Published in final edited form as:

Biom J. 2011 May ; 53(3): 512–524. doi:10.1002/bimj.201000200.

Analysis of covariance with pre-treatment measurements in randomized trials under the cases that covariances and post-treatment variances differ between groups

Takashi Funatogawa^{1,*}, Ikuko Funatogawa^{1,**}, and Yu Shyr¹

¹Cancer Biostatistics Center, Vanderbilt-Ingram Cancer Center, 691 Preston Building, Nashville, TN 37232-6838, U.S.A

Abstract

When primary endpoints of randomized trials are continuous variables, the ANCOVA with pre-treatment measurements as a covariate is often used to compare two treatment groups. In the ANCOVA, equal slopes (coefficients of pre-treatment measurements) and equal residual variances are commonly assumed. However, random allocation guarantees only equal variances of pre-treatment measurements. Unequal covariances and variances of post-treatment measurements indicate unequal slopes and, usually, unequal residual variances. For non-normal data with unequal covariances and variances of post-treatment measurements, it is known that the ANCOVA with equal slopes and equal variances using an ordinary least-squares method provides an asymptotically normal estimator for the treatment effect. However, the asymptotic variance of the estimator differs from the variance estimated from a standard formula, and its property is unclear. Furthermore, the asymptotic properties of the ANCOVA with equal slopes and unequal variances using a generalized least-squares method are unclear. In this paper, we consider non-normal data with unequal covariances and variances of post-treatment measurements, and examine the asymptotic properties of the ANCOVA with equal slopes using the variance estimated from a standard formula. Analytically, we show that the actual type I error rate, thus the coverage, of the ANCOVA with equal variances is asymptotically at a nominal level under equal sample sizes. That of the ANCOVA with unequal variances using a generalized least-squares method is asymptotically at a nominal level, even under unequal sample sizes. In conclusion, the ANCOVA with equal slopes can be asymptotically justified under random allocation.

Keywords

Analysis of covariance; Post-test; Pre-test; Unbalanced allocation; Unequal variance

*Corresponding author: ikuko.funatogawa@vanderbilt.edu, Phone: +1-615-936-2572, Fax: +1-615-936-2602.

**These authors contributed equally to this work.

Conflict of Interest

The authors have declared no conflict of interest.

1 Introduction

When primary endpoints of randomized clinical trials are continuous variables, an analysis of covariance (ANCOVA) on post-treatment measurements with pre-treatment measurements as a covariate is often used to compare treatment groups. The ANCOVA has higher power to detect a treatment effect than the t -test under standard assumptions (Feldt, 1958; Follmann, 1991). The Committee for Proprietary Medicinal Products (2004), on behalf of the European Agency for the Evaluation of Medicinal Products, has recommended using the baseline of a continuous endpoint as a covariate, and using equal slopes (coefficients of pre-treatment measurements) as a primary analysis of confirmatory (typically phase III) randomized trials.

Some researchers have studied the validity of the ANCOVA for the analysis of pre- and post-treatment measurements under random allocation. Crager (1987) showed that, if pre- and post-treatment measurements are bivariate normal and the covariance matrices are equal between groups, the ANCOVA with equal slopes and equal residual variances, using an ordinary least-squares method, provides a valid significance test for the treatment effect in finite sample situations. For non-normal data with unequal covariances and variances in post-treatment measurements, Yang and Tsiatis (2001) proved that the ANCOVA with equal slopes and equal variances using an ordinary least-squares method provides an asymptotically normal estimator for the treatment effect under random allocation and showed the asymptotic variance of the estimator. These findings are important, because random allocation guarantees only equal variances of pre-treatment measurements, and unequal values of covariances and variances of post-treatment measurements are often seen in actual clinical trials, especially in placebo-controlled trials. The normality is often doubtful in actual situations because of truncation caused by inclusion criteria, outliers, and skewed data.

However, the asymptotic variance for the estimator of the treatment effect derived by Yang and Tsiatis (2001) is different from the variance estimator based on the standard formula of an ordinary least-squares method provided by standard software. It is unknown whether the ANCOVA with the variance estimator based on the standard formula asymptotically keeps a nominal type I error rate. Moreover, the residual variances in the ANCOVA usually differ between groups for the data with unequal covariances and variances in post-treatment measurements. The ANCOVA with unequal residual variances using a generalized least-squares method may be appropriate for such data, as the Welch's t -test or t -test with Satterthwaite's approximation (Satterthwaite, 1946) is used instead of the Student's t -test. However, it is unknown whether the ANCOVA using a generalized least-squares method provides an asymptotically normal estimator for the treatment effect and the ANCOVA with the variance estimator based on the standard formula asymptotically keeps a nominal type I error rate.

In this paper, we examine the above unknown properties. The structure of this paper is as follows. In Section 2, we further investigate data-generating models under random allocation and show that the assumptions of the ANCOVA with equal slopes and equal residual variances are too strong in view of data-generating models. In Section 3, we analytically

consider asymptotic properties of the ANCOVA with equal slopes in view of statistical tests that compare treatment groups for data generated under random allocation. Sections 3.1 and 3.2 consider the ANCOVA with equal and unequal variances, respectively. In Section 4, we conduct simulation studies. In Section 5, we offer a discussion.

2 Data-generating model under random allocation and motivating data

Section 2.1 describes a data-generating model that assumes only random allocation. Section 2.2 describes conceptual data-generating models that use inter- and intra-individual variabilities. In Sections 2.3 and 2.4, we show that the assumptions of the ANCOVA with equal slopes and equal residual variances are too strong for randomized clinical trials, using an actual example.

2.1 Data-generating model under random allocation

We consider a parallel-group, randomized clinical trial of two treatments. Let Y_{ijk} denote the measurement of the k th ($k = 1, 2$) time for the j th ($j = 1, \dots, n_i$) subject in the i th ($i = 1, 2$) treatment group. Y_{ij1} and Y_{ij2} are pre- and post-treatment measurements, respectively. We assume that the pair (Y_{ij1}, Y_{ij2}) has finite second moments and is distributed with the following mean and the covariance matrix:

$$\text{Mean}(Y_{ij1}, Y_{ij2}) = \begin{pmatrix} \mu_{\text{pre}} \\ \mu_{\text{post}} \end{pmatrix} \text{ and } \text{Cov}(Y_{ij1}, Y_{ij2}) = \begin{pmatrix} \sigma_{\text{pre}}^2 & \sigma_{\text{icov}} \\ \sigma_{\text{icov}} & \sigma_{\text{ipost}}^2 \end{pmatrix}. \quad (1)$$

Random allocation will guarantee equal values for μ_{pre} and σ_{pre}^2 between the groups, but will not for μ_{post} , σ_{icov} , or σ_{ipost}^2 . $\mu_{2\text{post}} - \mu_{1\text{post}}$ can be considered as the treatment effect because μ_{pre} is equal between the groups. Here, we do not assume normality.

2.2 Conceptual data-generating model under random allocation

To conceptually clarify how the data are generated in randomized trials, we decompose the above covariance matrix to inter- and intra-individual variabilities. We consider the following data-generating model:

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix} = \begin{pmatrix} \mu_{\text{pre}} \\ \mu_{\text{post}} \end{pmatrix} + \begin{pmatrix} S_{ij1} \\ S_{ij2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix},$$

$$\text{Cov}(Y_{ij1}, Y_{ij2}) = \begin{pmatrix} \sigma_{\text{S pre}}^2 & \sigma_{\text{is cov}} \\ \sigma_{\text{is cov}} & \sigma_{\text{is post}}^2 \end{pmatrix} + \begin{pmatrix} \sigma_{\varepsilon \text{ pre}}^2 & 0 \\ 0 & \sigma_{\varepsilon \text{ post}}^2 \end{pmatrix}, \quad (2)$$

where $\mu_{\text{pre}} + S_{ij1}$ and $\mu_{\text{post}} + S_{ij2}$ represent the true pre- and post-treatment measurements, respectively, for the subject. The random subject effects S_{ij1} and S_{ij2} are distributed with mean zero and the above covariance matrix. ε_{ij1} and ε_{ij2} are random errors, independent of S_{ijk} and each other, distributed with mean zero and variances $\sigma_{\varepsilon \text{ pre}}^2$ and $\sigma_{\varepsilon \text{ post}}^2$, respectively. Random allocation will guarantee equal values for μ_{pre} , $\sigma_{\text{S pre}}^2$ and $\sigma_{\varepsilon \text{ pre}}^2$ between the groups,

but will not for μ_{post} , $\sigma_{S\text{ cov}}$, $\sigma_{iS\text{ post}}^2$ or $\sigma_{i\varepsilon\text{ post}}^2$. Additionally, $\sigma_{i\varepsilon\text{ post}}^2 = \sigma_{\varepsilon\text{ pre}}^2$ may be assumed. The number of the variance parameters in this model is too large to estimate from the pre- and post-treatment measurements data. Blomqvist (1977) used this model to evaluate associations between responses to treatments and pre-treatment measurements using external information on the variances of measurement errors.

Crager (1987) considered a model additionally that assuming $S_{ij1} = S_{ij2}$ and equal $\sigma_{i\varepsilon\text{ post}}^2$ between groups. $S_{ij1} = S_{ij2}$ means that $\sigma_{iS\text{ cov}} = \sigma_{S\text{ pre}}^2$ and $\sigma_{iS\text{ post}}^2 = \sigma_{S\text{ pre}}^2$. Because of these additional assumptions, the number of the variance parameters is three, and these are estimable from the data. In this model, the (unobserved) true changes $\mu_{post} + S_{ij1} - (\mu_{pre} + S_{ij1})$ are the same amongst all subjects in each group. Thus, no inter-individual variability exists in the true change, and the true changes are independent of pre-treatment measurements. These assumptions are often unnatural in actual clinical trials. Note that the assumptions in this model result in the equal covariance matrixes between groups, that is equal covariances and variances of post-treatment measurements.

2.3 ANCOVA model

If we fit regression lines in each group to data generated by Model (1), the estimates of the slopes and the residual variances are asymptotically $\sigma_{icov}/\sigma_{pre}^2$ and $\sigma_{i\text{ post}}^2 - \sigma_{icov}^2/\sigma_{pre}^2$, respectively. Hence, if σ_{icov} is equal between the groups, the estimates of the slopes are asymptotically equal. Additionally, if $\sigma_{i\text{ post}}^2$ is equal between the groups, the estimates of the residual variances are asymptotically equal. If the covariance matrices are equal between the groups, the ANCOVA with equal slopes and equal residual variances fit the data. However, the assumptions of equal slopes and equal residual variances are often too strong in view of data-generating models, as shown in Section 2.4.

2.4 Motivating data

For an actual example of unequal covariance and post-treatment variances between groups, we show the results of a placebo-controlled, randomized trial of succimer in children (Rogan *et al.*, 2001). The blood lead levels at baseline and one week after administration for a subsample of 100 children in each group (Fitzmaurice *et al.*, 2004) are shown in Figure 1. The means of pre- and post-treatment measurements in the succimer group are 27 and 14, respectively, and those in the placebo group are 26 and 25. The covariance matrixes are as follows:

$$\begin{pmatrix} 25 & 15 \\ 15 & 59 \end{pmatrix} \text{ in the succimer group and } \begin{pmatrix} 25 & 23 \\ 23 & 30 \end{pmatrix} \text{ in the placebo group.}$$

The means and variances of the pre-treatment measurements are almost same between the groups because of random allocation, but the other elements obviously differ. If we fit regression lines in each group separately, the respective estimates of the slopes and the residual variances are 15/25 and 50 in the succimer group and 23/25 and 8.8 in the placebo group.

The data show that changes after the intervention were small in the placebo group, and the correlation between the pre- and post-treatment measurements was relatively strong. The changes after the intervention were large in the succimer group, and the amounts of change differed amongst individuals. Then, the correlation of pre- and post-treatment measurements was weak. The variance in post-treatment measurement was larger than that of the placebo group, and also compared to the variance in the pre-treatment measurement. The reason may be that true post-treatment measurements depend not only on pre-treatment measurements, but also the individual response to succimer. In a clinical trial, the response to a particular treatment varies from person to person because of known and/or unknown factors, such as genetics, clinical history and body weight. The size of this variability and dependency on pre-treatment measurement will differ amongst treatments. Thus, we should assume

different values for parameters $\sigma_{S\text{ cov}}$ and $\sigma_{iS\text{ post}}^2$ between the groups. In this case, the ANCOVA model with equal slopes and equal residual variances does not fit the data.

3 ANCOVA with equal slopes

In this section, we consider analytically the asymptotic properties of the ANCOVA with equal slopes in view of statistical tests. The data are generated by Model (1) in Section 2.1, that is, non-normal data with unequal covariances and variances in post-treatment measurements under random allocation.

The ANCOVA model being used to analyze the data is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ where } \mathbf{y} = \begin{pmatrix} y_{112} \\ \vdots \\ y_{1n_12} \\ y_{212} \\ \vdots \\ y_{2n_22} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & 0 & y_{111} \\ \vdots & \vdots & \vdots \\ 1 & 0 & y_{1n_11} \\ 1 & 1 & y_{211} \\ \vdots & \vdots & \vdots \\ 1 & 1 & y_{2n_21} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_{\text{int}} \\ \beta_{\text{trt}} \\ \beta_{\text{slope}} \end{pmatrix}.$$

Here, the sample sizes are n_1 for Group 1, n_2 for Group 2, $N = n_1 + n_2$, $\mathbf{y}$ is an outcome vector, $\mathbf{X}$ is a design matrix and $\boldsymbol{\beta}$ is a parameter vector (β_{int} intercept: β_{trt} treatment effect: β_{slope} slope). In Section 3.1, the elements of $\boldsymbol{\varepsilon}$ are assumed to be distributed with a unique variance σ_{res}^2 . In Section 3.2, the elements of $\boldsymbol{\varepsilon}$ are assumed to be distributed with variances $\sigma_{1\text{res}}^2$ in Group 1 and $\sigma_{2\text{res}}^2$ in Group 2.

Section 3.1 shows that the actual type I error rate, thus the coverage, of the ANCOVA with equal variances using the standard formula is asymptotically at a nominal level in the case of equal sample sizes. Section 3.2 shows that the actual type I error rate of the ANCOVA with unequal variances is asymptotically at a nominal level even if sample sizes differ.

3.1 ANCOVA with equal variances

We first consider the properties of the ANCOVA with equal variances when used for analysis under random allocation. Because the ordinary least-squares estimators of $\boldsymbol{\beta}$ are obtained from $\hat{\boldsymbol{\beta}}_{\text{ols}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$, the estimators are expressed as follows:

$$\begin{aligned}
\hat{\beta}_{\text{ols_slope}} &= (n_1 \hat{\sigma}_{1\text{cov}} + n_2 \hat{\sigma}_{2\text{cov}}) (n_1 \hat{\sigma}_{1\text{pre}}^2 + n_2 \hat{\sigma}_{2\text{pre}}^2)^{-1}, \\
\hat{\beta}_{\text{ols_trt}} &= \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j2} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j2} - \hat{\beta}_{\text{ols_slope}} \left(\frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j1} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j1} \right), \\
\hat{\beta}_{\text{ols_int}} &= \frac{1}{n_1} \left(\sum_{j=1}^{n_1} y_{1j2} - \hat{\beta}_{\text{ols_slope}} \sum_{j=1}^{n_1} y_{1j1} \right).
\end{aligned} \tag{3}$$

Here, $\hat{\sigma}_{\text{icov}}$ and $\hat{\sigma}_{\text{ipre}}^2$ are as follows:

$$\hat{\sigma}_{\text{icov}} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left[\left(y_{ij1} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij1} \right) \left(y_{ij2} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij2} \right) \right] \text{ and } \hat{\sigma}_{\text{ipre}}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_{ij1} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij1} \right)^2.$$

Yang and Tsiatis (2001) showed that $\hat{\beta}_{\text{ols_trt}}$ is a consistent estimator of the treatment effect, $\beta_{\text{trt}} = \mu_{2\text{post}} - \mu_{1\text{post}}$ and $N^{1/2} (\hat{\beta}_{\text{ols_trt}} - \beta_{\text{trt}})$ converges to a normal distribution with zero mean, for non-normal data with unequal covariances and variances in post-treatment measurements under random allocation. From the work of Yang and Tsiatis (2001), the asymptotic variance of $\hat{\beta}_{\text{ols_trt}}$ is expressed as follows:

$$V_{\text{ols_a}} = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \times \left\{ \left(\frac{n_2 \sigma_{1\text{post}}^2 + n_1 \sigma_{2\text{post}}^2}{n_1 + n_2} \right) - \frac{1}{\sigma_{\text{pre}}^2} \left(\frac{n_1 \sigma_{1\text{cov}} + n_2 \sigma_{2\text{cov}}}{n_1 + n_2} \right) \left(\frac{p \sigma_{1\text{cov}} + q \sigma_{2\text{cov}}}{p + q} \right) \right\}, \tag{4}$$

where $p = 2n_2 - n_1$ and $q = 2n_1 - n_2$. Here and throughout, we neglect $o(1/N)$ terms.

Next we consider the variance estimator of $\hat{\beta}_{\text{ols}}$ based on the standard formula,

$\text{Cov}(\hat{\beta}_{\text{obs}}) = \sigma_{\text{res}}^2 (\mathbf{X}'\mathbf{X})^{-1}$. The variance of $\hat{\beta}_{\text{ols_trt}}$ is expressed as follows:

$$\hat{V}(\hat{\beta}_{\text{ols_trt}}) = \sigma_{\text{res}}^2 \left\{ (n_1 + n_2) \left(\sum_{j=1}^{n_1} y_{1j1}^2 + \sum_{j=1}^{n_2} y_{2j1}^2 \right) - \left(\sum_{j=1}^{n_1} y_{1j1} + \sum_{j=1}^{n_2} y_{2j1} \right)^2 \right\} \times \left\{ n_1 n_2 (n_1 \hat{\sigma}_{1\text{pre}}^2 + n_2 \hat{\sigma}_{2\text{pre}}^2) \right\}^{-1}.$$

Here, the residual variance $\hat{\sigma}_{\text{res}}^2$ is as follows:

$$\begin{aligned}
\hat{\sigma}_{\text{res}}^2 &= (n_1 + n_2)^{-1} \hat{\mathbf{e}}' \hat{\mathbf{e}} \\
&= \left(\frac{n_1 \hat{\sigma}_{1\text{post}}^2 + n_2 \hat{\sigma}_{2\text{post}}^2}{n_1 + n_2} \right) - \left(\frac{n_1 \hat{\sigma}_{1\text{pre}}^2 + n_2 \hat{\sigma}_{2\text{pre}}^2}{n_1 + n_2} \right)^{-1} \left(\frac{n_1 \hat{\sigma}_{1\text{cov}} + n_2 \hat{\sigma}_{2\text{cov}}}{n_1 + n_2} \right)^2,
\end{aligned}$$

where we ignore the adjustment of the degrees of freedom, and $\hat{\sigma}_{ipost}^2$ is as follows:

$$\hat{\sigma}_{ipost}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_{ij2} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij2} \right)^2.$$

Using $\hat{\sigma}_{1pre}^2 \rightarrow \sigma_{pre}^2$ and $\hat{\sigma}_{2pre}^2 \rightarrow \sigma_{pre}^2$, $\hat{V}(\hat{\beta}_{ols_trt})$ is asymptotically expressed as $V_{ols} = (1/n_1 + 1/n_2) \sigma_{res}^2$. This value is explicitly expressed as follows:

$$V_{ols} = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \left\{ \left(\frac{n_1 \sigma_{1post}^2 + n_2 \sigma_{2post}^2}{n_1 + n_2} \right) - \frac{1}{\sigma_{pre}^2} \left(\frac{n_1 \sigma_{1cov} + n_2 \sigma_{2cov}}{n_1 + n_2} \right)^2 \right\}. \quad (5)$$

Therefore, $V_{ols} - V_{ols_a}$ is expressed as follows:

$$V_{ols} - V_{ols_a} = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \times \left\{ \frac{(n_1 - n_2) \sigma_{1post}^2 + (n_2 - n_1) \sigma_{2post}^2}{n_1 + n_2} - \frac{2}{\sigma_{pre}^2} \frac{n_1 \sigma_{1cov} + n_2 \sigma_{2cov}}{n_1 + n_2} \frac{(n_1 - n_2) \sigma_{1cov} + (n_2 - n_1) \sigma_{2cov}}{n_1 + n_2} \right\}.$$

When $\sigma_{1post}^2 = \sigma_{2post}^2$ and $\sigma_{1cov} = \sigma_{2cov}$, V_{ols} is equal to valid V_{ols_a} in (4). Even if $\sigma_{1post}^2 \neq \sigma_{2post}^2$ or $\sigma_{1cov} \neq \sigma_{2cov}$, $V_{ols} = V_{ols_a}$ as long as the samples sizes are equal, $n_1 = n_2$. Therefore, the actual type I error rate, thus the coverage, of the ANCOVA with equal residual variances is asymptotically at a nominal level in the case of equal sample sizes.

3.2 ANCOVA with unequal variances

We now consider the properties of the ANCOVA with unequal variances when used for analysis under random allocation. The generalized least-squares estimators of β_{gls} are obtained by $\beta_{gls} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}$. Here, $\mathbf{V}$ is an $(n_1 + n_2) \times (n_1 + n_2)$ diagonal matrix whose l th diagonal elements are σ_{lres}^2 ($l=1, \dots, n_1$) and σ_{lres}^2 ($l=n_1+1, \dots, n_1+n_2$). The estimators are expressed as follows:

$$\begin{aligned} \hat{\beta}_{gls_slope} &= \left(\frac{n_1}{\sigma_{1res}^2} \hat{\sigma}_{1cov} + \frac{n_2}{\sigma_{2res}^2} \hat{\sigma}_{2cov} \right) \left(\frac{n_1}{\sigma_{1res}^2} \hat{\sigma}_{1pre}^2 + \frac{n_2}{\sigma_{2res}^2} \hat{\sigma}_{2pre}^2 \right)^{-1}, \\ \hat{\beta}_{gls_trt} &= \frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j2} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j2} - \hat{\beta}_{gls_slope} \left(\frac{1}{n_2} \sum_{j=1}^{n_2} y_{2j1} - \frac{1}{n_1} \sum_{j=1}^{n_1} y_{1j1} \right), \\ \hat{\beta}_{gls_int} &= \frac{1}{n_1} \left(\sum_{j=1}^{n_1} y_{1j2} - \hat{\beta}_{gls_slope} \sum_{j=1}^{n_1} y_{1j1} \right). \end{aligned} \quad (6)$$

We consider the asymptotic variance of $\hat{\beta}_{gls_trt}$. The equations of $\hat{\beta}_{ols_trt}$ in (3) and $\hat{\beta}_{gls_trt}$ in (6) are same except the derivations of $\hat{\beta}_{ols_slope}$ in (3) and $\hat{\beta}_{gls_slope}$ in (6). Therefore, by

using the similar manner in Section 3.1, we can show that $\hat{\beta}_{\text{gls_trt}}$ is a consistent estimator of the treatment effect, $\beta_{\text{trt}} = \mu_{2\text{post}} - \mu_{1\text{post}}$, and $N^{1/2} (\hat{\beta}_{\text{gls_trt}} - \beta_{\text{trt}})$ converges to a normal distribution with mean zero. We derived the asymptotic variance of $\hat{\beta}_{\text{gls_trt}}$ as follows:

$$V_{\text{gls_a}} = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \times \left\{ \left(\frac{n_2 \sigma_{1\text{post}}^2 + n_1 \sigma_{2\text{post}}^2}{n_1 + n_2} \right) - \frac{1}{\sigma_{\text{pre}}^2} \left(\frac{n_1^* \sigma_{1\text{cov}} + n_2^* \sigma_{2\text{cov}}}{n_1^* + n_2^*} \right) \left(\frac{p^* \sigma_{1\text{cov}} + q^* \sigma_{2\text{cov}}}{p^* + q^*} \right) \right\}, \quad (7)$$

where $n_1^* = n_1 / \sigma_{1\text{res}}^2$, $n_2^* = n_2 / \sigma_{2\text{res}}^2$, $p^* = 2n_2 - n_1^* (n_1 + n_2) / (n_1^* + n_2^*)$ and $q^* = 2n_1 - n_2^* (n_1 + n_2) / (n_1^* + n_2^*)$. Because $\sigma_{1\text{res}}^2$ and $\sigma_{2\text{res}}^2$ are unknown, we often estimate these variances by likelihood-based methods assuming normality. Using (6) and some algebra, $V_{\text{gls_a}}$ is expressed in terms of the residual variances as follows:

$$V_{\text{gls_a}} = \sigma_{1\text{res}}^2 / n_1 + \sigma_{2\text{res}}^2 / n_2. \quad (8)$$

Next we consider the variance estimator of $\hat{\beta}_{\text{gls}}$ based on the standard formula, $\text{Cov}(\hat{\beta}_{\text{gls}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. The variance of $\hat{\beta}_{\text{gls_trt}}$ is expressed as follows:

$$\begin{aligned} \hat{V}(\hat{\beta}_{\text{gls_trt}}) &= \left\{ \left(\frac{n_1}{\sigma_{1\text{res}}^2} + \frac{n_2}{\sigma_{2\text{res}}^2} \right) \left(\sum_{j=1}^{n_1} \frac{y_{1j1}^2}{\sigma_{1\text{res}}^2} + \sum_{j=1}^{n_2} \frac{y_{2j1}^2}{\sigma_{2\text{res}}^2} \right) - \left(\sum_{j=1}^{n_1} \frac{y_{1j1}}{\sigma_{1\text{res}}^2} + \sum_{j=1}^{n_2} \frac{y_{2j1}}{\sigma_{2\text{res}}^2} \right)^2 \right\} \\ &\times \left\{ \frac{n_1}{\sigma_{1\text{res}}^2} \frac{n_2}{\sigma_{2\text{res}}^2} \left(\frac{n_1}{\sigma_{1\text{res}}^2} \hat{\sigma}_{1\text{pre}}^2 + \frac{n_2}{\sigma_{2\text{res}}^2} \hat{\sigma}_{2\text{pre}}^2 \right) \right\}^{-1}. \end{aligned}$$

Using $\hat{\sigma}_{1\text{pre}}^2 \rightarrow \sigma_{\text{pre}}^2$, $\hat{\sigma}_{2\text{pre}}^2 \rightarrow \sigma_{\text{pre}}^2$ and some algebra, $\hat{V}(\hat{\beta}_{\text{gls_trt}})$ is asymptotically expressed as follows:

$$V_{\text{gls}} = \sigma_{1\text{res}}^2 / n_1 + \sigma_{2\text{res}}^2 / n_2. \quad (9)$$

V_{gls} in (9) is equal to valid $V_{\text{gls_a}}$ in (8) even if sample sizes differ. Thus, the actual type I error rate of the ANCOVA with equal residual variances is asymptotically at a nominal level even under unequal sample sizes.

4 Simulation studies

For continuous pre- and post-treatment measurements in parallel-group, randomized clinical trials, we often use the ANCOVA with equal slopes, the *t*-test on change from pre-treatment measurement or the *t*-test on post-treatment measurement to compare treatment groups. Hence, we compare actual type I error rates and powers of these three tests with a two-sided

type I error rate of 5% in simulated experiments. We examined these tests with equal and unequal residual variances. We simulated 100 000 data sets by the following equation:

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix} \sim MVN \left\{ \begin{pmatrix} \mu_{\text{pre}} \\ \mu_{\text{post}} \end{pmatrix}, \begin{pmatrix} \sigma_{\text{pre}}^2 & \sigma_{\text{cov}} \\ \sigma_{\text{cov}} & \sigma_{\text{post}}^2 \end{pmatrix} \right\},$$

where Y_{ijk} is the measurement of the k th ($k = 1, 2$) time for the j th ($j = 1, \dots, n_i$) subject in the i th ($i = 1, 2$) treatment group.

We assume the pair (Y_{ij1}, Y_{ij2}) is normally distributed with the following covariance matrix:

$$\begin{pmatrix} 25 & 15 \\ 15 & 59 \end{pmatrix} \text{ for Group 1 } (i=1) \text{ and } \begin{pmatrix} 25 & 23 \\ 23 & 30 \end{pmatrix} \text{ for Group 2 } (i=2).$$

For the parameter setting, we refer to the estimates of the blood lead data in Section 2. Note that the means and variances of pre-treatment measurements are equal between the groups. In this parameter setting, the residual variances of Group 1 and Group 2 are about 50 and 10 in the ANCOVA, 54 and 9 in the t -test on change and 59 and 30 in the t -test on post-treatment measurement. In each method, Group 1 has a larger variance. The slopes in the ANCOVA are $15/25 = 0.6$ and $23/25 \approx 0.9$ respectively, and the correlation coefficients

between pre- and post-treatment measurements are $15 / (\sqrt{25} \sqrt{59}) \approx 0.4$ and $23 / (\sqrt{25} \sqrt{30}) \approx 0.8$ respectively. The numbers in each group are set to be (75, 75), (100, 50) and (50, 100). The first is the case with equal sample sizes, the second is the case in which a group with a large sample size has a large variance and the last is the case in which a group with a large sample size has a small variance. Similarly, the numbers are also set to be (150, 150), (200, 100) and (100, 200). In each test with unequal variances, we use the Satterthwaite approximation for the degrees of freedom (Satterthwaite, 1946). To obtain actual type I error rates, we calculate the proportions of data sets detected with a significant difference under equal μ_{post} ; the means of the post-treatment are equal between the groups. Note that the proportions that the true values are included in the 95% confidence intervals (coverage proportion) are expressed by the values; 100% minus the actual type I error rates (%) under $\mu_{2\text{post}} - \mu_{1\text{post}}$. Similarly, to obtain actual powers, we calculate the proportions under $\mu_{2\text{post}} - \mu_{1\text{post}} = 1.75$; the difference in the means of the post-treatment value is 1.75. In this simulation study, the ANCOVA and t -test are performed by proc mixed and proc ttest of SAS 9.2, respectively. The programme codes of the ANCOVA are shown in the Appendix.

The results are shown in Table 1. The actual type I error rates of the all methods with equal residual variances are close to the nominal level only in cases of equal sample sizes. The actual rates are less than 5% when the group with a larger variance has a larger sample size, and these are more than 5% when the group with a larger variance has a smaller sample size. The actual type I error rates of the all methods with unequal variances are close to the nominal level, even in cases of unequal sample sizes. In each case, the powers of the

ANCOVA are biggest amongst the three methods. In each method, the type I error rates and powers are almost the same between equal and unequal variances for the case of equal sample sizes. In the ANCOVA and t -test on change with unequal variances, the powers are biggest for the case in which a group with a larger variance has a larger sample size. In the t -test on post-measurements with unequal variances, the powers are almost the same between the equal sample size situation and in the larger sample size group with a larger variance. These results do not depend on the total sample sizes; 150 and 300.

Moreover, we performed the simulation studies in the setting with truncated data. In the protocol of succimer, the inclusion criterion of the lead concentration is from 20 to 44 g/dL. Thus, we simulated data sets by truncated bivariate normal distributions. The variances, covariances and the difference of the means in the post-treatment measurement are set to be same as the former simulation setting. See Tallis (1961), Lee (1979) and Leppard and Tallis (1989) for the detail of the method to generate data sets following a truncated multivariate normal distribution. The percentage truncated by the lower limit 20 g/dL and upper limit 44 g/dL are 51% and 0.3%, respectively. The results are shown in Table 2. The results are almost same as those of Table 1.

5 Discussion

In this paper, we show the following two asymptotic properties of the ANCOVA with equal slopes when applied to non-normal data with equal means and variances in pre-treatment measurements and unequal covariances and variances in post-treatment measurements. The actual type I error rate of the ANCOVA with equal variances using the standard formula is asymptotically at a nominal level only in the case of equal sample sizes. The actual type I error rate of the ANCOVA with unequal variances is asymptotically at a nominal level even in the case of unequal sample sizes. In conclusion, the ANCOVA with equal slopes can be asymptotically justified under random allocation. In case of unequal sample sizes, an assumption of unequal residual variances is important.

If we consider a class of tests to compare the means of $y_{ij2} - \lambda y_{ij1}$, the tests of the treatment effect in the ANCOVA, t -test on change ($\lambda = 1$) and t -test on post-measurement ($\lambda = 0$) fall within it. The variances in $y_{ij2} - \lambda y_{ij1}$, v_i are $\lambda^2 \sigma_{i\text{pre}}^2 + \sigma_{i\text{post}}^2 - 2\lambda \sigma_{i\text{cov}}$ in each group. Under the equal variances assumption, the variance of the treatment effect, $(1/n_1 + 1/n_2)(n_1 v_1 + n_2 v_2)/(n_1 + n_2)$, takes a minimum value at $\lambda = (n_1 \sigma_{1\text{cov}} + n_2 \sigma_{2\text{cov}})(n_1 \sigma_{1\text{pre}}^2 + n_2 \sigma_{2\text{pre}}^2)^{-1}$, which is the slope derived from the ANCOVA with equal variances. Thus, the ANCOVA is the best test amongst the class, in view of power. However, under the unequal variances assumption, the variance of the treatment effect, $v_1/n_1 + v_2/n_2$, takes a minimum value at $\lambda = (n_2 \sigma_{1\text{cov}} + n_1 \sigma_{2\text{cov}})(n_2 \sigma_{1\text{pre}}^2 + n_1 \sigma_{2\text{pre}}^2)^{-1}$, which differs from the slopes derived from the ANCOVA with unequal variances. Thus, the ANCOVA with unequal variances will not be the best test in this class, in terms of power.

The difference of the asymptotic variances between the ANCOVA with equal and unequal residual variances, $V_{\text{ols_a}}$ in (4) and $V_{\text{gls_a}}$ in (7) is caused by the difference of weights in the weighted average for $\sigma_{1\text{cov}}$ and $\sigma_{2\text{cov}}$. Therefore, if the difference of the covariances is

small, the difference of V_{ols_a} and V_{gls_a} is small. When σ_{1cov} is equal to σ_{2cov} , V_{ols_a} and V_{gls_a} are same.

Unequal allocation of patients to treatments is applied in clinical trials because of ethical or economical reasons (Gail *et al.*, 1976; Nam, 1973; Pocock, 1983). The following properties are well known in the t -test with equal variances, the Student's t -test (Algina, 2005). In the case of unequal sample sizes and unequal variances, the actual type I error rate is not at a nominal level. The actual rate is over a nominal level in the case in which a group with a large sample size has a smaller variance, and that is under a nominal level in the case in which a group with a large sample size has a larger variance. These properties are applied to the ANCOVA with equal variances. The actual rates of the ANCOVA with equal variances in each situation of simulation studies were 12% and 1.5%, respectively, compared to the nominal rate of 5%. The false actual type I error rate can be expressed asymptotically as follows:

$$\alpha_{false} = 2 \times \left\{ 1 - \Phi^{-1} \left(z_{1-\alpha/2} \times \frac{\sqrt{\sigma_{res}^2(1/n_1 + 1/n_2)}}{\sqrt{\sigma_{1res}^2/n_1 + \sigma_{2res}^2/n_2}} \right) \right\},$$

where Φ is a cumulative standard normal distribution function, $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th quantile from the standard normal distribution and α is a nominal two-sided type I error rate. The values in the simulation study were almost the same as the values derived from this equation. Because large discrepancies from the nominal rate could occur, the ANCOVA with equal variances should not be used when the sample sizes and residual variances are both unequal between the groups.

Unequal covariances and variances of post-treatment measurements indicate unequal slopes and, usually, unequal residual variances in the ANCOVA. Because unequal slopes mean that the sizes of the treatment effect depend on the pre-treatment measurements, the ANCOVA with unequal slopes using several contrasts to compare treatment effects at given values of the pre-treatment measurements (Hendrix *et al.*, 1982) is used as explanatory analysis. Especially, the treatment effect is often estimated at the observed mean of pre-treatment measurements. However, the variances of the estimator based on the standard formula is underestimated because it is not corrected for estimating the unknown pre-treatment expectation (Chen, 2006; Winkens, *et al.*, 2007). Thus, the actual type I error rate is over a nominal level. When the average treatment effect is a primary concern, the ANCOVA with equal slopes is easier to interpret than the ANCOVA with unequal slopes. The ANCOVA with equal slopes is often used as a primary analysis.

There have been some controversies concerning a null hypothesis. The variances are same if treatments are completely identical. Therefore some researchers consider the variances should be same under a null hypothesis (Fisher, 1939; Brownie *et al.*, 1990). In our simulation study, we calculated the type I error rates under a null hypothesis: equal means and unequal variances. However, the coverage probability (%) is expressed as $100 - \text{the actual type I error rate}$. The simulation result is not limited under a null hypothesis, but it can be interpreted as the coverage probability.

Acknowledgments

We would like to thank the associate editor and the anonymous referees for their valuable comments, which have greatly improved the manuscript, and Dr. Lynne Berry and Yvonne Poindexter for their editorial suggestions.

References

- Algina, J. Aspin-Welch Test. In: Armitage, P., Colton, T., editors. Encyclopedia of Biostatistics. 2. Wiley; Chichester: 2005. p. 233-235.
- Blomqvist N. On the relation between change and initial value. Journal of American Statistical Association. 1977; 72:746-749.
- Brownie C, Boos DD, Hughes-Oliver J. Modifying the t and ANOVA F tests when treatment is expected to increase variability relative to controls. Biometrics. 1990; 46:259-266.
- Chen X. The adjustment of random baseline measurements in treatment effect estimation. Journal of Statistical Planning and Inference. 2006; 136:4161-4175.
- Committee for Proprietary Medicinal Products (CPMP). Points to consider on adjustment for baseline covariates. Statistics in Medicine. 2004; 23:701-709. [PubMed: 14981670]
- Crager MR. Analysis of covariance in parallel-group clinical trials with pretreatment baselines. Biometrics. 1987; 43:895-901. [PubMed: 3427174]
- Feldt LS. A comparison of the precision of three experimental designs employing a concomitant variable. Psychometrika. 1958; 23:335-353.
- Fisher RA. The comparison of samples with possibly unequal variances. Annals of Eutgenics. 1939; 9:174-180.
- Fitzmaurice, GM., Laird, NM., Ware, JH. Applied Longitudinal Analysis. Wiley; New York: 2004.
- Follmann DA. The effect of screening on some pretest-posttest test variances. Biometrics. 1991; 47:763-771. [PubMed: 1912270]
- Gail MWR, Byar DP, Brown C. How many controls? Journal of Chronic Disease. 1976; 29:723-731.
- Hendrix LJ, Carter MW, Scott DT. Covariance analyses with heterogeneity of slopes in fixed models. Biometrics. 1982; 38:641-650. [PubMed: 6756494]
- Lee LF. On the first and second moments of the truncated multi-normal distribution and a simple estimator. Economics Letters. 1979; 3:165-169.
- Leppard P, Tallis GM. Evaluation of the mean and covariance of the truncated multinormal. Applied Statistics. 1989; 38:543-553.
- Nam JM. Optimum sample sizes for the comparison of the control and the treatment. Biometrics. 1973; 29:101-108. [PubMed: 4691048]
- Pocock, S. Clinical Trials: A Practical Approach. John Wiley & Sons; Chichester: 1983.
- Rogan WJ, Dietrich KN, Ware JH, Dockery DW, Salganik M, Radcliffe J, Jones RL, Ragan NB, Chisolm JJ Jr, Rhoads GG. Treatment of Lead-Exposed Children Trial Group. The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. New England Journal of Medicine. 2001; 344:1421-1426. [PubMed: 11346806]
- Satterthwaite FW. An approximate distribution of estimates of variance components. Biometrics Bulletin. 1946; 2:110-114. [PubMed: 20287815]
- Tallis GM. The moment generating function of the truncated multinormal distribution. Journal of the Royal Statistical Society, Series B. 1961; 2:223-229.
- Winkens B, van Breukelen GJ, Schouten HJ, Berger MP. Randomized clinical trials with a pre- and a post-treatment measurement: repeated measures versus ANCOVA models. Contemporary Clinical Trials. 2007; 28:713-719. [PubMed: 17524958]
- Yang L, Tsiatis AA. Efficiency study of estimators for a treatment effect in a pretest-posttest trial. American Statistician. 2001; 55:314-321.

Appendix

We provide the SAS codes of the Mixed procedure to obtain estimates of the ANCOVA with equal slopes in Section 4. In the ANCOVA with equal residual variances, the estimator and the variance based on ordinary least-squares are the same as those based on restricted maximum likelihood.

```
/* equal variances */
proc mixed method = reml;
  class group;
  model post = pre group;
run;
/* unequal variances */
proc mixed method = reml;
  class group;
  model post = pre group / ddfm = satterth;
  repeated / group=group;
run;
```

Figure 1.
Blood lead levels at baseline and one week after administration of succimer and placebo.
The open and closed circles show the succimer and placebo groups, respectively.

Table 1

The actual type I error rates and powers in simulation studies of the ANCOVA with equal slopes, t-test on change from pre-treatment measurements, and t-test on post-treatment measurements for the data with unequal covariances and unequal variances of post-treatment measurements.

	sample size* total (n1, n2)	equal variance assumed			unequal variance assumed		
		ANCOVA with equal slopes	t-test on change	t-test on post	ANCOVA with equal slopes	t-test on change	t-test on post
type I error rate (%)	150	(75, 75)	5.14	5.15	5.02	5.07	5.06
		(100, 50)	1.53	1.36	3.02	5.06	5.13
		(50, 100)	12.20	12.56	7.73	5.01	5.00
	300	(150, 150)	5.00	5.09	5.07	5.00	5.05
		(200, 100)	1.45	1.28	2.94	5.10	5.06
		(100, 200)	12.23	12.59	7.85	4.98	5.03
power (%)	150	(75, 75)	48.8	47.4	35.7	48.3	47.1
		(100, 50)	34.7	32.6	27.4	54.5	53.3
		(50, 100)	53.9	53.0	37.1	36.5	35.6
	300	(150, 150)	78.4	76.6	62.0	77.9	76.5
		(200, 100)	68.3	65.7	52.9	84.1	82.9
		(100, 200)	78.4	77.1	60.7	63.1	61.9

*
n1:sample size of the group with a larger variance
n2:sample size of the group with a smaller variance

Table 2

The actual type I error rates and powers in simulation studies of the ANCOVA with equal slopes, t-test on change from pre-treatment measurements, and t-test on post-treatment measurements for the truncated data.

	sample size* total (n1, n2)	equal variance assumed			unequal variance assumed		
		ANCOVA with equal slopes	t-test on change	t-test on post	ANCOVA with equal slopes	t-test on change	t-test on post
type I error rate (%)	(75, 75)	4.94	5.07	5.03	4.91	4.98	5.01
	(100, 50)	1.46	1.31	2.92	5.06	5.08	4.92
	(50, 100)	12.24	12.50	7.82	4.96	4.99	5.01
300	(150, 150)	5.11	5.02	5.16	4.98	4.97	5.16
	(200, 100)	1.48	1.30	2.86	5.03	5.01	4.99
	(100, 200)	12.18	12.43	7.82	5.04	5.03	4.95
150	(75, 75)	49.2	48.0	35.8	48.7	47.7	35.8
	(100, 50)	34.7	32.8	27.6	54.6	53.7	35.4
	(50, 100)	53.8	53.2	37.3	36.8	35.9	29.8
power (%)	(150, 150)	78.3	76.8	62.1	78.0	76.6	62.1
	(200, 100)	68.3	65.9	52.9	83.9	82.9	61.9
	(100, 200)	78.6	77.5	61.0	63.7	62.3	52.7

* n1:sample size of the group with a larger variance
n2:sample size of the group with a smaller variance