SEQUENTIAL COMPARISON OF CHANGES
WITH REPEATED MEASUREMENTS DATA

Jae Won Lee
and
David L. DeMets

Department of Human Oncology
and
Department of Statistics

UNIVERSITY OF WISCONSIN-MADISON
SEQUENTIAL COMPARISON OF CHANGES WITH REPEATED MEASUREMENTS DATA

Jae Won Lee
and
David L. DeMets

University of Wisconsin-Madison
Biostatistics Center

Biostatistics Technical Report No. 53

October 1989
ABSTRACT

There are many clinical trials in which the patients enter sequentially, and a response variable is measured repeatedly over time for each patient. A group sequential procedure is proposed for comparing the rates of change between two treatment groups. Some existing procedures for testing the equality of means between two treatment groups with repeated measurements data, such as those proposed by Armitage et al. (1985) and Geary (1988), can be interpreted as special cases of the proposed procedure. Under a linear mixed effects model, the asymptotic joint distribution of the sequentially computed statistics is derived. Construction of the group sequential boundaries is based on this distribution theory. The extension to non-normal distributions of errors is also discussed. The proposed procedure allows for using either a sequential method of Slud & Wei (1982) or that of Lan & DeMets (1983). This procedure is then applied to repeated bone density measurements on a sample of 74 middle-aged women, and implemented by the method of Lan & DeMets (1983).

KEY WORDS: Asymptotic joint distribution; Group sequential testing; Linear mixed effects model; Rate of change; Repeated measurements.
1. Introduction

In clinical trials with sequential patient entry, fixed sample size designs are unjustified on ethical grounds and (continuous) sequential designs are often impracticable. One practical solution is a group sequential design so that the decision to stop the trial or continue is based on repeated significance tests of accumulated data after each group is evaluated. Thus, a trial that shows early benefit or unexpected toxicity mandates serious consideration for early termination. When statistical analyses are performed repeatedly, some adjustment to the size of the critical value at each analysis has to be made to maintain the probability of type I error at a specified level (cf. Armitage, McPherson & Rowe, 1969). Pocock (1977, 82) introduced and described the group sequential approach with constant level critical values or boundaries at each analysis. O'Brien & Fleming (1979) proposed a different method to construct discrete sequential boundaries for clinical trials where the critical values decrease as the number of analyses are performed. Sequential testing for equality of two survival distributions was considered by Slud & Wei (1982), who also developed a class of sequential boundaries. Lan & DeMets (1983) proposed a more flexible way to construct discrete sequential boundaries based on the choice of the type I error spending rate function, which does not require equal increments of information or that the number of analyses be specified in advance. A brief review of previous work on the group sequential procedure was provided
by DeMets (1987).

Most work related to group sequential testing assumes that individual measurements are statistically independent. When the patients enter the study sequentially and a response variable is measured for each patient at successive follow-up visits, the assumption of independence is no longer reasonable. An adjustment was made by Armitage, Stratton & Worthington (1985) to allow for the ratio of between-patient to within-patient variance and for possible first-order autocorrelation. Geary & Armitage (1987) and Geary (1988) proposed a sequential testing procedure using a 4-parameter model, more general than the model considered by Armitage et al. In their procedures, however, there are various strong assumptions. One is that all the patients enter the study at the same time. It is further assumed that there are no missing values in the data, and the measurements on the patients are assumed to be recorded at equally spaced times. There are many situations in which these assumptions are not satisfied. One interesting example is from the study of calcium supplement effects on bone density (cf. Smith et al., 1989), where 74 postmenopausal women are sequentially entered and assigned to either the calcium or the placebo groups. Here, the days on which the bone density measurements were taken vary from woman to woman, and thus the missing values and the staggered entry should be allowed. The first patient entered the study in April of 1978, and the last entered in July of 1978.
In this paper, we propose a sequential procedure, which does not require any
of above assumptions, to test the equality of the rates of change, or the means as
a special case, between two treatment groups. All the repeated measurements data
are generated from a linear mixed effects model, which has the 4-parameter model
of Geary as a special case. The estimation of the rate of change in linear mixed
effects model is briefly reviewed in § 2. In § 3, a group sequential procedure in a
linear mixed effects model is presented. The extension to non-normal distributions
of errors is discussed in § 4. In § 5, the proposed method is illustrated with the bone
density example.

2. Linear Mixed Effects Model

Most stochastic models for serial measurements can be classified as full
multivariate models or multi-stage random effects models. Multivariate models
with general covariance structure are often difficult to apply to highly unbalanced
data, whereas two-stage linear random effects models can be used easily (cf.
Ware, 1985). Two-stage linear random effects models are based on explicit
identification of individual and population characteristics. Laird & Ware (1982)
used a representation developed by Harville (1977) to define a family of models
for serial measurements that includes most repeated-measures models. Population
parameters, individual effects, and within-person variation are introduced at stage
1, and between-person variation at stage 2.
**Definition 1** (Linear Mixed Effects Model) Let $\alpha$ denote a $p \times 1$ vector of unknown population parameters and $X_i$ be a known $n_i \times p$ design matrix linking $\alpha$ to response $y_i = (y_{i1}, \ldots, y_{im})$. Let $b_i$ denote a $r \times 1$ vector of unknown individual effects and $Z_i$ a known $n_i \times r$ design matrix linking $b_i$ to response $y_i$. For measured multivariate normal data, we propose the following model:

**Stage 1.** For each individual unit $i$, $i = 1, \ldots, m$,

$$y_i = X_i \alpha + Z_i b_i + e_i$$

where $e_i$ is distributed as $N(0, R_i)$ (normal with mean 0 and covariance matrix $R_i$).

Here $R_i$ is an $n_i \times n_i$ positive-definite covariance matrix; it depends on $i$ through its dimension $n_i$, but the set of unknown parameters in $R_i$ will not depend on $i$. At this stage, $\alpha$ and $b_i$ are considered fixed, and the $e_i$ are assumed to be independent.

**Stage 2.** The $b_i$ are distributed as $N(0, D)$, independently of each other and of the $e_i$. Here $D$ is $r \times r$ positive-definite covariance matrix. The population parameters, $\alpha$, are treated as fixed effects.

Let $\theta$ be defined as a $q$-vector of variance parameters found in $R_i$, $i = 1, \ldots, m$, and $D$. When all covariance parameters are known, the standard estimators for $\alpha$
and $b_i$, denoted by $\hat{\alpha}$ and $\hat{b}_i$ respectively, are given by

$$\hat{\alpha} = \left( \sum_{i=1}^{m} X_i^T W_i X_i \right)^{-1} \sum_{i=1}^{m} X_i^T W_i y_i$$

and

$$\hat{b}_i = D Z_i^T W_i (y_i - X_i \hat{\alpha})$$

where $W_i^{-1} = V_i = \text{Var}(y_i) = R_i + Z_i D Z_i^T$. $\hat{\alpha}$ maximizes the likelihood based on the marginal distribution of the data. The expression for $\hat{b}_i$ was derived by Harville (1976). $\hat{b}_i$ is an empirical Bayes estimator since it has the form $\hat{b}_i = E(b_i | y_i, \hat{\alpha}, \theta)$.

When the covariance matrices are unknown, and thus the covariance component $\theta$ is unknown, we can estimate $\alpha$ and $b_i$ by using above equation, replacing each $W_i$ by $\hat{W}_i = \hat{V}_i^{-1} = (\hat{R}_i + Z_i \hat{D} Z_i^T)^{-1}$. We denote these estimates by $\hat{\alpha}(\hat{\theta})$ and $\hat{b}_i(\hat{\theta})$. In estimating the variance component $\theta$, we consider both maximum likelihood (ML) and restricted maximum likelihood (RML) estimator. The ML estimate for $\theta$, say $\hat{\theta}_M$, is obtained by maximizing the log-likelihood corresponding to the marginal density of $y$ for $\alpha$ and $\theta$. The maximum likelihood estimates $(\hat{\alpha}_M, \hat{\theta}_M)$ satisfy $\hat{\alpha}_M = \hat{\alpha}(\hat{\theta}_M)$, and setting $b^T = (b_1^T, \ldots, b_m^T)$ and $\hat{b}_M = E(b | y, \hat{\alpha}_M, \hat{\theta}_M)$ gives $\hat{b}_M = \hat{b}(\hat{\theta}_M)$ (cf. Laird and Ware, 1982). The ML estimate for $\theta$ is biased downwards because it fails to take into account the degrees of freedom lost in estimating $\alpha$. The RML estimate for $\theta$, say $\hat{\theta}_R$, is not biased.
The RML estimate is obtained by maximizing the likelihood of $\theta$ based, not on $y$ as in maximum likelihood, but on any full-rank set of error contrasts.

The ML, or RML, estimates for $\alpha$ and $\theta$ of the random effects model can be obtained either by the EM algorithm or by such second-order algorithms as Newton-Raphson. Lindstrom & Bates (1988) developed an efficient implementation of Newton-Raphson algorithm and concluded that a well implemented NR algorithm is preferable to the EM algorithm.

3. Sequential Comparison of Changes in Linear Mixed Effects Model

We now discuss how to compare sequentially the rates of change between two treatment groups. Let $K$ be the maximum number of interim analyses. The $m_{1k}$ subjects receive a treatment 1 and the $m_{2k}$ subjects receive a treatment 2 up to $k$-th, $k = 1, \ldots, K$, interim analysis. We postulate a linear mixed effects model with different slopes and intercepts for two treatments. That is, for subjects who receive treatment 1,

$$y_i = X_i \alpha_1 + Z_i b_i + e_i \quad i = 1, \ldots, m_{1K}$$

(3.1)
and for subjects who receive treatment 2,

$$y_i = X_i \alpha_2 + Z_i b_i + e_i \quad i = m_{1K} + 1, \ldots, m_{1K} + m_{2K}$$

where

$$\alpha_1 = (\beta_{01}, \beta_1)^T, \alpha_2 = (\beta_{02}, \beta_2)^T, X_i = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}$$

and $e_i = (e_{i1}, \ldots, e_{in_i})^T \sim N_{m_i}(0, R_i)$. $b_i = (b_{i1}, \ldots, b_{ir})^T$ is distributed as $N_r(0, D)$ and $Z_i$ is a known $n_i \times r$ design matrix linking $b_i$ to $y_i$. When $Z_i = X_i$, the above model is called a linear growth curve model. Here, $x_{ij}$ represents the $j$-th measurement time of the $i$-th subject. Let $t_k, k = 1, \ldots, K$, be defined as the time when $k$-th interim analysis is done. Suppose that the $i$-th subject ($i = 1, \ldots, m_{1k}$, for treatment 1 and $i = m_{1K} + 1, \ldots, m_{1K} + m_{2k}$, for treatment 2 at $k$-th interim analysis) visits the clinic $n_{i(k)}$ times between $t_{k-1}$ and $t_k$ (see Figure 3.1 for the data structure). Note that there is no need for the $n_{i(k)}$'s be the same for all subjects,
interim analysis

which means that we allow missing values. Staggered entry and unequally spaced measurements times are also allowed. We further define $\hat{\beta}_1(k) \{ \hat{\beta}_2(k) \}, k = 1, \ldots, K$ as the generalized least square estimate of the expected rate of change $\beta_1 \{ \beta_2 \}$ from the accumulated data up to the $k$-th analysis, say, until $t_k$, in treatment group 1 {treatment group 2 }. For example, $\hat{\beta}_1(2)$ is estimated from the first
\( \sum_{i=1}^{m_i} (n_{i(1)} + n_{i(2)}) \) observations up to the second analysis in the treatment group 1.

Our goal is to test the null hypothesis of no difference in slopes, that is, \( H_0 : \beta_1 = \beta_2 \).

We use the normalized version of \( \hat{\beta}_1(k) - \hat{\beta}_2(k) \) as our test statistic, say \( S(k) \), at the \( k \)-th interim analysis (\( 1 \leq k \leq K \)). We reject the null hypothesis that the two treatment effects are equal at the \( k \)-th interim analysis if \( |S(k)| > c_k, k = 1, \ldots, K \), where \( c_k \) is the critical value at the \( k \)-th interim analysis.

When we assume that the variance component \( \theta \) is known, under the null hypothesis that slope is unaffected by the choice of treatment,

\[
(\hat{\beta}_1(1) - \hat{\beta}_2(1), \hat{\beta}_1(2) - \hat{\beta}_2(2), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T \sim N_K(0, \Sigma) \tag{3.2}
\]

where \( K \times K \) matrix \( \Sigma = MV_1^* M^T + LV_2^* L^T \), and \( M, L, V_1^* \) and \( V_2^* \) are defined in Appendix A. Proof is also given in Appendix B. Now consider the case in which the variance component \( \theta \) is unknown and needs to be estimated. As discussed in §2, \( \hat{\alpha}_M \) is equal to \( \hat{\alpha}(\hat{\theta}_M) \) if the ML estimate is used, and \( \hat{\alpha}_R \) is equal to \( \hat{\alpha}(\hat{\theta}_R) \) if RML estimate is used. Since the maximum likelihood estimators \( \hat{\theta}_M \) and \( \hat{\theta}_R \) are consistent estimators of \( \theta \), it can be easily shown by Slutsky theorem and the continuous mapping theorem (i.e. a continuous function of the consistent estimator is also consistent estimator) that both \( \hat{\alpha}_M \) and \( \hat{\alpha}_R \) converge in probability to \( \alpha \) for sufficiently large \( m_{1k} \) and \( m_{2k} \), \( k = 1, \ldots, K \), the number
of subjects in each treatment group at \( k \)-th interim analysis, where \( \hat{\alpha} \) is the estimate of \( \alpha \) in the case of known variance component. Note that \( \hat{\beta}_1 = (0,1)\hat{\alpha}_1 \) and \( \hat{\beta}_2 = (0,1)\hat{\alpha}_2 \), where \( \alpha_1 \) and \( \alpha_2 \) are defined in (3.1). By the continuous mapping and Slutzky theorems, under the null hypothesis \( H_0 : \beta_1 = \beta_2 \), we can show that \( (\hat{\beta}_1(1) - \hat{\beta}_2(1), \hat{\beta}_1(2) - \hat{\beta}_2(2), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T \) has asymptotically a multivariate normal distribution with mean \( \mathbf{0} \) and variance–covariance matrix \( \hat{\Sigma} \) as \( m_{1k}, m_{2k} \to \infty \), \( k = 1, \ldots, K \), where \( \hat{\Sigma} \), a consistent estimate of \( \Sigma \), is obtained by replacing \( \mathbf{V}_i = \text{var}(y_i) = \mathbf{R}_i + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T \) by its consistent estimate \( \hat{\mathbf{V}}_i = \mathbf{R}_i + \mathbf{Z}_i \hat{\mathbf{D}} \mathbf{Z}_i^T \).

The exit probabilities \( \pi_1, \ldots, \pi_K \), summing to the overall significance level \( \alpha \), are defined such that under \( H_0 \):

\[
P_0\{\left| S(1) \right| \leq c_1, \ldots, \left| S(k-1) \right| \leq c_{k-1}, \left| S(k) \right| > c_k \} = \pi_k
\]

and hence \( P_0\{\left| S(k) \right| > c_k \text{ for some } 1 \leq k \leq K \} = \alpha \). Lan & DeMets (1983) proposed a flexible way to compute these exit probabilities through their type I error spending function. From the knowledge of the multivariate normal distribution of

\[
\left( \frac{\hat{\beta}_1(1) - \hat{\beta}_2(1)}{\hat{\Sigma}_{11}}, \ldots, \frac{\hat{\beta}_1(K) - \hat{\beta}_2(K)}{\hat{\Sigma}_{KK}} \right)^T = (S(1), \ldots, S(K))^T
\]

we can construct critical values \( \{c_k, k = 1, \ldots, K\} \) recursively to obtain a desired
type I error (see Appendix B).

Note that if \( X_i = Z_i = 1 = (1, \ldots, 1)^T \), and \( \alpha_1, \alpha_2 \) and \( b_i \) is the scalar \( \mu_1, \mu_2 \) and \( a_i \) for all \( i \), respectively, then model (3.1) reduces to a 4-parameter model, which was used by Geary (1988). Hence, testing the null hypothesis of no difference in means between two treatments in a 4-parameter model is a special case of testing the null hypothesis of no difference in slopes between two groups in a linear mixed effects model.

4. Extension to non-normal distributions of errors

The theory presented so far assumed normal distributions of error. In this section, our proposed group sequential testing procedure is extended to non-normal situations. For simplicity, we assume that the number of subjects in each group is fixed in the course of the trial, that is, \( m_{1k} = m_1 \) and \( m_{2k} = m_2 \) for all \( k = 1, \ldots, K \). Generalization to the sequential entry of the subjects is straightforward. The same model as that in § 3 is assumed, that is (3.1), except that the distributions of \( b_i \) and \( e_i \), \( i = 1, 2, \ldots, m_1 + m_2, k = 1, \ldots, K \) are unknown. However we still assume that \( b_i \) has variance \( \mathbf{D} \) and \( e_i \) has variance \( \mathbf{R}_i \) for all individuals, i.e. \( i = 1, 2, \ldots, m_1 + m_2 \).

To derive the asymptotic normality of our test statistics, we need the following two conditions: (multivariate version of the Lindberg–Feller conditions):
\[(C1) \quad \sum_{i=1}^{m_1} A_i V_i A_i^T / m_1 \to \Sigma_1 \quad \text{and} \quad \sum_{i=m_1+1}^{m_1+m_2} A_i V_i A_i^T / m_2 \to \Sigma_2 \]

\[(C2) \quad \frac{1}{m_1} \sum_{i=1}^{m_1} \int_{\|z - A_i X_i \alpha_i\| > \varepsilon \sqrt{m_1}} \|z - A_i X_i \alpha_1\|^2 dF_i(z) \to 0 \text{ as } m_1 \to \infty \text{ for each } \varepsilon > 0 \]

and

\[\frac{1}{m_2} \sum_{i=m_1+1}^{m_1+m_2} \int_{\|z - A_i X_i \alpha_2\| > \varepsilon \sqrt{m_1}} \|z - A_i X_i \alpha_2\|^2 dF_i(z) \to 0 \text{ as } m_2 \to \infty \text{ for each } \varepsilon > 0 \]

where $F_i, i = 1, \ldots, m_1 + m_2$, is the distribution function of $z_i$. $A_i$ and $z_i$ are defined in Appendix C.

When we assume that the variances are known, under the null hypothesis that the slope is unaffected by the choice of treatment, i.e., $H_0 : \beta_1 = \beta_2$, and under the conditions above (C1 & C2),

\[(\hat{\beta}_1(1) - \hat{\beta}_2(1), \hat{\beta}_1(2) - \hat{\beta}_2(2), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T \xrightarrow{d} N_k(0, \Sigma) \text{ as } m_1, m_2 \to \infty \quad (4.1)\]

where $\Sigma = MV_1^* M^T + LV_2^* L^T$, and $M, L, V_1^*$ and $V_2^*$ are defined in Appendix A.

Proof is also given in Appendix C.
In the case that variances, $\mathbf{R}_i$ and $\mathbf{D}$, are unknown, but the consistent estimates of variances are available, we use $\hat{V}_i = \hat{\mathbf{R}}_i + \mathbf{X}_i \hat{\mathbf{D}} \mathbf{X}_i^T = \hat{\mathbf{W}}_i^{-1}$ instead of $V_i$. As discussed at the end of § 3, under the null hypothesis $H_0: \beta_1 = \beta_2$,

$$(\hat{\beta}_1(1) - \hat{\beta}_2(1), \hat{\beta}_1(2) - \hat{\beta}_2(2), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T \overset{d}{\to} N_k(\mathbf{0}, \hat{\Sigma}) \text{ as } m_1, m_2 \to \infty \quad (4.2)$$

where $\hat{\Sigma}$ is a consistent estimate of $\Sigma$. In conclusion, the results for the normal case, discussed in § 3, can be applied as a reasonable approximation to the non-normal case.

5. A Numerical Example

The data were taken from a study of calcium supplement effects on bone density (Smith et al., 1989), where 37 postmenopausal women received calcium and the other 37 women received a placebo. Ten or eleven observations are available for each woman, and the days on which the bone density measurements were taken vary from woman to woman. We assume that bone density is decreasing linearly in time. First four subjects in each group are shown in Figure 5.1. While this trial was conducted as a fixed sample size trial, we shall use these data to illustrate how it may have been conducted as a group sequential design using the method of this paper. We postulated a linear growth curve model, which is a special case of the linear mixed effects model, with different slopes and intercepts for two treatments. This model
is expressed in (3.1), where \( Z_i = X_i \) for all \( i = 1, \ldots, 74 \). Here, we assume that 
\[
b_i = (b_{i1}, b_{i2})^T \sim N_2(0, D) \quad \text{and} \quad e_i = (e_{i1}, \ldots, e_{in_i})^T \sim N_{n_i}(0, I_{n_i \times n_i} \sigma^2),
\]
and \( \sigma^2 \) are unknown. Note that \( R_i \), defined in (3.1), is restricted to \( I_{n_i \times n_i} \sigma^2 \), which means the conditional independence between successive measurements. Assume that the data monitoring committee met once a year, and each interim analysis was done on the first day of the year, i.e. \( t_1 \) is January 1 of 1979 and \( t_2 \) is January 1 of 1980 etc.. Since it is not so difficult to estimate the maximum possible number of observations, \( N_{max} \), at the start of trial in this case, one choice of the group sequential method would be due to Lan & DeMets (1983). We need to predetermine the ‘error spending’ function \( \alpha(t), 0 \leq t \leq 1, \) such that \( \alpha(0) = 0 \) and \( \alpha(1) = \alpha \). This function allocates the amount of type I error that can be spent until scaled time \( t \). Defining the ‘information’ time \( t_k^* = n(k)/N_{max} \), where \( n(k) \) is total number of observations up to \( k \)-th interim analysis, we set \( \pi_k = \alpha(t_k^*) - \alpha(t_{k-1}^*) \) and solve sequentially for \( c_1, \ldots, c_k \). Note that here \( c_k \) depends only on \( n_1, \ldots, n_k \), where \( n_i \) is the number of observations in the \( i \)-th group, and there is no need to specify \( K \), the total number of interim analyses, in advance (cf. Lan & DeMets, 1983).

We estimated the maximum number of observations as 814 (=74 \times 11) and set 
\( \alpha(t) = \alpha t^{2.5} \), which gives similar boundaries to those of O’Brien & Fleming (1979) and smaller expected sample size (cf. Jennison & Turnbull, 1989). The design consideration of the choice of \( \alpha(t) \) was also discussed by Kim & DeMets (1987).
The experiment started in the spring of 1978 and ended in the summer of 1982, i.e. \( K = 5 \), and all 74 women entered the trial in 1978, i.e. before the first interim analysis.

The efficient implementation of Newton–Raphson (NR) algorithm, developed by Lindstrom & Bates (1988), was used to estimate the parameters, i.e. \( \alpha_1, \alpha_2 \) and \( b_i, i = 1, \ldots, 74 \) in (3.1). As discussed in §3, under the null hypothesis that two slopes are equal, \((\hat{\beta}_1(1) - \hat{\beta}_2(1), \ldots, \hat{\beta}_1(5) - \hat{\beta}_2(5))^T\) have asymptotically multivariate normal distribution with mean \( 0 \) and covariance matrix \( \Sigma \), a consistent estimate of \( 5 \times 5 \) matrix \( \Sigma \). We can construct critical values \( \{c_k, k = 1, \ldots, 5\} \) at each yearly analysis recursively, from the knowledge of the joint distribution of our statistics \((S(1), \ldots, S(K))^T\), proposed in § 3, by using the subroutine MULNOR developed by Schervish (1984).

Let the significance level \( \alpha \) be equal to 0.05. For the first interim analysis, the exit probability \( \pi_1 = 0.0017 \) was calculated and we estimated the test statistic \(|S(1)|\) to be 0.3768 which is less than the critical value \( c_1 = 3.1382 \). For the second interim analysis, \( \pi_2 = 0.0086 \) and \(|S(2)|\) is equal to 3.1392 which exceeds the critical value \( c_2 = 2.6244 \). The above result shows the significant differences in slopes between two treatment groups after the second interim analysis, i.e. after two years, and early stopping of the trial could have been considered if this procedure had been available. Note that the null hypothesis of equal slopes is also rejected when we conduct the
fixed sample size test at the end of trial, since $|S(5)| = 2.1895 > 1.96 = Z_{0.025}$.

![Graphs showing bone density over time for calcium and placebo groups.]

**Figure 5.1** First four subjects in each group

6. Concluding Remarks

In this paper, we have proposed a group sequential procedure to test the equality of the rates of change (slopes) between two treatment groups with repeated measurements data, generated from a linear mixed effects model. The linear mixed effects model for repeated measurements data has been successful because it can handle unbalanced data, missing data, and jointly dependent random effects.

The proposed procedure can be widely applicable for its flexibility. Staggered entry of subjects into the trial and missing values in the data are allowed. Repeated measurements on the subjects can be recorded at unequally spaced times. Further-
more, the missing data need not to be completely at random. The non-response mechanism is said to be \textit{ignorable}, when the probability of non-response depends on covariates and the observed response, but not the unobserved response. The proposed method is still valid as long as the non-response mechanism is ignorable, since ignorable non-response mechanism does not have any effect on the likelihood based inferences (cf. Laird, 1988).

The proposed procedure is also of practical use in the sense that all the estimates needed to perform interim analyses can be obtained by using existing algorithms (cf. Schervish, 1984; Lindstrom & Bates, 1988) and a simple FORTRAN program. The ‘conditional-independence’ model, i.e. $R_i = I\sigma^2$, was used in the example. The efficient implementation of Newton–Raphson algorithm can be easily extended to more general form of $R_i$ (cf. Lindstrom & Bates, 1988). It can also be shown that $\text{cov}(S(k), S(k')) = \text{var}(S(k'))$ for $k \leq k'$, so that only double integrals are needed in the computation.
References


Appendix A: Some Necessary Notation

For each individual, \( i, y_i | b_i \sim N(X_i \alpha + X_i b_i, R_i) \) and \( b_i \sim N(0, D) \)

\( \alpha \) is defined as \( \alpha_1 \) in treatment 1 (\( i = 1, \ldots, m_{1K} \)) and as \( \alpha_2 \) in treatment 2

(\( i = m_{1K} + 1, \ldots, m_{1K} + m_{2K} \)). We can write the model for all data in a matrix

form by defining

\[
\begin{align*}
    y_1^* &= (y_1^T, \ldots, y_{m_{1K}}^T)^T, \\
    X_1^* &= (X_1^T, \ldots, X_{m_{1K}}^T)^T, \\
    y_2^* &= (y_{m_{1K}+1}^T, \ldots, y_{m_{1K}+m_{2K}}^T)^T, \\
    X_2^* &= (X_{m_{1K}+1}^T, \ldots, X_{m_{1K}+m_{2K}}^T)^T, \\
    V_1^* &= \text{Diag}(V_1, \ldots, V_{m_{1K}}), \\
    V_2^* &= \text{Diag}(V_{m_{1K}+1}, \ldots, V_{m_{1K}+m_{2K}})
\end{align*}
\]

where \( V_i = \text{Var}(y_i) = R_i + X_i D X_i^T \), \( i = 1, \ldots, m_{1K} + m_{2K} \). Thus the marginal

distributions of \( y_q^* \), \( q = 1, 2 \), are

\[
y_q^* \sim N(X_q^* \alpha_q, V_q^*).
\]

When the variance component \( \theta \) is known, the generalized least squares
estimator for $\alpha_q, q = 1, 2$, is

$$\hat{\alpha}_q = ((X_q^*)^T W_q^* X_q^*)^{-1} (X_q^*)^T W_q^* y_q^*$$

where $W_q^* = V_q^*$. These are equivalent to $\hat{\alpha}$, defined in §2. We define $n_i(k), i = 1, \ldots, m_{1K} + m_{2K}, k = 1, \ldots, K$ as the number of observations of individual $i$ in the $k$-th group, i.e. between the $(k-1)$-th interim analysis and the $k$-th interim analysis. We also define $S_i(k)$ as the accumulated number of observations of individual $i$ up to the $k$-th interim analysis, i.e. $S_i(k) = n_i(1) + \ldots + n_i(k)$. Note that $n_i = n_i(1) + \ldots + n_i(K) = S_i(K), i = 1, \ldots, m_{1K} + m_{2K}$. We further define $X_i^{(k)}, X_1^{*(k)}, X_2^{*(k)}, W_1^{(k)}, W_2^{(k)}, y_1^{(k)}$ and $y_2^{(k)}$ from the accumulated data up to the $k$-th analysis in the same way as $X_i, X_1, X_2, W_1, W_2, y_1$ and $y_2$, respectively, from the entire data. That is

$$X_i^{(k)} = \begin{bmatrix} 1 & x_{i,1} \\ \vdots & \vdots \\ 1 & x_{i,S_i(k)} \end{bmatrix} \quad i = 1, \ldots, m_{1K} + m_{2K}$$

$$X_1^{*(k)} = (X_1^{(k)T}, \ldots, X_m^{(k)T})^T,$$

$$X_2^{*(k)} = (X_{m_{1K}+1}^{(k)T}, \ldots, X_{m_{1K}+m_{2K}}^{(k)T})^T,$$
\[
\begin{align*}
y_i^{(k)} &= (y_i, 1, \ldots, y_i, S_i(k))_T, \\
y_1^{*(k)} &= (Y_1^{(k)} T, \ldots, Y_{m_{1K}^*}^{(k)} T)_T, \\
y_2^{*(k)} &= (Y_{m_{1K}^*+1}^{(k)} T, \ldots, Y_{m_{1K}^*+m_{2K}^*}^{(k)} T)_T.
\end{align*}
\]

Note also that

\[
X_i^{(K)} = X_i, \ y_i^{(K)} = y_i, \ i = 1, \ldots, m_{1K} + m_{2K}, \ X_1^{*(K)} = X_1^*, \ X_2^{*(K)} = X_2^*.
\]

\[
W_1^{*(K)} = W_1^*, \ W_2^{*(K)} = W_2^*, \ y_1^{*(K)} = y_1^* \text{ and } y_2^{*(K)} = y_2^*.
\]

We can find that \(\hat{\beta}_1(k)\) and \(\hat{\beta}_2(k)\), defined in model (3.1), can be expressed as follows;

\[
\begin{align*}
\hat{\beta}_1(k) &= (0, 1) \times ((X_1^{*(k)} T W_1^{*(k)} X_1^{*(k)})^{-1} (X_1^{*(k)} T W_1^{*(k)} y_1^{*(k)}), \\
\hat{\beta}_2(k) &= (0, 1) \times ((X_2^{*(k)} T W_2^{*(k)} X_2^{*(k)})^{-1} (X_2^{*(k)} T W_2^{*(k)} y_2^{*(k)}).
\end{align*}
\]

We introduce the special matrices \(J_{1(k)}\) and \(J_{2(k)}\) for each of the two cases; all the subjects enter the trial before the first interim analysis is conducted or not. Consider first the case in which all the subjects enter the trial before the first analysis is done,
i.e. \( m_{1k} = m_1 \) and \( m_{2k} = m_2 \) for all \( k = 1, \ldots, K \). \( \mathbf{J}_{1(k)} \) and \( \mathbf{J}_{2(k)} \) are defined such that
\[
\mathbf{J}_{1(k)} = \text{Diag}(\mathbf{P}_{1(k)}, \ldots, \mathbf{P}_{m_1(k)}) \quad \text{and} \quad \mathbf{J}_{2(k)} = \text{Diag}(\mathbf{P}_{m_1+1(k)}, \ldots, \mathbf{P}_{m_1+m_2(k)})
\]
where \( \mathbf{P}_{i(k)} = [I_{S_i(k) \times S_i(k)} : \mathbf{O}_{S_i(k) \times (n_i-S_i(k))}] \), \( i = 1, \ldots, m_1 + m_2 \). Note that \( \mathbf{J}_{1(k)} \) has \( S_1(k) + \cdots + S_{m_1(k)} \) rows and \( n_1 + \cdots + n_{m_1} \) columns, and that \( \mathbf{J}_{2(k)} \) has \( S_{m_1+1(k)} + \cdots + S_{m_1+m_2(k)} \) rows and \( n_{m_1+1} + \cdots + n_{m_1+m_2} \) columns. Note also that \( \mathbf{J}_{1(K)} = \mathbf{I}_{n_1+\cdots+n_{m_1}} \) and \( \mathbf{J}_{2(K)} = \mathbf{I}_{n_{m_1+1}+\cdots+n_{m_1+m_2}} \). Now consider the case in which the subjects enter sequentially throughout the trial even after the first interim analysis is done. Let \( l_{1k} \) and \( l_{2k} \) be the number of subjects who entered the trial before \( k \)-th interim analysis in treatment group 1 and 2, respectively. Note that for any \( k \), \( k = 1, \ldots, K \),

\[
S_{i(k)} > 0 \quad \text{if} \quad 1 \leq i \leq l_{1k}, \quad m_{1K} + 1 \leq i \leq m_{1K} + l_{2k}
\]

\[
S_{i(k)} = 0 \quad \text{if} \quad l_{1k} + 1 \leq i \leq m_{1K}, \quad m_{1K} + l_{2k} + 1 \leq i \leq m_{1K} + m_{2K}
\]

where \( S_{i(k)} \) is defined as the accumulated number of observations of \( i \)-th individual up to the \( k \)-th interim analysis. Let \( \mathbf{J}_{1(k)} \), \( \mathbf{J}_{2(k)} \) and \( \mathbf{P}_{i(k)} \), \( k = 1, \ldots, K \), \( i = 1, \ldots, m_{1K} + m_{2K} \), be defined as follows:

\[
\mathbf{P}_{i(k)} = [I_{S_i(k) \times S_i(k)} : \mathbf{O}_{S_i(k) \times (n_i-S_i(k))}] , \quad 1 \leq i \leq l_{1k}-1, \quad m_{1K}+1 \leq i \leq m_{1K}+l_{2k}-1
\]
\[ \mathbf{P}_{l_{1k}}(k) = \begin{bmatrix} \mathbf{I}_{s_{l_{1k}}(k) \times s_{l_{1k}}(k)} : \mathbf{O}_{s_{l_{1k}}(k) \times (m_{1k} - s_{l_{1k}}(k) + \sum_{i=1}^{i_{1k} + 1} n_i)} \end{bmatrix}, \]

\[ \mathbf{P}_{m_{1K} + i_{2k}}(k) = \begin{bmatrix} \mathbf{I}_{s_{m_{1K} + i_{2k}}(k) \times s_{m_{1K} + i_{2k}}(k)} : \mathbf{O}_{s_{m_{1K} + i_{2k}}(k) \times (m_{1K} + i_{2k} - s_{m_{1K} + i_{2k}}(k) + \sum_{i=m_{1K} + i_{2k} + 1}^{m_{1K} + m_{2K}} n_i)} \end{bmatrix}, \]

\[ \mathbf{J}_{1(k)} = \text{Diag}(\mathbf{P}_{1(k)}, \ldots, \mathbf{P}_{l_{1k}(k)}), \]

\[ \mathbf{J}_{2(k)} = \text{Diag}(\mathbf{P}_{m_{1K} + 1(k)}, \ldots, \mathbf{P}_{m_{1K} + i_{2k}(k)}) \]

where \( n_i, i = 1, \ldots, m_{1K} + m_{2K} \) is the total number of observations of \( i \)-th individual. We further define \( \mathbf{M}(k) \) and \( \mathbf{L}(k) \) as follows:

\[ \mathbf{M}(k) = (0, 1) \times \left( (\mathbf{X}_1^{* (k)})^T \mathbf{W}_1^{* (k)} \mathbf{X}_1^{* (k)} \right)^{-1} (\mathbf{X}_1^{* (k)})^T \mathbf{W}_1^{* (k)} \]

and

\[ \mathbf{L}(k) = (0, 1) \times \left( (\mathbf{X}_2^{* (k)})^T \mathbf{W}_2^{* (k)} \mathbf{X}_2^{* (k)} \right)^{-1} (\mathbf{X}_2^{* (k)})^T \mathbf{W}_2^{* (k)} \]

It is also defined that \( \mathbf{M} \) and \( \mathbf{L} \) have \( \mathbf{M}(k) \) and \( \mathbf{L}(k), k = 1, \ldots, K \), as their \( k \)-th rows respectively.
Appendix B: Proof of (3.2)

Since \( y_1^{* (k)} = J_{1(k)}y_1^* \) and \( y_2^{* (k)} = J_{2(k)}y_2^* \), it follows that

\[
\hat{\beta}_1(k) = (0,1) \times ((X_1^{*(k)})^T W_1^{*(k)}X_1^{*(k)})^{-1}(X_1^{*(k)})^T W_1^{*(k)}J_{1(k)}y_1^* = M(k)y_1^*
\]

and

\[
\hat{\beta}_2(k) = (0,1) \times ((X_2^{*(k)})^T W_2^{*(k)}X_2^{*(k)})^{-1}(X_2^{*(k)})^T W_2^{*(k)}J_{2(k)}y_2^* = L(k)y_2^*
\]

where \( M(k) \) and \( L(k) \) are defined in Appendix A. Thus \( (\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K))^T = My_1^* \) and \( (\hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T =Ly_2^* \) where \( M \) and \( L \) have \( M(k) \) and \( L(k), k = 1, \ldots, K \), as their \( k \)-th rows respectively.

The fact that \( y_1^* \sim N_{n_{11} + \ldots + \sum_{k=1}^{K} m_{1k}}(X_1^* \alpha_1, V_1^*) \) and \( y_2^* \sim N_{n_{21} + \ldots + \sum_{k=1}^{K} m_{2k}}(X_2^* \alpha_2, V_2^*) \)

imply that

\[
(\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K))^T \sim N_K(MX_1^* \alpha_1, MV_1^* M^T)
\]

and

\[
(\hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T \sim N_K(LX_2^* \alpha_2, LV_2^* L^T)
\]

where \( \alpha_1 \) and \( \alpha_2 \) are defined in (3.2), i.e. \( \alpha_1 = (\beta_0^1, \beta_1)^T \) and \( \alpha_2 = (\beta_0^2, \beta_2)^T \). It
follows from the independence among individuals that,

\[(\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K), \hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T \sim N_{2K}(\begin{bmatrix} MX_1^* \alpha_1 \\ LX_2^* \alpha_2 \end{bmatrix}, \begin{bmatrix} MV_1^* M^T & 0 \\ 0 & LV_2^* L^T \end{bmatrix})].\]

Note that \(\text{Cov}[(\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K))^T, (\hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T] = \text{Cov}(My_1^*, Ly_2^*) = M\text{Cov}(y_1^*, y_2^*)L^T = 0\) since \(\text{Cov}(y_i, y_j) = 0\) if \(i \neq j\). Hence from the fact that \(MX_1^* = LX_2^* = [0_{K \times 1}, 1_K]\), we have \(MX_1^* \alpha_1 - LX_2^* \alpha_2 = (\beta_1 - \beta_2)1_K\), where \(1_K\) is a \(K \times 1\) vector \((1, \ldots, 1)^T\), and we can find the distribution of our test statistics:

\[
(\hat{\beta}_1(1) - \hat{\beta}_2(1), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T
= [I_{K \times K} : -I_{K \times K}](\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K), \hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T
\sim N_K((\beta_1 - \beta_2)1_K, MV_1^* M^T + LV_2^* L^T).
\]

We can conclude that, under \(H_0: \beta_1 = \beta_2\), our test statistics have a multivariate normal distribution with mean 0 and covariance matrix \(\Sigma_{K \times K} = MV_1^* M^T + LV_2^* L^T\).

This completes the proof of (3.2).

**Appendix C: Proof of (4.1)**

We need a transformation of \(y_i\) to apply the multivariate central limit theorem.
(CLT). Let

\[ z_i = A_i y_i \quad i = 1, \ldots, m_1 + m_2 \]

where \( A_i, \ i = 1, \ldots, m_1 + m_2, \) is \( K \times n_i \) matrix, hence \( z_i, \ i = 1, \ldots, m_1 + m_2, \) is \( K \times 1 \) vector. If we choose \( A_i \) such that

\[ M = [A_1 : A_2 : \ldots : A_{m_1}] \]

and

\[ L = [A_{m_1+1} : A_{m_1+2} : \ldots : A_{m_1+m_2}] \]

where \( M \) and \( L \) are defined in Appendix A, then we have that

\[ \sum_{i=1}^{m_1} z_i = \sum_{i=1}^{m_1} A_i y_i = [A_1 : \ldots : A_{m_1}] y_1^* = My_1^* \]

and

\[ \sum_{i=m_1+1}^{m_1+m_2} z_i = \sum_{i=m_1+1}^{m_1+m_2} A_i y_i = [A_{m_1+1} : \ldots : A_{m_1+m_2}] y_2^* = Ly_2^* \]

By the multivariate extension of Lindberg–Feller central limit theorem (cf. Serfling, R. J., 1980, p.30),

\[ \sum_{i=1}^{m_1} z_i \overset{d}{\rightarrow} N_K \left( \sum_{i=1}^{m_1} A_i X_i \alpha_1, \sum_{i=1}^{m_1} A_i V_i A_i^T \right) \]
and

\[ \sum_{i=m_1+1}^{m_1+m_2} z_i \xrightarrow{d} N_K(\sum_{i=m_1+1}^{m_1+m_2} A_i X_i \alpha_2, \sum_{i=m_1+1}^{m_1+m_2} A_i V_i A_i^T) \]

since we assume the following two conditions are satisfied:

(C1) \[ \sum_{i=1}^{m_1} A_i V_i A_i^T / m_1 \to \Sigma_1 \quad \text{and} \quad \sum_{i=m_1+1}^{m_1+m_2} A_i V_i A_i^T / m_2 \to \Sigma_2 \]

(C2) \[ \frac{1}{m_1} \sum_{i=1}^{m_1} \int_{||z - A_i X_i \alpha_1|| > \epsilon / \sqrt{m_1}} \|z - A_i X_i \alpha_1\|^2 dF_i(z) \to 0 \text{ as } m_1 \to \infty \text{ for each } \epsilon > 0 \]

and

\[ \frac{1}{m_2} \sum_{i=m_1+1}^{m_1+m_2} \int_{||z - A_i X_i \alpha_2|| > \epsilon / \sqrt{m_1}} \|z - A_i X_i \alpha_2\|^2 dF_i(z) \to 0 \text{ as } m_2 \to \infty \text{ for each } \epsilon > 0 \]

where \( F_i, i = 1, \ldots, m_1 + m_2 \), is the distribution function of \( z_i \).

Therefore, we find that \( My_i^* \) and \( Ly_2^* \) have asymptotically multivariate normal distributions as \( m_1, m_2 \to \infty \). That is

\[ My_i^* \xrightarrow{d} N_K(\beta_1 1_K, M V_i^* M^T) \]
and

\[ \text{Ly}_2^* \overset{d}{\sim} N_K(\beta_21_K, \text{LV}_2^*\text{L}^T). \]

Note that

\[ \sum_{i=1}^{m_1} A_iX_i\alpha_1 = \beta_11_K, \quad \sum_{i=1}^{m_1} A_iV_iA_i^T = MV_1^*M^T \]

and

\[ \sum_{i=m_1+1}^{m_1+m_2} A_iX_i\alpha_2 = \beta_21_K, \quad \sum_{i=m_1+1}^{m_1+m_2} A_iV_iA_i^T = \text{LV}_2^*\text{L}^T. \]

Since \( M \) and \( L \) are defined such that

\[ (\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K))^T = My_1^* \]

and

\[ (\hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T = Ly_2^*, \]

it follows from the independence among individuals that

\[ (\hat{\beta}_1(1), \ldots, \hat{\beta}_1(K), \hat{\beta}_2(1), \ldots, \hat{\beta}_2(K))^T \overset{d}{\sim} N_{2K} \left( \begin{bmatrix} MX_1^*\alpha_1 \cr LX_2^*\alpha_2 \end{bmatrix}, \begin{bmatrix} MV_1^*M^T & 0 \\ 0 & \text{LV}_2^*\text{L}^T \end{bmatrix} \right) \]
and thus

$$(\hat{\beta}_1(1) - \hat{\beta}_2(1), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T \overset{d}{\rightarrow} N_K((\beta_1 - \beta_2)1_K, \textbf{M}V_1^*\textbf{M}^T + \textbf{L}V_2^*\textbf{L}^T).$$

We conclude that, under $H_0: \beta_1 = \beta_2$, our test statistics $(\hat{\beta}_1(1) - \hat{\beta}_2(1), \ldots, \hat{\beta}_1(K) - \hat{\beta}_2(K))^T$ have an asymptotically multivariate normal distribution with mean 0 and covariance matrix $\Sigma_{K\times K} = \textbf{M}V_1^*\textbf{M}^T + \textbf{L}V_2^*\textbf{L}^T$ as $m_1,m_2 \to \infty$. This completes the proof of (4.1).