

## Original Research

# Design and Statistical Methods for Handling Covariates Imbalance in Randomized Controlled Clinical Trials: Dilemmas Resolved

Bolaji E. Egbewale, PhD\*

Department of Community Medicine, College of Health Sciences, Ladoké Akintola University of Technology, Ogbomoso, Oyo State, Nigeria

\*Corresponding author

Bolaji E. Egbewale, PhD

Associate Professor of Biostatistics, Department of Community Medicine, College of Health Sciences, Ladoké Akintola University of Technology, Ogbomoso, Oyo State, Nigeria; E-mail: [beegbewale@lutech.edu.ng](mailto:beegbewale@lutech.edu.ng)

### Article information

**Received:** October 1<sup>st</sup>, 2021; **Revised:** October 28<sup>th</sup>, 2021; **Accepted:** November 5<sup>th</sup>, 2021; **Published:** November 15<sup>th</sup>, 2021

### Cite this article

Egbewale BE. Design and statistical methods for handling covariates imbalance in randomized controlled clinical trials: Dilemmas resolved. *Clin Trial Pract Open J.* 2021; 1(1): 95-102. doi: [10.17140/CTPOJ-1-121](https://doi.org/10.17140/CTPOJ-1-121)

## ABSTRACT

### Introduction

In practice, between groups baseline imbalance following randomization not only opens effect estimate to bias in controlled trials, it also has certain ethical consequences. Both design and statistical approaches to ensure balanced treatment groups in prognostic factors are not without their drawbacks. This article identified potential limitations associated with design and statistical approaches for handling covariate imbalance in randomized controlled clinical trials (RCTs) and proffered solutions to them.

### Methods

A careful review of literatures coupled with a robust appraisal of statistical models of methods involved in a way that compared their strength and weaknesses in trial environments, was adopted.

### Results

Stratification breaks down in small sample size trials and may not accommodate more than two stratification factors in practice. On the other hand, minimization that balances for multiple prognostic factors even in small trials is not a pure random procedure and in addition, could present with complexities in computations. Overall, either minimization or stratification factors should be included in the model for statistical adjustment. Statistically, estimate of effect by change score analysis (CSA) is susceptible to direction and magnitude of imbalance. Only analysis of covariance (ANCOVA) yields unbiased effect estimate in all trial scenarios including situations with baseline imbalance in known and unknown prognostic covariates.

### Conclusion

Design methods for balancing covariates between groups are not without their limitations. Both direction and size of baseline imbalance also have profound consequence on effect estimate by CSA. Only ANCOVA yields unbiased treatment effect and is recommended at all trial scenarios, whether or not between groups covariate imbalance matters.

### Keywords

Randomization; Covariate imbalance; Stratification; Minimization; Change score analysis; ANCOVA.

## INTRODUCTION

The essence of randomization exercise is to bring about comparable treatment groups in a controlled trial. Treatment groups are expected to be similar in factors - known or unknown, that are prognostic of outcome of interest for investigator to draw a valid inference on treatment effect.<sup>1</sup> However, in practice, balance in covariates between groups is often not attained with randomization.<sup>1,2</sup> The resultant imbalance subtly opens the trial intervention to a degree of misrepresentation of estimates of effect.<sup>3</sup> The need

for a correct and more reliable inference on the effect of interventions under trial has led to efforts to ensure that balance is achieved in the distribution of covariates between treatment groups. Potentially great studies at onset had ended up been declared inconclusive owing to issues related to improper design, in particular, imbalance in risk factors between treatment groups. For example, in their study, Rosenberger et al<sup>4</sup> recall an abrupt termination of a trial on the role of erythropoietin in maintaining normal hemoglobin concentrations in patients with metastatic cancer. The trial which was supposed to be a major study involving 139 clinical sites

and 939 patients was declared inconclusive owing to issues related to covariate handling.

Dealing with covariate imbalance between treatment groups is both design and statistical in nature and needs given consideration when planning a controlled experiment by researchers. However, it still appears that the statistical community is unclear on how to deal with covariates at the design stage, especially on the first line strategy for balancing important prognostic factors at the design of randomized controlled clinical trials (RCTs).<sup>4-8</sup> The general consensus seems to be that, whichever method is employed at the design stage to attempt balance in covariate distribution, an adjusted statistical analysis that takes into account important covariate imbalance should take precedence over the unadjusted analysis.<sup>4,6,7,9-11</sup> Nonetheless, crude unadjusted analysis by analysis of variance (ANOVA) is still common and statistically adjusted analysis is either model-based analysis of covariance (ANCOVA) or analysis based on change between the pre and post-treatment score. Thus, leaving researchers options to choose from at any clinical trial scenario. This study aimed at exploring the design and statistical methods involved in handling covariate in RCTs with a view to addressing associated limitations. This will guide researchers' choice and preparations while designing future controlled experiments.

## METHODOLOGY

This article presents known design and statistical methods used in dealing with covariates imbalance in randomized controlled trials involving a single post treatment assessment of a continuous outcome variable. Existing facts on appropriate design and statistical methods being used in dealing with covariate imbalance scattered in previous literatures were reviewed and synthesized in this paper. In addition, a careful appraisal of mathematical models or equations of the standard statistical methods being used at different trial scenarios was considered in a way that makes the work easily accessible by non-mathematics specialists. The article used illustrations and mathematical examples to describe mechanisms of covariate handling by these methods for ease of comprehension.

### State of the art Narrative Synthesis and Model Equations

**Baseline imbalance:** A particular class of covariate imbalance between treatment groups is that which involves baseline difference in an outcome variable of interest. In primary care setting, randomized controlled trials often involve quantifying a numerical outcome variable at baseline and repeating the same after treatment. Measurement of treatment effect often depends on the observed changes from the baseline value within a designated period of time after treatments have been administered. At such time, it is the change in mean value of an outcome variable that is under investigation. For example, two or more diets may be compared for the mean change in body weight they produce,<sup>12</sup> two or more treatments for hypertension may be compared for the mean changes in diastolic or systolic blood pressure which they produce; two or more cancer therapies may be with respect to the mean changes in tumour size they produce<sup>13</sup> and finally, the effect of exercise and diet on obesity may also be compared in osteoarthritis patients in terms of mean change in body mass index. In all of these empirical

examples, the difference in baseline score of the outcome variable between treatment groups has a direct influence on treatment effect. If one group has a higher mean score at baseline, then, an unfair advantage/disadvantage arise for that group in relation to treatment effect compared to the other group.

**Adjusting for baseline imbalance – the analogy:** Not adjusting for baseline covariate imbalance in an RCT setting could be likened to two athletes who prepare to run a 100 m race but start at different points on the track. This gives one an unfair advantage over the other, and the winner of the race may not be the truly faster runner over 100 m. The level of unfairness (and implication on the result) ties in with the size of the difference in starting position. Clearly, the result or outcome would not be a precise or correct reflection of their true performance because the baseline difference is a factor that has a direct relationship with the outcome - in this case time to finish. So therefore, it is only expected in the interest of a fair result and correct measurement of performance, that the baseline difference at the starting point be 'accounted for' in the system so that the measure of true performance (time to finish) of the two athletes would be a valid measure. Nobody bothers of course if the two athletes differ at baseline in some respects that do not affect the outcome; for example, colour of their outfit.

In a two parallel-arm RCT setting, the two athletes in the above scenario represent the two treatment groups (treated and control), while difference in starting point is analogous to baseline imbalance as it has an established relationship with the outcome variable, and difference in time-to-finish the race represents treatment effect. This analogy of course, may not completely represent what transpires in a trial setting, where the average scores of various responses (rather than individual responses) within each treatment group is compared.

### Design Methods

Various methods used at the design stage to attempt balance in prognostic factors between treatment groups include: stratification and minimization. Also, commonly used is the basic simple randomization; the principle being that between-group inequalities are reduced through chance correction with increased sample units. This approach is known to yield an unbalanced treatment groups, especially when the design is implemented on a study with small sample size. The usual practice then is to unduly increase sample size by randomizing more patients into treatment groups over and above the minimum number required to have a level of study power.<sup>14</sup> The issue here is that in certain instances when the true effect of the compound being tried has not been completely ascertained or when there is an indication that the drug being tested may still have some side effects, more patients for example would invariably be enrolled either as recipients of such compound that may turn out to be ineffective or in the end found to adversely affects their health. When this happens, it begs the question of when should a researcher allow the desire to attain balance in prognostic factors to override the ethical responsibility of patients' safety and protection? A possible way out of this scenarios when they exist would be for researchers to accept the outcome of simple randomization based on minimum sample size associated with a reasonable level

of statistical power and account for the between groups imbalance in prognostics factors during statistical analysis.

Furthermore, with stratified randomization, stratification breaks down in trials with small sample size and especially when there are many stratification factors to consider for balance. When there are large numbers of covariates with each presenting with multiple levels, stratification procedure would require that separate allocation lists be prepared at each level of an identified covariate.<sup>15</sup> Inevitably, such multi-stratification may pose logistical problems, making the whole exercise almost impracticable. For example, in a trial with 5 prognostic factors at levels 2, 2, 3, 3 and 4 respectively, 144 separate allocating lists have to be prepared and maintained for as long as the study lasts. To avoid such scenarios where stratified randomisation would be rather difficult and almost impracticable researchers are advised to keep strata few. Previous authors have argued that in practice it is rarely possible to stratify for more than two prognostics factors, especially in small sample trials.<sup>16-18</sup>

On the other hand, minimization as a design approach uses information about patients who are already in the trial to determine treatment assignment for the incoming participants, such that differences between groups are minimized.<sup>5</sup> Thus, the next patient is usually assigned to the treatment group with the lower covariate marginal total. As was noted previously,<sup>19</sup> various authors have submitted that minimization proffers a solution to the limitations of stratification in balancing for multiple prognostic factors in small trials, as the procedure makes treatment groups similar in several important features even with small samples.<sup>16-18,20</sup> However, it has been argued that minimization is open to predictability of assignment<sup>3,20</sup> and researchers can therefore add a random element to the procedure at least to reduce prediction of assignment.<sup>3</sup> Another drawback with minimization is the complex computation process involved; however, a user-friendly program that manages this has been developed.<sup>20</sup>

**Statistical Methods**

Statistical methods for handling baseline imbalance for a single post-treatment assessment of a continuous outcome variable are change score analysis (CSA), percentage change score and ANCOVA. However, the use of percentage change score for the evaluation of treatment effect in clinical trial settings has been shown not to be statistically efficient.<sup>21</sup> Percentage change score analysis presents large error variance of the estimator and as a result has poor power to detect a difference in treatment effect when one exists. Since percentage change score has been found to be grossly inefficient, it would not be taken further in this article. Also, since crude analysis by analysis of variance ANOVA is still popular in evaluating treatment effect in trial scenarios irrespective of the distribution of baseline covariates, it is also considered alongside with CSA and ANCOVA in this paper.

**The Traditional Analysis of Variance**

The simplest and perhaps the most common approach to estimating treatment effect between treatment groups is the crude com-

parison of post-test scores using statistical tests, such as *t*-test or ANOVA for quantitative outcome variables.

The underlying model representation for a two-group trial, alluding, is given as:

$$Y_{ij} = \beta_0 + \lambda_i + \epsilon_{ij} \dots \dots \dots (1.0)$$

$i=1,2; j=1 \dots n$ , or

$$Y_{ij} = \mu + \lambda_i + \epsilon_{ij} \dots \dots \dots (1.1)$$

where  $Y_{ij}$  is the posttreatment score for the  $j^{th}$  patient in the  $i^{th}$  group,  $\beta_0$  or  $\mu$  is the common mean value of the outcome variable,  $\lambda_i$  is the treatment effect in the group and  $\epsilon_{ij}$  is the error term. There is clearly no term in the model for ANOVA of post-test to accommodate any systematic variation in the groups that is related with the outcome as ANCOVA does and this explains the larger error term associated with the estimate from ANOVA.

Essentially, with respect to ANCOVA, the model extends to:

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 Z_{ij} + \epsilon_{ij} \dots \dots \dots (1.2)$$

$G_{ij}$  is a treatment indicator,  $\beta_1$  is the group difference in  $Y$  adjusted for differences on  $Z$ .

When  $\beta_2$  is close to 0, then it approximates the ANOVA model. It becomes obvious, therefore, that the difference between the statistical methods under investigation in this study actually lies in the different ways in which each of them responds to the presence of baseline imbalance. For example, as mentioned earlier, with ANOVA of post-test,  $\beta_2=0$ , for ANOVA of change  $\beta_2=1$ , and with ANCOVA  $\beta_2$  is computed such that the residual post-test variance is minimized, thereby minimizing the standard error of the treatment effect estimate.<sup>22</sup>

The basis for the statistical procedure of ANOVA on post-test is that baseline scores of the outcome are comparable between the treatment arms by randomization. In other words, the statistical procedure assumes that baseline data for the groups to be compared are sufficiently similar and thus only the post treatment score is entered into the analysis.

In a RCT, let the baseline measurement from the control group be represented by random variable  $Z_c$  and the outcome variable by  $Y_c$ ; the corresponding measurements for the intervention group are  $Z_T$  and  $Y_T$  for baseline and outcome respectively;

Thus,

$$E(Y_c) = \mu \text{ and } E(Y_T) = \mu + \lambda \dots \dots \dots (1.3)$$

and since by randomization the baselines have a common mean

value:

$$E(Z_c)=E(Z_t)=\mu_z \dots \dots \dots (1.4)$$

$$E(\bar{Y}_t)-E(\bar{Y}_c)=\lambda \dots \dots \dots (1.5)$$

From the above, the sample mean of the outcome in the control group  $\bar{Y}_c$  will have expectation of  $\mu$  and in the intervention group it will be  $\mu+\lambda$ ; hence, the difference in means will have expectation  $\lambda$  as required. This shows that the analysis based on post score is unbiased yielding an unbiased estimate of treatment effect. Previous authors,<sup>23</sup> have used simulation study to demonstrate the fact that both other methods, change score analysis CSA and ANCOVA, at least when the treatment groups are balanced at baseline are expected to yield the same unbiased estimate as ANOVA.

However, when treatment groups are not comparable at baseline and such that there exists a correlation between baseline and outcome scores, then,

$$E(Z_t) \neq E(Z_c) \dots \dots \dots (1.6)$$

Direct comparison of outcomes from the groups becomes invalid and the resultant estimate is not unbiased.

Thus the true effect is modelled as,

$$E(\bar{Y}_t - \bar{Y}_c | \bar{Z}_t, \bar{Z}_c) = \lambda + \rho(\bar{Z}_t - \bar{Z}_c) \dots \dots \dots (1.7)$$

However, since ANOVA model does not have such a term that accounts for baseline imbalance, its estimate of treatment effect will not respond to baseline-outcome correlation, direction and magnitude of baseline imbalance as in the last equation. Again, this fact has been variously demonstrated,<sup>23,24</sup> their simulation studies highlight the non-responsiveness of ANOVA to various degrees of baseline imbalance and prognostic strength.

**Change Score Analysis**

If the analysis is based on ANOVA of change from baseline, there is a conscious effort to bring about balance in baseline data in the treatment groups by analysing the absolute difference between the baseline and the post-test score in the groups; however, baseline scores are not included in the analysis as independent variables.

Here, analysis concerns (Y-Z)s in the two groups.

The underlying model is given as:

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + Z_{ij} + \epsilon_{ij} \dots \dots \dots (1.8)$$

Where  $Z_{ij}$  is the baseline value for the  $j^{th}$  patient in the group. For change score analysis, the regression coefficient for the covariate is equal to 1.

Again, supposing treatment groups are comparable by

randomization, the expectation will be,

$$E(\bar{Y}_t - \bar{Z}_t) - E(\bar{Y}_c - \bar{Z}_c) = (\mu + \lambda - \mu_z) - (\mu - \mu_z) = \lambda \dots \dots \dots (1.9)$$

and this demonstrates the fact that CSA yields unbiased estimate when treatment groups are comparable.

$$\text{Var}(Y_t) = \text{Var}(Y_c) = \text{Var}(Z_t) = \text{Var}(Z_c) = \sigma^2 \dots \dots \dots (1.10)$$

However, the associated variance differs from the variance of the unadjusted analysis. Whereas the variance of the unadjusted analysis is completely independent of the baseline outcome correlation and if we assume randomisation makes groups similar, then<sup>15</sup>

(where  $Y_t, Y_c$  are the outcome variables for both treated and control groups and  $Z_t$  and  $Z_c$  are the baseline variables for both treated and control groups)

but the variance of the CSA is given as;

$$\text{Var}(Y_t - Z_t) = \text{Var}(Y) + \text{Var}(Z) - 2\text{cov}(Y, Z) = \sigma^2 + \sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1 - \rho) \dots \dots \dots (1.11)$$

where  $\rho$  is the correlation between Y and Z, assumed to be same for both groups.

The above presentation shows that the analysis of change scores from baseline has an entirely different variance structure compared with analysis from post-score comparison, and this has implications for the precision of the effect estimate. For example, if the correlation  $\rho$  exceeds 0.5 then a small variance (standard error) results and the analysis becomes more powerful than the comparison of post-test outcomes. However, if the correlation is below 0.5, using analysis of change from baseline (CSA) will bring about increased variance - a large standard error and less power to detect a real difference between groups. This fact was observed,<sup>25</sup> who argued that the estimate by analysis of change would not always have a lesser magnitude of associated variability compared with that from an unadjusted analysis - crude comparison of post-treatment scores. He states that precision will be lost by change score analysis if the baseline-outcome correlation is less than 0.5, he further argued that only ANCOVA should be used if chance imbalance in treatment groups is to be taken into consideration since ANCOVA takes account of regression to mean whereas CSA does not. Similarly, ANOVA will usually fail to detect a bias in an effect estimate since there is no term in the ANOVA model that takes account of the baseline difference in the treatment groups.<sup>26</sup> These considerations suggest that various methods used for the analysis of clinical trials can have a very profound effect on the estimate of treatment effect. In fact, under the same experimental conditions, ANOVA, CSA and ANCOVA have been observed to yield estimates of effect that are conspicuously different in size and precision.<sup>22,23,27,28</sup>

Although the estimate of effect by CSA may not be

markedly affected by the degree of baseline-outcome correlation, its regression coefficients are markedly influenced by both the magnitude and direction of imbalance. When imbalance is in the opposite direction to that of the treatment effect, that is, in this case, the control group have lower mean value (i.e. are better) at baseline, the absolute value of the effect estimate by CSA increases in relation to the underlying treatment effect. Here, the higher the level of imbalance the wider the distance between the estimated effect and zero, and the more likely it is to infer a significant result by CSA. The reason for this seeming exaggeration is because the control group is treated by change score analysis as if it enjoys a level of treatment which was never assigned to it, giving rise to false positives. On the other hand, if the imbalance is in the same direction as the treatment effect, overall, there is a masking of the treatment effect by change score. This is a consequence of the way change is computed.

For example, if  $\bar{Z}_T, \bar{Y}_T$  represents the baseline and outcome score for the treatment group and  $(\bar{Z}_C, \bar{Y}_C)$  represents the baseline and outcome for the control group.

Thus, with change (C) given as;

C=baseline–outcome,

for an absolute imbalance of 0.09 in the same direction as an effect size of (-0.2), the arrangement would be (note that reduction implies treatment effect and imbalance in same direction as treatment implies the treated group has a better prognosis at baseline):

$$CSA_{\text{treatment effect}} = (\bar{Z}_T - \bar{Y}_T) - (\bar{Z}_C - \bar{Y}_C) - 0.09 - (-0.2) - (0 - 0) = -0.09 + 0.2 = 0.11$$

Whereas, if the imbalance of 0.09 is in the opposite direction of effect size of (-0.2), then:

$$CSA_{\text{treatment effect}} = 0 - (-0.2) - (-0.09) - 0 = 0.29 + 0.09 = 0.29$$

These arrangements explain three points;

1. CSA yields estimates of effect in the opposite direction to the effect (improvement) to be determined. This is the reason for the positive sign on the estimate of effect that is expected to be negative.
2. Change score assumes the baseline-outcome correlation to be 1. Thus, estimates of effect are the same across all levels of correlation.
3. Summarily, the computation of effect by CSA when imbalance is in the same direction as treatment effect is such that the magnitude of this imbalance is subtracted from the absolute value of the treatment group's effect. On the other hand, when imbalance is in the opposite direction of treatment, the computation of effect by CSA is such that the magnitude of imbalance is added to the absolute value of the treatment group's effect.

When imbalance is in the same direction as the treatment effect, the estimate from CSA is seen to converge to a zero value,

indicating no effect. This phenomenon ultimately depends on the size of imbalance; the larger the imbalance the closer to zero is the estimate of effect by CSA. This tapering of effect size relative to size of imbalance is due to the deduction of the size of the imbalance from the treatment effect in the treatment group resulting in the loss of some effect. This then means that though some treatment effects exist, they will not be detected by CSA and thus, false negatives will result. Therefore, depending on direction, the larger the imbalance the larger the exaggerating or masking effect by CSA on its estimate of effect.

### Analysis of Covariance

Analysis of covariance is a statistical technique that makes use of the distribution of baseline scores and disparity in this between treatment groups to explain the overall treatment effect. ANCOVA conspicuously features baseline score as a covariate in its model equation and thus accounts for the imbalance during the analysis. Thus, because the model incorporates additional information, there is already an expectation of efficiency in the estimation of the effect. This extra or ancillary information accounts for the reduction in residual variance by ANCOVA.

Similar to other authors on this subject, Van Breukelen<sup>22</sup> presents ANCOVA models as;

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 Z_{ij} + \epsilon_{ij}$$

equivalently as,

$$Y_{ij} - \beta_2 Z_{ij} = \beta_0 + \beta_1 G_{ij} + \epsilon_{ij} \dots \dots \dots (1.12)$$

This, though, presents the method as removing all the effect of the covariate from the outcome. However, Rutherford<sup>29</sup> argues that outcome variables are not adjusted to completely remove the effect of the covariate but rather, adjustment is done such that all patients obtain a covariate score equal to the general covariate mean. In other words, ANCOVA uses the general covariate score to equalize the covariate distribution in the treatment groups. Thus, if a treatment group has a group mean at baseline that is greater than the grand or general covariate mean, the average treatment outcome for that group is adjusted downward. On the other hand, if a group has a mean score at baseline that is lower than the grand mean, then, the group average treatment outcome will be adjusted upward. The issue here is more of semantic (language) than concept. When ANCOVA equalizes the covariate distribution in the treatment groups by using the grand covariate mean, baseline imbalance is inevitably removed and thus offers a platform for a justifiable comparison of groups' treatment effect.

Thus, Rutherford expresses the ANCOVA model following adjustment as:

$$Y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 (Z_{ij} - \bar{Z}) + \epsilon_{ij} \dots \dots \dots (1.13)$$

Equivalently as;

$$Y_{ij} - \beta_2(Z_{ij} - \bar{Z}) = \beta_0 + \beta_1 G_{ij} + \epsilon_{ij} \dots \dots \dots (1.14)$$

$\beta_2$ , represents the degree of linear relationship between the covariate and the outcome and is empirically determined from the data - again, in ordinary language, this represents the portion of the post treatment outcome that is explained by the baseline difference. This must be separated from the main effect otherwise it biases the estimate of effect.  $\bar{Z}$  represents the grand covariate mean (average of all the baseline score).

And following certain algebraic processes, deliberately skipped in this paper, the adjusted estimate of effect  $\lambda$  as by ANCOVA is given as:

$$\lambda = (\bar{Y}_T - \bar{Y}_C) - \hat{\beta}(Z_T - Z_C) \dots \dots \dots (1.14)$$

Only ANCOVA yields an unbiased estimate of effect (with respect to a covariate) when baseline imbalance in the prognostic baseline variable is accounted for.

This then suggests that the estimate of treatment effect by ANCOVA approximates that of ANOVA if the mean baseline score for the two groups is similar. Alternatively, both analyses are equal if ( $\rho=0$ ) irrespective of the size and direction of imbalance. If, however, the baseline score for the control group is greater than the baseline score for the treated group in absolute value, then the overall treatment effect by ANCOVA is expected to be greater than that of the ANOVA (also in absolute value). Similarly, if the treatment group is higher in baseline absolute score than the control group, then the overall treatment effect by ANCOVA will be smaller in absolute score compared to that from ANOVA of post treatment scores. There is directionality in how treatment outcome is affected by covariate imbalance between the treatment groups. For example, if the treated group has a lower mean value at baseline and reduction in baseline score implies that treatment is effective, then the unadjusted treatment effect will fail to identify the possible exaggeration on the overall treatment effect. The overall treatment effect will not reflect the undue advantage of a better prognosis that the treated group had at baseline (that is, when the effect is in the same direction as the baseline imbalance). Conversely, if a lower mean value is recorded at baseline for the control group, then the masking effect on the overall treatment effect will still not be identified and the unadjusted analysis will yield an overall under-estimate of effect (analysis being carried out as if baseline prognosis of both treatment groups is the same). Thus, whether imbalance is in the same direction as treatment or opposite the crude unadjusted analysis will give the same (biased) estimate of effect. With respect to the direction of baseline imbalance, change score analysis will yield an exaggerated treatment effect when baseline imbalance is in the opposite direction of the treatment, that is, the control group has a better prognostic status (lower baseline score) than the treated group. The overall treatment effect however, will be masked by using change score analysis if the imbalance is in the same direction as treatment.

This situation may be overcome by ANCOVA accounting for the imbalance at baseline, thus reducing the systematic

variation in the interests of a less biased and more precise estimate of treatment effect. ANCOVA does not crudely compare the treatment groups' outcomes, but first adjusts the outcomes in relation to the covariate level in the groups. Thus, the procedure of covariate adjustment by ANCOVA, as explained<sup>29</sup> usually involves two stages: 1) ANCOVA determines the co-variation between the covariate(s) and the outcome variable, that is, the influence that the group imbalance has on the treatment outcome for that group, and 2) it removes that variance associated with the covariates from the outcome variable scores (adjusts in a way that the covariate mean value is made equal between the groups). These two stages occur prior to determining whether there is difference in outcome. So, essentially ANCOVA compares two adjusted outcome values. In their study,<sup>30</sup> observe that the precision of the adjusted estimate of treatment effect increases as a function of the correlation between the response variable and the covariate. This implies that as correlation between the covariate and outcome variable increases, the precision of the estimate by ANCOVA also increases.<sup>23,24</sup>

**CONCLUSION**

Covariate imbalance is a real phenomenon in randomized controlled trials and its potentially capable of distorting estimate of treatment effect. Design methods at balancing covariates between groups are not without their flaws. It remains unethical for researchers to deliberately increase sample size in a controlled experiment in which the true effect of the compound has not been ascertained or when there is an indication that the drug being tested may still have some side effects. In the event either stratified randomization or minimization is used, stratification or minimization factors are to be treated as covariates during statistical analysis. The direction and size of baseline imbalance have profound effect on treatment effect estimate by CSA. Only ANCOVA yields unbiased estimate of effect and is recommended at all trial scenarios in which there are concerns about the distribution of prognostic variables between treatment groups.

**ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

Not applicable.

**DATA AVAILABILITY**

Not applicable.

**FUNDING STATEMENT**

Not funded.

**CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest.

**REFERENCES**

1. Altman DG, Doré CJ. Randomisation and baseline compar-

- isons in clinical trials. *Lancet*. 1990; 335(8682): 149-153. doi: [10.1016/0140-6736\(90\)90014-v](https://doi.org/10.1016/0140-6736(90)90014-v)
2. Tu D, Shalay K, Pater J. Adjustment of treatment effect for covariates in clinical trials: Statistical and Regulatory issues. *Drug Inf J*. 2000; 34: 511-523. doi: [10.1177/009286150003400220](https://doi.org/10.1177/009286150003400220)
  3. Hewitt CE, Torgerson DJ. Is restricted randomisation necessary? *BMJ*. 2000; 332(7556): 1506-1508. doi: [10.1136/bmj.332.7556.1506](https://doi.org/10.1136/bmj.332.7556.1506)
  4. Rosenberger WF, Sverdlov O. Handling covariates in the design of clinical trials. *Stat Sci*. 2008; 23(3): 404-419. doi: [10.1214/08-STS269](https://doi.org/10.1214/08-STS269)
  5. Kernan WN, Makuch RM. Letter. *J Clin Epidemiol*. 2001; 54: 104-107.
  6. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. A review. *Control Clin Trials*. 2002; 23(6): 662-674. doi: [10.1016/s0197-2456\(02\)00242-8](https://doi.org/10.1016/s0197-2456(02)00242-8)
  7. Hagino A, Hamada C, Yoshimura I, Ohashi Y, Sakamoto J, Nakazato H. Statistical comparison of random allocation methods in cancer clinical trials. *Control Clin Trials*. 2004; 25(6): 572-584. doi: [10.1016/j.cct.2004.08.004](https://doi.org/10.1016/j.cct.2004.08.004)
  8. Taves DR. Faulty assumptions in Atkinson's criteria for clinical trial design. *J R Stat Soc*. 2004; 167(1): 179-181.
  9. Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol*. 2004; 57(5): 454-460. doi: [10.1016/j.jclinepi.2003.09.014](https://doi.org/10.1016/j.jclinepi.2003.09.014)
  10. Moore KL, Vanderlaan MJ. Covariates adjustment in randomised trials with binary outcomes: Targeted maximum likelihood estimation. Web site. <http://www.bepress.com/ucbbiostat/paper215/>. Accessed September 30, 2021.
  11. Kent DM, Trikalinos TA, Hill MD. Are unadjusted analyses of clinical trials inappropriately biased toward the null? *Stroke*. 2009; 40(3): 672-673. doi: [10.1161/STROKEAHA.108.532051](https://doi.org/10.1161/STROKEAHA.108.532051)
  12. Sacks FM, Bray GA, Carey VJ, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med*. 2009; 360(9): 859-873. doi: [10.1056/NEJMoa0804748](https://doi.org/10.1056/NEJMoa0804748)
  13. Wieder HA, Beer AJ, Lordick F, et al. Comparison of changes in tumor metabolic activity and tumor size during chemotherapy of adenocarcinomas of the esophagogastric junction. *J Nucl Med*. 2005; 46(12): 2029-2034.
  14. Lachin JM. Statistical properties of randomisation in clinical trials. *Control Clin Trials*. 1988; 9(4): 289-311. doi: [10.1016/0197-2456\(88\)90045-1](https://doi.org/10.1016/0197-2456(88)90045-1)
  15. Matthews JNS. *An Introduction to randomised Controlled Clinical Trials*. Arnold, London: Chapman and Hall/CRC; 2000.
  16. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet*. 2002; 359(9300): 57-61. doi: [10.1016/S0140-6736\(02\)07283-5](https://doi.org/10.1016/S0140-6736(02)07283-5)
  17. Altman DG, Bland JM. How to randomise. *BMJ*. 1999; 319(7211): 703-704. doi: [10.1136/bmj.319.7211.703](https://doi.org/10.1136/bmj.319.7211.703)
  18. Roberts C, Torgerson D. Understanding controlled trials: Randomisation methods in controlled trials. *BMJ*. 1998; 317: 1301. doi: [10.1136/bmj.317.7168.1301](https://doi.org/10.1136/bmj.317.7168.1301)
  19. Egbewale BE. Random allocation in controlled clinical trials: A review. *J Pharm Pharm Sci*. 2014; 17(2): 248-253. doi: [10.18433/j3sw36](https://doi.org/10.18433/j3sw36)
  20. Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Train*. 2008; 43(2): 215-221. doi: [10.4085/1062-6050-43.2.215](https://doi.org/10.4085/1062-6050-43.2.215)
  21. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Med Res Methodol*. 2001; 1: 6. doi: [10.1186/1471-2288-1-6](https://doi.org/10.1186/1471-2288-1-6)
  22. Van Breukelen GJP. ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *J Clin Epidemiol*. 2006; 59: 920-925. doi: [10.1016/j.jclinepi.2006.02.007](https://doi.org/10.1016/j.jclinepi.2006.02.007)
  23. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: A simulation study. *BBMC Med Res Methodol*. 2014; 14: 49. doi: [10.1186/1471-2288-14-49](https://doi.org/10.1186/1471-2288-14-49)
  24. Egbewale BE. Differences in sample size requirements of statistical methods involved in clinical trials with baseline imbalance demonstrated and quantified: a simulation study. *J Clin Trials*. 2015; 5: 225. doi: [10.4172/2167-0870.1000225](https://doi.org/10.4172/2167-0870.1000225)
  25. Fleiss JL. *The Design and Analysis of Clinical Experiments*. New York, USA: John Wiley; 1986.
  26. Camilli G, Shepard LA. The inadequacy of ANOVA for detecting test bias. *J Educ Stat*. 1987; 12(1): 87-99. doi: [10.3102/10769986012001087](https://doi.org/10.3102/10769986012001087)
  27. Christensen E, Neuberger J, Crowe J, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cir-

- rhosis. Final results of an international trial. *Gastroenterology*. 1985; 89(5): 1084-1091. doi: [10.1016/0016-5085\(85\)90213-6](https://doi.org/10.1016/0016-5085(85)90213-6)
28. Piantadosi S. *Clinical Trials: A Methodological Perspective*. NJ, USA: John Wiley & Sons; 1997: 302.
29. Rutherford A. *Introducing ANOVA and ANCOVA: A GLM Approach*. London, England: Sage Publications; 2001.
30. Wang SJ, Hung HM. Adaptive covariate adjustment in clinical trials. *J Biopharm Stat*. 2005; 15(4): 605-611. doi: [10.1081/BIP-200062280](https://doi.org/10.1081/BIP-200062280)