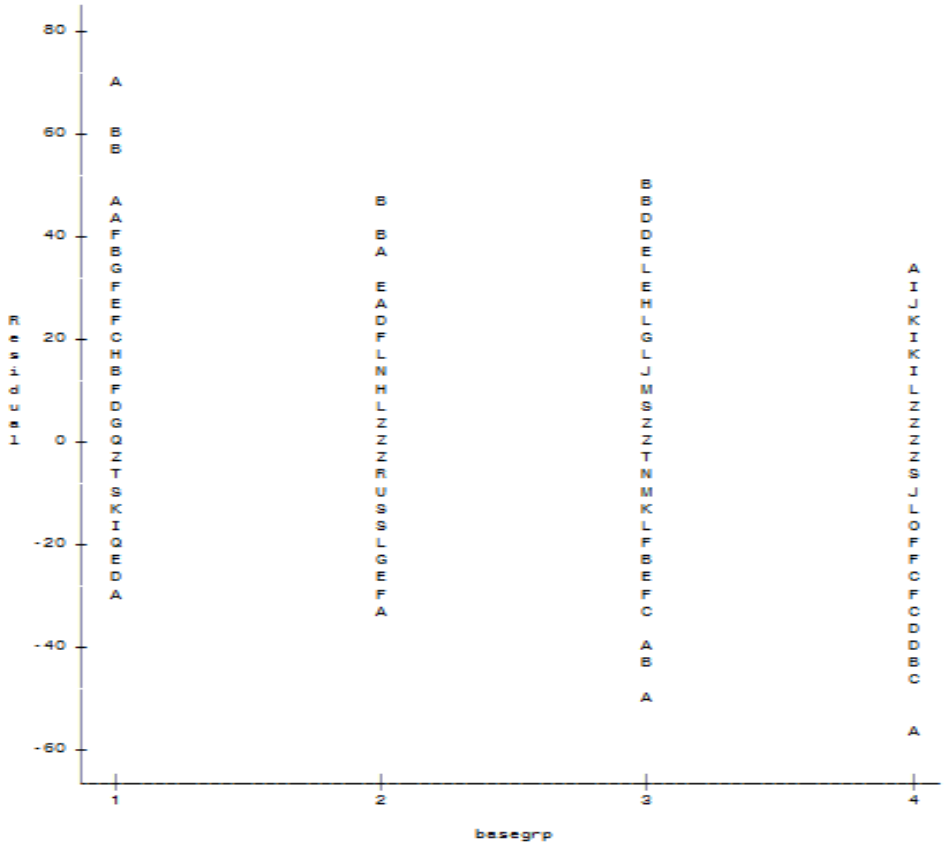
calculated for these variables, and they are shown in Table 2. The third column shows the approximate P-value for the score statistic based on the Chi-square distribution with the degrees of freedom equal to one.

Table 2. Score statistic for subject-level covariate $\boldsymbol{z_i}$

Subject-level covariate ($\boldsymbol{z_i}$)	Score statistic	P-value based on $\chi^2_{(1)}$
age group	2.252	> 0.10
BMI	2.936	> 0.05
site	0.471	> 0.30
baseline mobile score	15.202	< 0.001

The score statistic shows that the error variance may be a function of baseline mobility score, which is confirmed from the plot of residuals against baseline mobility scores. Figure 3 below shows the plot of residuals against four groups of baseline mobility score. The plot indicates that the error variance may decrease as baseline mobility score increases.

Figure 3. Plot of residual against baseline mobility score group



More accurate estimate of treatment effect may be obtained when we use the expanded model for the within-group variance, $\sigma^2 \exp(\boldsymbol{\lambda}^T \boldsymbol{z}_i)$ with baseline mobility score as $\boldsymbol{z}_i$. However, it involves calculating the maximum likelihood estimate of $\boldsymbol{\lambda}$ which may be difficult with the commercially available software. Instead, we used the Group option in PROC Mixed which is often used to handle the heterogeneity problem. When we use the

Group option, we obtain the following SAS output from PROC Mixed in Table 3. The P-value for the treatment effect (0.0402) indicates that the treatment effect is now statistically significant at the significance level of 5%.

Table 3. SAS output from PROC Mixed with Group Option

Effect	Week	Estimate	StdErr	DF	tValue	Probt
Intercept	_	15.8948	2.8318	15	5.61	<.0001
TRT	_	3.4886	1.6980	1053	2.05	0.0402
Week	5	-10.5932	1.1891	1053	-8.91	<.0001
Week	10	2.9775	1.1891	1053	2.50	0.0124
Week	15	2.8714	1.1891	1053	2.41	0.0159
Week	20	1.4452	1.1891	1053	1.22	0.2245
Week	25	0	.	.	.	.
B_Score	_	0.8025	0.06245	1053	12.85	<.0001

6. Conclusion

In linear mixed models, the homogeneity for the error structure is almost automatically taken for granted, and little attention has been devoted to checking the impact of the incorrect correlation structure on the parameter estimates.

We've shown that the standard error of the regression coefficient may be over-estimated if the within-cluster variance depends on a cluster-level covariate and this dependence is not modeled. We developed a diagnostic test based on score statistic, and it appears that it can be useful in detecting heterogeneous within-cluster variances. The future work may include investigating the small-sample behaviour of the null distribution of score statistic developed in this paper.

Acknowledgements

The author would like to express thanks to Nelson Lu who helped with SAS Macro, and Kevin Lee who provided the dataset from a medical device trial.

References

Cook, R.D. and Weisberg, S. (1983). Diagnostics for Heteroscedasticity in Regression. Biometrika. 70, 1-10.

Cox, D.R. and Hinkley (1974). Theoretical Statistics, London: Chapman Hall.

Heagerty, P. (1999) Marginally specified logistic-normal models for longitudinal binary data. stimates and generalized linear mixed models, Biometrics, 55, 688-98.

Heagerty, P. and Kurland, B. (2001). Misspecified maximum likelihood estimates and generalized linear mixed models, Biometrika. 88, 973-85.

James, W. and Stein, C. (1960). Estimation with quadratic loss. In Proc. Fourth Berkley Symp. Math. Statis. Probab. 1. 361-380. University of California Press, Berkley.

Laird, N. and Ware, J. (1982). Random-Effects Models for Longitudinal Data, Biometrics 38, 963-974.

Lin, X, Raz, J. and Harlow, S. (1997). Linear Mixed Models with Heterogeneous within-Cluster Variances, Biometrics, Vol. 53, No. 3, 910-923.

Serfling (1980). Approximation Theorem of Mathematical Statistics, New York: Wiley.