

# Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods

Nathaniel S. O'Connell, Lin Dai, Yunyun Jiang, Jaime L. Speiser, Ralph Ward, Wei Wei, Rachel Carroll and Mulugeta Gebregziabher\*

Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

## Abstract

Often repeated measures data are summarized into pre-post-treatment measurements. Various methods exist in the literature for estimating and testing treatment effect, including ANOVA, analysis of covariance (ANCOVA), and linear mixed modeling (LMM). Under the first two methods, outcomes can either be modeled as the post treatment measurement (ANOVA-POST or ANCOVA-POST), or a change score between pre and post measurements (ANOVA-CHANGE, ANCOVA-CHANGE). In LMM, the outcome is modeled as a vector of responses with or without Kenward-Rogers adjustment. We consider five methods common in the literature, and discuss them in terms of supporting simulations and theoretical derivations of variance. Consistent with existing literature, our results demonstrate that each method leads to unbiased treatment effect estimates, and based on precision of estimates, 95% coverage probability, and power, ANCOVA modeling of either change scores or post-treatment score as the outcome, prove to be the most effective. We further demonstrate each method in terms of a real data example to exemplify comparisons in real clinical context.

**Keywords:** Analysis of variance; Analysis of covariance; Linear mixed model; Pre-post; Rrandomized trial; Repeated measures

**Abbreviations:** ANCOVA: Analysis of covariance; ANCOVA-CHANGE: Analysis of covariance using change score as the outcome; ANCOVA-POST: Analysis of covariance using post-treatment as the outcome; ANOVA: Analysis of variance; ANOVA-CHANGE: Analysis of variance using change score as the outcome; ANOVA-POST: Analysis of variance using post-treatment as the outcome; CI: Confidence interval; HCS: Heterogeneous compound symmetry; KR: Kenward-Rogers; LMM: Linear mixed model; MAR: Missing at random; MCAR: Missing completely at random; MNAR: Missing not at random; REML: Restricted maximum likelihood; SD Standard deviation.

## Introduction

In clinical research, it is common to record repeated measurements for subject responses across multiple occasions. There exist a variety of analysis methods, including repeated measures analysis of variance (RANOVA), multivariate ANOVA (MANOVA) and linear mixed modeling (LMM). Researchers often simplify repeated measures data by using summary data to quantify pre-and post-treatment outcomes, allowing for a more intuitive and easier interpretation of treatment comparisons [1]. When appropriately applied, simplifying repeated measure outcomes to only two time measurements in the design phase improves efficiency and cost effectiveness, particularly in circumstances when responses are expensive to measure. In many instances this is done because the response of patients at a certain final time point is more clinically relevant than trends over time. Rather than comparing trends over time within each treatment group, the pre-post treatment summary method also simplifies data analysis to standard t-test procedures.

Decades of literature exists exploring and comparing methods for pre-post analysis, in both theory and application. The goal of this paper is not in developing new methods of analysis, but to review and succinctly tie together existing literature into a cohesive comparison of common methods often discussed and employed. We revisit and review the basic methods of pre-post data analysis discussed in the literature, and then exemplify the results through simulation and real data examples to corroborate existing knowledge. The rest of the paper

is structured as follows: we first provide a review of key literature in pre-post analysis. We then outline the models to be compared and set up a simple simulation study to demonstrate the comparison of methods, and discuss simulated results in conjunction with the theoretical expectations of variance and related implicit measures. A real data example is used to exemplify the difference in methods in practice, and highlight the importance of a proper analysis method. We finish with a discussion of results and further present ideas for future avenues of research in the area of pre-post data analysis.

## Review of Literature

Frison and Pocock [2] discuss three methods for analyzing data from pre-post designs: a) ANOVA with the post measurement as the response variable (ANOVA-POST), b) ANOVA with the change from pre-treatment to post-treatment as the response variable (ANOVA-CHANGE), and c) ANCOVA with the post measurement as the response variable (ANCOVA-POST), adjusting for the pre-treatment measurement. Brogan and Kutner [3] compare the use of ANOVA-CHANGE with RANOVA. However, Huck and McLean [4] criticize the latter method due to its frequent misinterpretation in practice. Furthermore, they note the F-test in an RANOVA interaction is equivalent to the F-test in change score analysis. RANOVA provides the same conclusion as ANOVA-CHANGE, but use of ANOVA-CHANGE is simpler and more accurately interpreted compared to RANOVA. These conclusions are defended by Jennings [5], who asserts RANOVA is not recommended for pre-post analysis given the simpler alternatives presented.

\*Corresponding author: Gebregziabher M, Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon St, MSC 835, Charleston, SC 29425, USA, Tel: 843-876-1112, Fax: 843-876-1126; E-mail: [gebregz@musc.edu](mailto:gebregz@musc.edu)

Received January 23, 2017; Accepted February 21, 2017; Published February 24, 2017

Citation: O'Connell NS, Dai L, Jiang Y, Speiser JL, Ward R, et al. (2017) Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods. J Biom Biostat 8: 334. doi:10.4172/2155-6180.1000334

Copyright: © 2017 O'Connell NS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Among the methods, ANCOVA-POST is generally regarded as the preferred approach, given that it typically leads to unbiased treatment effect estimate with the lowest variance relative to ANOVA-POST or ANOVA-CHANGE [1-6]. However, ANCOVA has been criticized as being biased in the case of unequal pre-treatment mean measurements between groups [7,8]. This conundrum, known as "Lord's Paradox," was first documented in 1967 Lord [9], and has been discussed in the literature extensively. Among a detailed examination of various methods of repeated measures data analysis for pre-post outcomes, Liang and Zeger [10] note in the simple case with only two responses (i.e. pre- and post-treatment measurements), ANCOVA-POST produces an unbiased estimate only in the case of equal pre-treatment measurements, whereas ANOVA-CHANGE leads to unbiased estimates that are only slightly less efficient than ANCOVA-POST. Senn [11] discusses these criticisms at length, providing various conditions for which these claims do not hold, ultimately concluding ANCOVA should be used with caution in the case of unequal pre-treatment measurements, but ANOVA-CHANGE is not impervious to bias either.

A recent simulation study by Egbewale, Lewis, and Sim [12] over varying degrees of pre-treatment imbalance and pre-post treatment correlations, demonstrates that a comparison of the methods is not straightforward in the presence of unequal pre-treatment measures between groups. They recommend ANCOVA when pre-treatment measurements are equal in expectation across groups, as should be the case in properly designed randomized trials [12]. ANOVA-POST has a larger variance because it allows for possible random baseline imbalance for which it cannot adjust. ANCOVA allows adjustment for baseline differences and thus has a smaller variance than ANOVA. In further support for ANCOVA, Vickers and Altman [13] note that ANCOVA achieves the greatest power relative to ANOVA-CHANGE or ANOVA-POST, but the power of ANOVA-CHANGE approaches ANCOVA as correlation between pre-post measures approaches one.

Combining analysis of change scores with adjustments for pre-treatment measures, Laird [14] offers a slight modification to ANCOVA, in which the change score is incorporated as the outcome and pre-treatment measures as covariate. Compared with traditional ANCOVA, this ANCOVA-CHANGE leads to equal results in terms of variance of treatment effect, although Laird [14] asserts the latter method allows one to assess whether change occurred in individual treatment groups. Despite this possible advantage, this appears less frequently used or discussed in the literature.

In the rest of the paper, we will discuss and compare results between the five common methods ANOVA and ANCOVA modeling both the post-treatment response only and the change score, along with a linear mixed model (LMM) modeling the pre-post treatment response vector,  $Y_{ij}$ . We use simulations over a range of sample sizes and pre-post measurement correlations to corroborate the comparison of methods with the existing literature and theoretical expectations of variance.

## Methods

To set up the modeling framework, let  $Y_i$  be the continuous response variable from a randomized trial, for  $i=1, \dots, n$  patient responses from samples  $n_1$  and  $n_2$  from each treatment group. Let the group assignment be designated by the indicator,  $X_p$ , such that for the  $i^{th}$  patient,  $X_i=1$  for the active treatment and  $X_i=0$  in the control/placebo group. Responses for each treatment are each sampled from a Gaussian distribution with mean  $\mu_x$  and variance  $\sigma^2$ , where  $\mu_x = \beta_0 + \beta_1 X_i$ . To distinguish between

post-treatment and change score measures as outcomes, let  $Y_i^{[p]}$  represent the post treatment response and  $Y_i^{[c]}$  represent the change from pre-treatment to post-treatment measurements. In the case of ANCOVA, let  $Y_{0i}$  be the pre-treatment measurement for which the model is adjusted for. Lastly, let  $\epsilon_i$  represent the random error terms for each of the models. Maximum likelihood is used to estimate the parameters corresponding to each model except for the variance in LMM which are estimated using restricted maximum likelihood (REML). Derivations of variance for each of the estimators from the different methods are given in the Appendix.

### Method 1: ANOVA-POST

Method 1 uses linear regression to compare treatment effects. Formally, the model is as follows:

$$Y_i^{[p]} = \beta_0^{[p]} + \beta_1^{[p]} X_i + \epsilon_i^{[p]}.$$

It is assumed that  $\epsilon_i$  are independently and identically normally distributed with mean 0 and variance  $\sigma^2$ .  $\beta_1^{[p]}$  is interpreted as the difference in the post-treatment mean between treatment groups. The variance of the estimated treatment effect is given by:

$$\text{var}(\hat{\beta}_1^{[p]}) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}.$$

### Method 2: ANOVA-CHANGE

Similar to ANOVA-POST, ANOVA-CHANGE employs a simple ANOVA framework, but instead models the outcome,  $Y_i^{[c]}$  without adjustment for pre-treatment values. Formally, the model with  $\epsilon_i$  assumed to be independently and identically normally distributed with mean 0 and variance  $\sigma^2$  is given as follows:

$$Y_i^{[c]} = \beta_0^{[c]} + \beta_1^{[c]} X_i + \epsilon_i^{[c]}.$$

Here  $\beta_1^{[c]}$  is interpreted as the difference in the change score mean of the treatment groups. Under an unstructured covariance structure assumption, the variance of  $\hat{\beta}_1^{[c]}$  is given by:

$$\text{var}(\hat{\beta}_1^{[c]}) = (\sigma_{pre}^2 + \sigma_{post}^2 - 2\rho\sigma_{pre}\sigma_{post}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right).$$

Under a compound symmetry assumption, where pre- and post-treatment variance is assumed to be equal, the variance is given by:

$$\text{var}(\hat{\beta}_1^{[c]}) = 2(1 - \rho) \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \sigma^2.$$

### Method 3: ANCOVA-CHANGE

Method 3 employs an ANCOVA model to analyze the change score as an outcome, adjusting for the pre-treatment values. Essentially, ANCOVA-CHANGE is equivalent to ANOVA-CHANGE, with an added adjustment for the pre-treatment measurement for every patient. Formally, the model is as follows:

$$Y_i^{[c]} = \beta_0^{[c2]} + \beta_1^{[c2]} X_i + \beta_2^{[c2]} Y_{0i} + \epsilon_i^{[c2]}.$$

It is assumed that  $\epsilon_i$  are independently and identically normally distributed with mean 0 and variance  $\sigma^2$ .  $\beta_1^{[c2]}$  is interpreted as the difference in the change score mean of the treatment groups, given the pre-treatment measurement and the variance of its estimator is given by

$$\text{var}(\hat{\beta}_1^{[c2]}) = \frac{n_1 + n_2 - 2}{n_1 + n_2 - 3} \sigma_{post}^2 (1 - \rho^2) \left[ \frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{Y}_{T0} - \bar{Y}_{P0})^2}{(n_1 + n_2 - 2)\sigma_{pre}^2} \right]$$

Which, as the sample size increases, simplifies to:

$$\text{var}(\beta^{[1]}) = (1 - \rho^2) \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \sigma_{post}^2.$$

#### Method 4: ANCOVA-POST

Method 4 employs an ANCOVA model to analyze the post-treatment measurements as the outcome, adjusting for the pre-treatment values. In the context of previous methods, ANCOVA-POST is essentially ANOVA-POST (method 1) with pre-treatment measurement included as a covariate. Formally, the model is as follows:

$$Y_i = \beta_0^{[a]} + \beta_1^{[a]} X_i + \beta_2^{[a]} Y_{0i} + \varepsilon_i^{[a]}.$$

It is assumed that  $\varepsilon_i$  are independently and identically normally distributed with mean 0 and variance  $\sigma^2$ .  $\beta_1^{[a]}$  is interpreted as the difference in the post treatment score mean of the treatment groups, given the pre-treatment measurement. Since this method is equivalent to method 3 Laird [14], results including the variance of the estimated treatment effect  $\beta_1^{[a]}$  is the same.

#### Method 5: LMM

Method 5 consists of employing a linear mixed model (LMM) to analyze a vector of the pre-and post-measurements as the outcome.  $Y_{ij}$  denotes the  $j^{\text{th}}$  measure of the  $i^{\text{th}}$  subject. Formally, the model is as follows:

$$Y_{ij} = \beta_0^{[b]} + \beta_1^{[b]} X_i + \beta_2^{[b]} t_{ij} + \beta_3^{[b]} t_{ij} X_i + \varepsilon_{ij}^{[b]},$$

Where  $t_{ij}$  is an indicator for pre-treatment measurement (coded 0) or post-treatment measurement (coded 1). In LMM, it is assumed that  $\varepsilon_{ij}$  are bivariate normally distributed with means 0 and heterogeneous compound symmetric (HCS) covariance matrix and correlation,  $\rho$ . The term  $\beta_1^{[b]} + \beta_3^{[b]}$  is interpreted as the mean difference between treatment groups post-treatment, and  $\beta_1^{[b]}$  is the mean difference between treatment groups pre-treatment. LMM allows for pre-treatment mean differences between the groups.

The variance of  $\hat{\beta}_1 + \hat{\beta}_3$  under HCS covariance matrix for the error term is given by:

$$\text{Var}(\hat{\beta}_1 + \hat{\beta}_3) = \text{Var}(\hat{\beta}_1) = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \hat{\sigma}_{post}^2 = \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \sigma^2.$$

If assuming compound symmetry (CS) is assumed, the variance is:

$$\text{Var}(\hat{\beta}_1 + \hat{\beta}_3) = (1 - \rho) \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \sigma^2$$

The LMM was evaluated under REML estimation in PROC MIXED (SAS 9.3, SAS Institute Inc, Cary, NC). Acknowledging that many degree of freedom adjustments may be employed in mixed models, we choose to evaluate this approach with the conservative, and widely used Kenward and Rogers (KR) adjustment, as well as an unadjusted model. The KR adjustment Kenward and Roger [15]; Schaalje, McBride and Fellingham [16]; Senn [17] appropriately inflates the variance-covariance matrix, along with an adjusted degree of freedom estimate (KR degree of freedom adjustment) when making inference on fixed effects which rely on asymptotic distributions that can lead to biased variance estimates when sample sizes are small. No adjustment, according to Senn [17], leads to negligible difference with the first scenario as sample size grows reasonably large (e.g.  $n > 40$ ).

#### Simulation Study

Data are simulated using SAS 9.3. Simulations are designed to represent a variety of situations which are plausible in pre-post

studies. Using 1000 repetitions, we compare models under three sample sizes ( $n=50, 100, \text{ and } 200$ ), under three pre-post correlations ( $\rho=0.1, 0.5, \text{ and } 0.8$ ), and six  $\beta_1$  coefficients for treatment effect ( $\beta_1=-1.5, -0.1, -1.0, 0.1, 1.0, \text{ and } 1.5$ ). Covariates are generated assuming  $X \sim \text{uniform}(0,1)$ , and  $Y_0 \sim N(0,1)$ ; the post treatment response  $Y_1$  is generated using:  $Y_1 = 10 + 1.5 \cdot \{X \geq 0.5\} + 1.5 \cdot Y_0 + \varepsilon$ , such that  $\{X \geq 0.5\}$  represents treatment 1 (i.e.  $X_i=1$ ) and  $\{X < 0.5\}$  represents control/placebo (i.e.  $X_i=0$ ). To generate correlation between pre- ( $Y_0$ ) and post- ( $Y_1$ ) treatment measures, we use the relationship between correlation and slope:  $\rho = \beta \frac{\sigma_{y0}}{\sigma_{y1}}$ , where  $\sigma_{y0}$  and  $\sigma_{y1}$  are the standard deviations for pre- and post-treatment responses, respectively.  $\beta$  is fixed at 1.5 and  $\sigma_{y1}$  is calculated for each combination of  $\sigma_{y0}$  and  $\rho$ . Random errors are generated such that  $\varepsilon \sim N(0, \sigma^2)$ . The corresponding residual variance is calculated using the relationship between  $\sigma_{y1}$  and the variance of  $\varepsilon$  for different  $\beta_1$  coefficients:  $\sigma_\varepsilon^2 = \sigma_{y1}^2 - \beta_1^2 \times 0.25 - 2.25 \times \sigma_{y0}^2$ . There are a total of 108 simulated scenarios among the combinations of  $n, \rho$ , and  $\beta_1$ . Estimates for the parameter ( $\beta_1$ ), its variance, bias, power, and nominal 95% coverage probability are computed for each simulation scenario, and the results are compared across the five methods. Code used to implement this may be found in supplementary material available online.

#### Simulation Results

In this section, findings from the simulations are discussed and compared to expected theoretical results. Treatment effect parameter estimates and associated standard deviations are reported in Table 1, while bias and power are presented in Tables 2 and 3 shows the 95% coverage probability. The results, consistent with theoretical expectations, show that all methods produce equally unbiased estimates of the treatment effect across equivalent combinations of  $\rho, \beta_1$ , and  $n$ , with accuracy of the estimates improving with sample size. In general, as the number of observations increases, the bias converges to zero for each of the methods across all of the simulated scenarios. When  $\rho$  increases, the bias decreases, regardless of sample size or true  $\beta_1$ .

Since all methods result in unbiased estimates for the treatment effect, we use variance and other implicit measures of the estimates to compare the five methods (Table 1 and Figure 1). In general, when the correlation between pre-treatment and post-treatment values is high, the variance of the estimates are relatively small, regardless of method and the value of  $\beta_1$ . However, for higher correlation values, differences in variability between methods become more apparent when the sample size and value of  $\beta_1$  are fixed. For example, in the scenario when  $\beta_1=0.1$  and  $n=200$ , the difference between the largest and smallest standard deviations among the 5 methods with low correlation ( $\rho=0.1$ ), is 0.006. However, under the same scenario with high correlation ( $\rho=0.8$ ), the difference between the largest and smallest standard deviations is 0.109. While the individual standard deviation estimates are greater in low correlation scenarios, the difference between estimates of differing methods are much more pronounced in scenarios with high correlation, i.e. the greatest variability in  $\beta_1$  occurs when  $\rho$  is close to 0 (i.e.  $\rho=0.1$ ). Furthermore, as sample size increases, variability decreases as expected theoretically. To summarize in terms of correlation and sample size, variability of  $\beta_1$  estimates are greatest when  $n=50$  and  $\rho=0.1$ , and they are lowest when  $n=200$  and  $\rho=0.8$ .

Comparing the two ANOVA methods specifically, ANOVA-CHANGE produces approximately equal or less variability compared to ANOVA-POST, the difference of which increases as  $\rho$  approaches one. Intuitively, it follows that ignoring pre-treatment observations in ANOVA-POST causes a loss of information which leads to an increase

Simulated $\beta_1$ Values across methods		$\rho=0.1$			$\rho=0.5$			$\rho=0.8$		
		$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$
0.1	ANOVA-POST	0.095(4.240)	0.013(3.065)	0.206(2.121)	0.114(0.856)	0.117(0.601)	0.110(0.447)	0.110(0.532)	0.096(0.386)	0.102(0.268)
	ANOVA-CHANGE	0.095(4.220)	0.018(3.055)	0.198(2.115)	0.107(0.768)	0.110(0.537)	0.105(0.396)	0.104(0.343)	0.096(0.246)	0.101(0.177)
	ANCOVA-CHANGE	0.110(4.261)	0.022(3.075)	0.201(2.119)	0.106(0.767)	0.106(0.535)	0.102(0.387)	0.101(0.312)	0.095(0.222)	0.100(0.159)
	ANCOVA-POST	0.110(4.261)	0.022(3.075)	0.201(2.119)	0.106(0.767)	0.106(0.535)	0.102(0.387)	0.101(0.312)	0.095(0.222)	0.100(0.159)
	LMM	0.095(4.240)	0.013(3.065)	0.206(2.121)	0.114(0.856)	0.117(0.601)	0.110(0.447)	0.110(0.532)	0.096(0.386)	0.102(0.268)
1.0	ANOVA-POST	1.104(4.424)	0.997(3.197)	0.912(2.075)	1.038(0.833)	1.007(0.598)	1.006(0.424)	1.028(0.504)	0.992(0.351)	1.006(0.256)
	ANOVA-CHANGE	1.105(4.411)	1.008(3.178)	0.917(2.066)	1.031(0.723)	1.000(0.525)	1.005(0.379)	1.010(0.311)	0.988(0.219)	1.002(0.159)
	ANCOVA-CHANGE	1.098(4.448)	1.022(3.188)	0.913(2.059)	1.025(0.712)	0.997(0.519)	1.004(0.373)	0.999(0.278)	0.986(0.199)	0.999(0.143)
	ANCOVA-POST	1.098(4.448)	1.022(3.188)	0.913(2.059)	1.025(0.712)	0.997(0.519)	1.004(0.373)	0.999(0.278)	0.986(0.199)	0.999(0.143)
	LMM	1.104(4.424)	0.997(3.197)	0.912(2.075)	1.038(0.833)	1.007(0.598)	1.006(0.424)	1.028(0.504)	0.992(0.351)	1.006(0.256)
1.5	ANOVA-POST	1.557(4.295)	1.668(3.048)	1.577(2.094)	1.525(0.831)	1.520(0.578)	1.522(0.412)	1.503(0.504)	1.492(0.347)	1.487(0.242)
	ANOVA-CHANGE	1.567(4.277)	1.650(3.039)	1.579(2.083)	1.526(0.724)	1.518(0.508)	1.526(0.361)	1.502(0.290)	1.494(0.199)	1.489(0.140)
	ANCOVA-CHANGE	1.556(4.313)	1.638(3.057)	1.586(2.086)	1.528(0.718)	1.516(0.502)	1.530(0.354)	1.500(0.250)	1.496(0.170)	1.491(0.119)
	ANCOVA-POST	1.556(4.313)	1.638(3.057)	1.586(2.086)	1.528(0.718)	1.516(0.502)	1.530(0.354)	1.500(0.250)	1.496(0.170)	1.491(0.119)
	LMM	1.557(4.295)	1.668(3.048)	1.577(2.094)	1.525(0.831)	1.520(0.578)	1.522(0.412)	1.503(0.504)	1.492(0.347)	1.487(0.242)

Parameter estimates (standard deviations) are presented in Table 1 for the five methods for values of  $\beta_1$ ,  $n$ , and  $\rho$ . All estimates are unbiased, so comparing the standard deviations of the estimates allows for comparison of the methods.

Table 1: Mean Parameter and standard deviation estimates across 1000 simulations for positive  $\beta_1$  values.

Simulated $\beta_1$ Values across methods		$\rho=0.1$			$\rho=0.5$			$\rho=0.8$		
		$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$
0.1	ANOVA-POST	-0.005	-0.087	0.106	0.014	0.017	0.010	0.010	-0.004	0.002
		0.056	0.062	0.049	0.053	0.060	0.070	0.056	0.060	0.069
	ANOVA-CHANGE	-0.005	-0.082	0.098	0.007	0.010	0.005	0.004	-0.004	0.001
		0.056	0.061	0.049	0.051	0.061	0.066	0.058	0.064	0.091
	ANCOVA-CHANGE	0.010	-0.078	0.101	0.006	0.006	0.002	0.001	-0.005	0.000
		0.057	0.062	0.045	0.054	0.060	0.060	0.071	0.068	0.107
	ANCOVA-POST	0.010	-0.078	0.101	0.006	0.006	0.002	0.001	-0.005	0.000
		0.057	0.062	0.045	0.054	0.060	0.060	0.071	0.068	0.107
	LMM	-0.005	-0.087	0.106	0.014	0.017	0.010	0.010	-0.004	0.002
		0.056	0.062	0.049	0.053	0.060	0.070	0.056	0.060	0.069
1.0	ANOVA-POST	0.104	-0.003	-0.088	0.038	0.007	0.006	0.028	-0.008	0.006
		0.058	0.067	0.088	0.211	0.382	0.644	0.490	0.794	0.976
	ANOVA-CHANGE	0.105	0.008	-0.083	0.031	0.000	0.005	0.010	-0.012	0.002
		0.059	0.066	0.086	0.267	0.470	0.756	0.856	0.988	1.000
	ANCOVA-CHANGE	0.098	0.022	-0.087	0.025	-0.003	0.004	-0.001	-0.014	-0.001
		0.060	0.065	0.085	0.275	0.479	0.772	0.920	0.999	1.000
	ANCOVA-POST	0.098	0.022	-0.087	0.025	-0.003	0.004	-0.001	-0.014	-0.001
		0.060	0.065	0.085	0.275	0.479	0.772	0.920	0.999	1.000
	LMM	0.104	-0.003	-0.088	0.038	0.007	0.006	0.028	-0.008	0.006
		0.058	0.067	0.088	0.211	0.382	0.644	0.490	0.794	0.976
1.5	ANOVA-POST	0.057	0.168	0.077	0.025	0.020	0.022	0.003	-0.008	-0.013
		0.058	0.094	0.102	0.412	0.713	0.949	0.848	0.987	1.000
	ANOVA-CHANGE	0.067	0.150	0.079	0.026	0.018	0.026	0.002	-0.006	-0.011
		0.058	0.094	0.105	0.522	0.825	0.985	1.000	1.000	1.000
	ANCOVA-CHANGE	0.056	0.138	0.086	0.028	0.016	0.030	0.000	-0.004	-0.009
		0.057	0.088	0.107	0.527	0.836	0.987	1.000	1.000	1.000
	ANCOVA-POST	0.056	0.138	0.086	0.028	0.016	0.030	0.000	-0.004	-0.009
		0.057	0.088	0.107	0.527	0.836	0.987	1.000	1.000	1.000
	LMM	0.057	0.168	0.077	0.025	0.020	0.022	0.003	-0.008	-0.013
		0.058	0.094	0.102	0.412	0.713	0.949	0.848	0.987	1.000

Bias (top number) and power (bottom number) are presented in Table 2 for the five methods for values of  $\beta_1$ ,  $n$ , and  $\rho$ . Generally, all methods produced estimates which were unbiased for the parameter of interest. Power was typically marginally higher for ANCOVA models compared to ANOVA and LMM, and increased for higher values of  $n$  and  $\rho$ .

Table 2: Power and bias estimates from 1000 simulations for true values of  $\beta_1$ .

in variance estimates when pre-treatment and post-treatment values are correlated. However, when correlation is low, results are dependent the on pre-post covariance structure. These discrepancies are addressed

and explained in detail in following paragraphs. Table 1 displays the difference between variance estimates which grows with increasing correlation, holding sample size and true  $\beta_1$  parameters constant.

Simulated $\beta_1$ Values Across Methods		$\rho=0.1$			$\rho=0.5$			$\rho=0.8$		
		$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$	$n=50$	$n=100$	$n=200$
0.1	ANOVA-POST	0.939	0.949	0.952	0.957	0.960	0.939	0.959	0.943	0.953
	ANOVA-CHANGE	0.936	0.950	0.950	0.949	0.947	0.946	0.940	0.936	0.956
	ANCOVA-CHANGE	0.935	0.950	0.949	0.947	0.950	0.952	0.944	0.942	0.948
	ANCOVA-POST	0.935	0.950	0.949	0.947	0.950	0.952	0.944	0.942	0.948
	LMM	0.939	0.949	0.952	0.957	0.960	0.939	0.959	0.943	0.953
1.0	ANOVA-POST	0.937	0.951	0.945	0.944	0.951	0.955	0.954	0.941	0.960
	ANOVA-CHANGE	0.941	0.952	0.947	0.950	0.954	0.951	0.962	0.952	0.959
	ANCOVA-CHANGE	0.945	0.952	0.947	0.943	0.950	0.951	0.955	0.947	0.956
	ANCOVA-POST	0.945	0.952	0.947	0.943	0.950	0.951	0.955	0.947	0.956
	LMM	0.937	0.951	0.945	0.944	0.951	0.955	0.954	0.941	0.96
1.5	ANOVA-POST	0.954	0.947	0.951	0.957	0.957	0.951	0.964	0.951	0.948
	ANOVA-CHANGE	0.953	0.948	0.952	0.954	0.953	0.954	0.961	0.952	0.950
	ANCOVA-CHANGE	0.950	0.942	0.953	0.953	0.947	0.949	0.960	0.954	0.952
	ANCOVA-POST	0.950	0.942	0.953	0.953	0.947	0.949	0.960	0.954	0.952
	LMM	0.954	0.947	0.951	0.957	0.957	0.951	0.964	0.951	0.948
-0.1	ANOVA-POST	0.948	0.954	0.959	0.952	0.961	0.952	0.964	0.949	0.944
	ANOVA-CHANGE	0.947	0.953	0.962	0.951	0.961	0.950	0.965	0.952	0.945
	ANCOVA-CHANGE	0.950	0.952	0.959	0.953	0.954	0.950	0.959	0.955	0.944
	ANCOVA-POST	0.950	0.952	0.959	0.953	0.954	0.950	0.959	0.955	0.944
	LMM	0.948	0.954	0.959	0.952	0.961	0.952	0.964	0.949	0.944
-1.0	ANOVA-POST	0.953	0.954	0.957	0.955	0.959	0.953	0.957	0.959	0.955
	ANOVA-CHANGE	0.953	0.953	0.956	0.955	0.959	0.948	0.957	0.948	0.955
	ANCOVA-CHANGE	0.954	0.951	0.956	0.952	0.948	0.95	0.962	0.950	0.949
	ANCOVA-POST	0.954	0.951	0.956	0.952	0.948	0.95	0.962	0.950	0.949
	LMM	0.953	0.954	0.957	0.955	0.959	0.953	0.957	0.959	0.955
-1.5	ANOVA-POST	0.958	0.948	0.946	0.956	0.950	0.950	0.954	0.954	0.948
	ANOVA-CHANGE	0.958	0.947	0.944	0.957	0.949	0.952	0.957	0.953	0.950
	ANCOVA-CHANGE	0.958	0.951	0.942	0.96	0.949	0.947	0.955	0.950	0.951
	ANCOVA-POST	0.958	0.951	0.942	0.96	0.949	0.947	0.955	0.950	0.951
	LMM	0.958	0.948	0.946	0.956	0.950	0.950	0.954	0.954	0.948

Confidence interval coverage probabilities are presented in Table 3 for the five methods for values of  $\beta_1$ ,  $n$ , and  $\rho$

**Table 3:** 95% confidence interval coverage probabilities from 1000 simulations.

The two scenarios of ANCOVA (ANCOVA-POST and ANCOVA-CHANGE) give identical measures of variability, regardless of differing combinations of  $\rho$ ,  $\beta_1$  and  $n$ . These results can be seen numerically in Table 1 and visually assessed in Figure 1, and are further supported by theoretical derivation of variance. Lastly, the LMM analysis results are equivalent to the ANOVA-POST in terms of estimates and their standard deviations, regardless of the assumed within-subject covariance structure,  $R_r$ . The LMM approach is evaluated with and without a KR adjustment, ultimately showing no difference. These results are consistent with the literature, given the smallest simulated sample size was  $n=50$ .

Across all methods, ANCOVA models consistently performed best compared to the other methods, regardless of  $\rho$ ,  $\beta_1$  and  $n$ , as has been demonstrated in existing literature [18]. The ANCOVA methods (ANCOVA-CHANGE and ANCOVA-POST) are compared with ANOVA-POST, ANOVA-CHANGE and LMM in terms of variance of the estimate of  $\beta_1$  (Figure 1). Except at the lowest combination of  $\rho$  and  $n$  values where discrepancies between differing method variances is negligible, ANCOVA had the lowest variability. Similarly, the 95% coverage intervals are smallest under ANCOVA, and widest in ANOVA-POST (Table 4). Compared to ANOVA-CHANGE, ANCOVA models have less variation. These results are also similar when ANCOVA is compared with LMM, except when the correlation is close to zero.

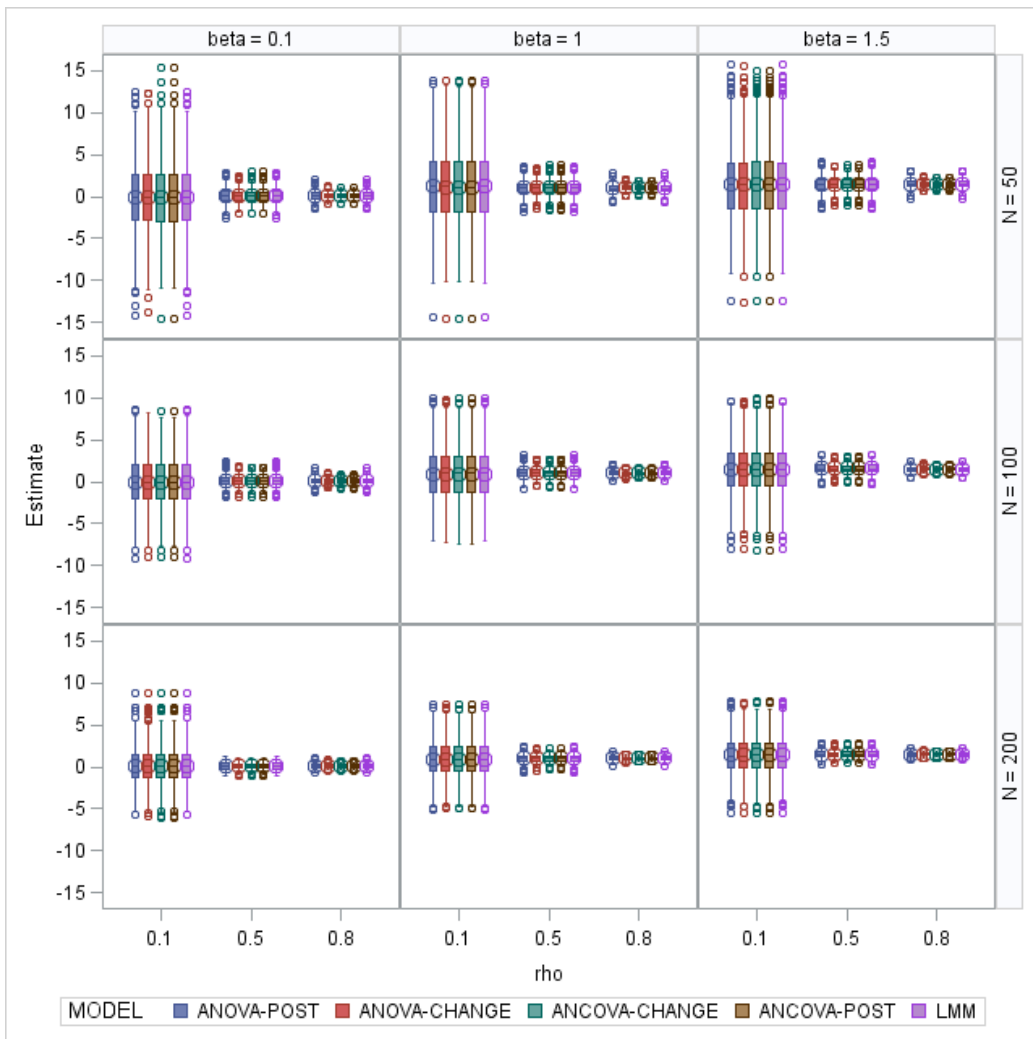
It is of interest to compare the simulations to theoretical

Method	$\hat{\beta}$	95% Confidence Interval		SE( $\hat{\beta}$ )	P-value
<b>1<sup>st</sup> Session</b>					
ANOVA-POST	0.16	-0.09	0.41	0.123	0.192
ANOVA-CHANGE	0.14	0.05	0.24	0.047	0.005
ANCOVA-CHANGE	0.14	0.05	0.24	0.046	0.004
ANCOVA-POST	0.14	0.05	0.24	0.046	0.004
LMM	0.16	-0.09	0.41	0.123	0.191
<b>4<sup>th</sup> Session</b>					
ANOVA-POST	-0.08	-0.28	0.12	0.099	0.409
ANOVA-CHANGE	0.09	-0.02	0.21	0.056	0.107
ANCOVA-CHANGE	0.07	-0.05	0.19	0.059	0.243
ANCOVA-POST	0.07	-0.05	0.19	0.059	0.243
LMM	-0.08	-0.28	0.12	0.099	0.409

Table 4 presents parameter estimates, their standard errors, 95% confidence intervals and p-values for the dental data example. Results are consistent with simulation data conclusions since all methods produced unbiased estimates for the treatment effect, and ANCOVA models had smaller standard errors compared to ANOVA and LMM models.

**Table 4:** Estimates of toothbrush effect on bacterial plaque index in 1<sup>st</sup> and 4<sup>th</sup> sessions.

expectations for the variance of the treatment effect estimates. As the correlation between pretreatment and post-treatment observations approaches zero, the variance for ANOVA-POST, ANCOVA-POST, ANCOVA-CHANGE, and LMM should be approximately equivalent, but the variance for ANOVA-CHANGE is two times that of the others.



Boxplots for parameter estimates for the 1000 simulations for the combinations of  $\beta_1$ ,  $n$ , and  $\rho$  are displayed in Figure 1. Consistent with the data tables, all parameter estimates are unbiased, and the boxplots highlight differences in variability for the models. In general, variance was much larger for small values of  $\rho$  and small  $n$ . ANCOVA models have smaller variances compared to ANOVA and LMM, though differences are quite small.

**Figure 1:** Distribution of treatment effects estimates varied by correlation, sample size and true positive  $\beta_1$  values under  $Y_0 \sim N(0, 1)$ .

When  $\rho=0.1$  and assuming equal variance, the variance of ANOVA-CHANGE should theoretically approach  $2(1-\rho=0.1)$ , or 1.8 times that of ANOVA-POST, and ANCOVA methods should approach  $(1-\rho^2)$ , or 0.99 of ANOVA-POST. Similarly, for  $\rho = 0.1$ , the ratio of ANOVA-POST to ANOVA-CHANGE should be 1 and ANOVA-POST to ANCOVA should be 0.75. Finally, at  $\rho=0.8$ , ANOVA-CHANGE and both ANCOVA models should be at 0.4 and 0.36 times ANOVA-POST respectively. However, permitting pre and post measurements to have different variances leads to differing results, particularly apparent when correlation is low, such that the variance of ANOVA-CHANGE approaches the variance of ANOVA-POST as  $\rho$  approaches zero and the variance of post-measures becomes increasingly greater than pre-measures. Presented simulations are for data simulated assuming this unstructured covariance matrix. For example, the variance of ANOVA-CHANGE and ANOVA-POST are approximately equal when  $\rho=0.1$ ; had we assumed equal variance, ANOVA-CHANGE would be almost twice that of ANOVA-POST. These are made apparent given the following ratio of variances:

As a result, the variance of ANOVA-CHANGE approaches the variance of ANOVA-POST as  $\rho$  approaches zero and  $\beta_1 > 0.5$ . Under the given method of simulating pre and post treatment variances where  $\sigma_{pre} = 1, \beta_1 = 1.5$ , and  $\sigma_{post} = \beta_1 \sigma_{pre} / \rho$ , the ratio of the variances is given by:

$$\frac{(\sigma_{pre}^2 + \sigma_{post}^2 - 2\rho\sigma_{pre}\sigma_{post})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma_{post}^2} \approx 0.99, \text{ for } \rho = 0.1.$$

Since the HCS structure is more conservative in its assumptions (permitting the pre- and post-treatment effects to have different variances), it is used to report the main simulation results of the study. As correlation increases beyond 0.5, results become less sensitive to the pre-post measure covariance structure.

Power for testing  $\beta_1 = 0$  was assessed for the methods under the simulated conditions (Table 2). Under the primary simulation method assuming  $Y_0 \sim N(0, 1)$ , power across methods did not vary by a large

degree. Observing power across simulated scenarios at the lowest sample size ( $n=50$ ), there is marginally higher power in ANCOVA and ANOVA-CHANGE methods over ANOVA-POST and LMM. Overall, ANCOVA methods achieve the greatest power, intuitively so given ANCOVA leads to the lowest variability. As correlation between pre-and post-measurements increase, the difference in power between ANCOVA and ANOVA-CHANGE compared to ANOVA-POST and LMM, grows appreciably, while ANOVA-CHANGE nears that of ANCOVA as correlation approaches one. Finally, increases in sample size leads to increased power for detecting a significant treatment effect similarly across methods, meaning that an increase in sample size does not appear to affect any single methods statistical power more than other methods. Additional results from simulations with an increased variance (assuming  $Y_0 \sim N(0,9)$ ), are reported.

### Data Example

To illustrate the application of the five methods, we consider data from a dental hygiene study characterized by small sample size [19]. There were a total of 32 subjects, randomized to two treatment groups based on type of toothbrush, and effectiveness was measured by reduction in bacterial plaque index over time. Of the original four independent sessions, we analyzed data from the first and last sessions, comparing pre and post treatment outcomes. The results for the first and last session are summarized in Table 4. In the first session, the pre-treatment mean (standard error) for treatment group 1 and treatment group 2 are 1.31 (0.35) and 1.33 (0.38), respectively, and similarly for the last session, 1.54 (0.26) and 1.36 (0.27), respectively. The pre-treatment measures between groups show no significant difference for either of the sessions. The correlations in the pre and post treatment measures are 0.91 and 0.82 for the first and last sessions, respectively. In line with simulations and theoretical expectations, it follows that under the first session, ANCOVA-CHANGE and ANCOVA-POST performed equally well among methods, presenting the lowest standard error for treatment effect. However, in the last session, ANOVA-CHANGE actually presented a slightly lower standard error compared to ANCOVA methods (0.0564 compared to 0.0592). Given the high correlation between pre and post measurements in this data set, it follows theoretical expectations that ANOVA-CHANGE is extremely close to that of ANCOVA models. In the case where  $\rho=0.91$ , ANOVA-CHANGE should theoretically produce variance that is 1.05 times ANCOVA. In both sessions, ANOVA and LMM exhibit larger variance than ANCOVA methods. The variance estimates from ANOVA and LMM are 2.7 times greater in the first session and 1.7 times greater in the second session than ANCOVA.

### Discussion

This paper compares four traditional approaches (ANOVA-POST, ANOVA-CHANGE, ANCOVA-POST, and ANCOVA-CHANGE) and a more modern approach (LMM) used in the analysis of pre-post data. These five methods are compared via simulated data from a typical clinical trials setting, where pre-treatment groups are assumed equally allocated through proper randomization, and the primary interest is to examine estimates of treatment effect. Comparisons of these methods have been investigated in theoretical framework (Brogan and Kutner [3]; Dimitrov and Rumrill [4]; Frison and Pocock [2]; Huck and McLean [4]; Laird [14]), and in a similar manner as Egbewale et al. [12] we review these methods and discuss them in terms of several simulated circumstances, as well as a real data application.

Overall, all of the five methods in the simulated scenarios yield equally unbiased treatment effect estimate. However, their performance

(in terms of variance, type-I error and 95% CI coverage) varies with pre-treatment group differences as indicated in previous literature [7,11,12]. For example, LMM is found to be more conservative compared to the ANCOVA methods. Consistent with previous literature, ANCOVA models have the smallest variance, highest power, and nominal 95% confidence interval coverage compared to ANOVA-POST, ANOVA-CHANGE, and LMM. Similar to conclusions reached by Vickers and Altman [13], in our simulation study as correlation between pre-and post-treatment measures increase, ANOVA-CHANGE approaches ANCOVA in both variance and power. However, in all but the most extreme cases (i.e. when  $\rho \approx 0$  or 1), ANCOVA methods are the most optimal, achieving the greatest power and lowest variability. Thus, in the case of balanced pre-treatment data, our results are consistent with most existing literature, in that ANCOVA is a preferred method. This may not hold in situations with some degree of imbalance between treatment groups at baseline and different levels of pre-post correlation [12].

We also examine the robustness of these methods when the pre-post measures are simulated under CS versus HCS covariance structures. The data simulated under HCS produces the greatest effect on the ANOVA-CHANGE results, where treatment effect variances are particularly influenced at lower correlations. When pre and post treatment measures have equal variance and low correlation, ANOVA-POST outperforms ANOVA-CHANGE, but as the imbalance between pre-post variance grows, the two methods perform similarly. In practical applications, when one does not have control over pre-and post- treatment variances, results demonstrate that one could reasonably expect ANOVA-CHANGE to consistently perform better than ANOVA-POST when  $\rho \leq 0.5$  regardless of equality of variances. When  $\rho \leq 0.5$ , the best performing method will depend on the degree of equality of the variances in pre-and post-measurements. Nevertheless, both methods are still consistently outperformed by ANCOVA.

In our simulation study, the LMM approach performs only as well as ANOVA-POST. However, these simulations assume no missing data. In clinical trials evaluating patients over time, on the other hand, missing data are common (i.e. some patients are lost to follow-up and post-treatment measurements are never recorded). In such cases, the mechanism of missing data as defined by Little and Rubin [20], along with method of analysis are important in reaching unbiased results. When data are missing completely at random (MCAR), the method of analysis makes little difference, i.e. the generalized least squares (GLS) based methods and LMM should provide equivalent results [21]. However, in the case when data are missing at random (MAR), GLS can lead to biased inference on effects, whereas the LMM approach which relies on likelihood provides unbiased results when the within-subject covariance matrix,  $R_i$ , is specified correctly [21]. When data are missing not at random (MNAR), all five methods may lead to biased results. Thus, conclusions made in this study with regards to the LMM approach compared to the other approaches hold only when data are MCAR or not missing. In the case of MAR data, LMM may be more optimal.

### Conclusion

Despite decades of long scrutiny of this topic and our extensive simulation study under a wide range of scenarios, there still remain several avenues of future work. For example, in the clinical trials setting, there are often many more important covariates which are included to deal with baseline covariate imbalance. Additionally, this study assumed that the treatment assignment was random, which is usually the case in clinical trials, resulting in equal pre-treatment values among

treatment groups. As it was discussed in the introduction, results may differ when pre-treatment measures are unequal, particularly affecting the bias of ANCOVA-POST. Future work is needed to assess how ANCOVA-CHANGE as presented by Laird [14], performs under varying degrees of pre-treatment imbalance as in the simulation study performed by Egbewale [11]. Finally, these methods could be explored under the generalized linear mixed model framework with non-Gaussian pre-post data.

#### Acknowledgement

The study concept and design and acquisition of data is by MG. Analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content and final approval of manuscript is made by all co-authors. The authors thank (names omitted to maintain integrity of the review process) for helpful comments. The manuscript represents the views of the authors. It is a byproduct of group project work by graduate students in an advanced statistical methods course.

#### References

1. Matthews J, Altman DG, Campbell M, Royston P (1990) Analysis of serial measurements in medical research. *BMJ: British Medical Journal* 300: 230.
2. Frison L, Pocock SJ (1992) Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in medicine* 11: 1685-1704.
3. Brogan DR, Kutner MH (1980) Comparative analyses of pretest-posttest research designs. *The American Statistician* 34: 229-232.
4. Huck SW, McLean RA (1975) Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin* 82: 511.
5. Jennings E (1988) Models for Pretest-Posttest Data: Repeated Measures ANOVA Revisited. *Journal of Educational and Behavioral Statistics* 13: 273-280.
6. Dimitrov DM, Rumrill J, Phillip D (2003) Pretest-posttest designs and measurement of change. *Work: A Journal of Prevention, Assessment and Rehabilitation* 20: 159-165.
7. Samuels ML (1986) Use of analysis of covariance in clinical trials: a clarification. *Controlled clinical trials* 7: 325-329.
8. Van Breukelen GJ (2006) ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of clinical epidemiology* 59: 920-925.
9. Lord FM (1967) A paradox in the interpretation of group comparisons. *Psychological Bulletin* 68: 304-305.
10. Liang KY, Zeger SL (2000) Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhyā: The Indian Journal of Statistics Series B*: 134-148.
11. Senn S (2006) Change from baseline and analysis of covariance revisited. *Statistics in medicine* 25: 4334-4344.
12. Egbewale BE, Lewis M, Sim J (2014) Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. *BMC medical research methodology* 14: 49.
13. Vickers AJ, Altman DG (2001) Statistics notes: analysing controlled trials with baseline and follow up measurements. *BMJ: British Medical Journal* 323: 1123.
14. Laird N (1983) Further comparative analyses of pretest-posttest research designs. *The American Statistician* 37: 329-330.
15. Kenward MG, Roger JH (1997) Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics* 53: 983-997.
16. Schaalje GB, McBride JB, Fellingham GW (2002) Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models. *Journal of Agricultural, Biological, and Environmental Statistics* 7: 512-524.
17. Senn S (2014) Various varying variances: The challenge of nuisance parameters to the practising biostatistician. *Stat Methods Med Res*.
18. Betticher DC, Delmore G, Breitenstein U, et al (2013) Efficacy and tolerability of two scalp cooling systems for the prevention of alopecia associated with docetaxel treatment. *Support Care Cancer* 21: 2565-2573.
19. Parizotto SPC, Rodrigues CRMD, Singer JM, Sef HC (2003) Effectiveness of low cost toothbrushes, with or without dentifrice, in the removal of bacterial plaque in deciduous teeth. *Pesquisa Odontológica Brasileira* 17: 17-23.
20. Little RJA, Rubin DB (2002) *Statistical Analysis with Missing Data*: Wiley.
21. Fitzmaurice GM, Laird NM, Ware JH (2012) *Applied longitudinal analysis*: John Wiley & Sons.