

**Statistical Issues in Randomized Controlled Trials: an editorial**Umesh Wadgave¹, Mahesh Ravindra Khairnar¹, Yogesh Wadgave²

¹ Assistant Professor, Department of Public Health Dentistry, Bharati Vidyapeeth Deemed University Dental College and Hospital, Sangli, Maharashtra, India

² Special Medical Officer, Department of Community Medicine, Seth G.S. Medical College, Mumbai, Maharashtra, India

Type of article: Editorial

Abstract

Randomization is the bedrock of randomized controlled trials, which ensures the elimination of selection bias and also to some extent the homogenous distribution of covariates between the intervention arms. Randomization does not always guarantee the baseline balance, and hence makes the statistical analysis more complex. Several published clinical trials have employed test of significance to compare baseline measures between the groups. However, such practice has been criticized by several authors and CONSORT statement also discourages it. This overview discusses various statistical designs that were employed in published trials. Post intervention data (follow up score) comparison between the arms was common practice in published RCTs. However, this approach fails to adjust baseline imbalance. Both Change score and Percentage change methods adjust the baseline imbalance. Both of the approaches give precise estimates when there is a high correlation between baseline and follow-up score. However, when correlation is low they both give biased and less precise estimates of treatment effect. Analysis of covariance (ANCOVA) is a regression method, which maintains high statistical power and gives less biased and more precise estimates of treatment effect regardless of correlation level. Understanding strengths and limitations of different statistical designs of RCTs will prevent statistical errors, which can yield an accurate estimate of treatment effect.

Keywords: Biostatistics, Randomized controlled trials, Covariate adjustment, Analysis of covariance

Note:

This is a solicited editorial. Because of the increasing number of Randomized Controlled Trials (RCTs) submitted to Electronic Physician, the journal has invited Associate Professor Dr. Umesh Wadgave, who is an expert in statistical analysis of RCTs, and his team, to review and discuss on the timely topic of statistical errors in this field. Dr. Umesh Wadgave is an Associate Editor of Electronic Physician.

1. Introduction

Randomized controlled trials (RCTs) generate the least biased estimate of intervention and hold the highest position among the primary studies in the level of evidence pyramid (1). Randomization is the bedrock of RCTs, which removes the selection bias and lowers the probability of baseline differences between the intervention arms. However, on few occasions, randomization fails to equally distribute some important covariates, which can influence the results (2). So, it is essential to identify such differences and adjust them during the statistical analysis (2, 3). Adjusting the baseline differences (Covariate adjustment) increases the power of the study, and can yield an accurate estimate of treatment effect (3-6). This overview critically evaluates different statistical designs employed

Corresponding author:

Assistant Professor Dr. Umesh Wadgave, Department of Public Health Dentistry, Bharati Vidyapeeth Deemed University Dental College and Hospital, Walnesswadi, Sangli, Maharashtra, India (Pin: 416416).

Tel: +91 8805835602, Email: dr.w.umesh@gmail.com

Received: September 12, 2018, Accepted: October 03, 2018, Published: November 17, 2018

iThenticate screening: October 03, 2018, English editing: October 08, 2018, Quality control: October 22, 2018

This article has been reviewed / commented by three experts

© 2018 The Authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

in RCTs and attempts to suggest the best methods for future researchers. Box 1 includes the important highlights of the article, and Table 1 presents the main terminologies used in this context.

Box 1. Important highlights of the review

<ul style="list-style-type: none"> ❖ Highlights the limitation of baseline comparison using inferential statistics in randomized controlled trials. ❖ Discusses the influence of different statistical designs employed in RCTs on covariate adjustment. ❖ Justifies the utility of Analysis of Covariance (ANCOVA) statistical method over other approaches. ❖ To the best of our literature search, this is the only article that discusses most of the possible statistical designs employed in RCTs. ❖ This article helps researchers in understanding the limitations and strengths of the different statistical designs of RCTs.

Table 1. Important terminologies

Term	Definition	Ref. no.
Randomization	Ensures that research participants are divided into groups by chance (random allocation), like in a lottery. It eliminates selection bias and minimizes the differences between the intervention groups.	7
Regression to mean	A statistical phenomenon that can make natural variation in repeated data look like real change. It can result in wrongly concluding that an effect is due to treatment when it is due to chance.	8
Covariates	Any variable that is measurable and considered to have a statistical relationship with the dependent variable would qualify as a potential covariate. A covariate is thus a possible predictive or explanatory variable of the dependent variable.	9
Covariate adjustment	The statistical procedure of eliminating the influence of baseline imbalance of covariates that are correlated to outcome, leading to a more precise estimate of treatment effect.	6
Significance tests	Statistical tests that indicate whether observed differences between assessment results occur because of sampling error or by chance.	10

2. Baseline comparison

In RCTs, comparability of intervention arms at baseline in every aspect except the intervention they receive is essential for detecting the true estimates of any intervention (11). Theoretically, randomization procedure distributes both the known and unknown covariates equally between the intervention arms. But in reality, the distribution of covariates may not always be homogenous after randomization (10, 12). It has been observed that researchers employ the significance tests to check whether the heterogeneity in covariates distribution after randomization is statistically significant or not (13-15). This practice has been criticized by many scholars since any imbalance after randomization simply occurs because of chance, and hence it is futile to apply significance tests (13, 16, 17). Consolidated Standards of Reporting Trials (CONSORT) statement also discourages this practice, stating “tests of baseline differences are not necessarily wrong, just illogical. Such hypothesis testing is superfluous and can mislead investigators and their readers” (18). However, CONSORT Statement recommends reporting of a table describing baseline demographic and clinical characteristics of each intervention arm (18). Altman stated that “performing a significance test to compare baseline variables is to assess the probability of something having occurred by chance when we know that it did occur by chance” (16).

Significance tests for baseline comparison have been used to assess the success of randomization (17). However, even after a successful randomization, we can expect 5% of variables to differ significantly between the intervention arms at baseline (19, 20). Therefore, the success or the failure of randomization cannot be judged based on presence or absence of statistically significant differences. It is advisable to check for meaningful differences rather than statistically significant differences (17, 19). Meaningful differences are those differences at baseline that can impact [confound] the results of a trial. They can be identified with sound clinical knowledge, research evidence, and common sense (16). These meaningful differences should be adjusted during the statistical analysis.

3. Statistical adjustment of baseline imbalances

Randomization is not the guaranteed method for homogeneous distribution of both observed and unobserved covariates among the interventional arms. As discussed above, any baseline imbalances should be identified and adjusted during the statistical analysis. In every clinical trial, researchers collect data at baseline and post-intervention in all the arms and subject it to statistical analysis. There are five different techniques of data analysis practiced in RCTs, namely (Table 2):

- 1) Analyzing the difference between baseline and post intervention
- 2) Analyzing post-intervention data
- 3) Analyzing the change scores between post-intervention and baseline
- 4) Analyzing the percentage change between baseline and post intervention
- 5) Analysis of covariance (ANCOVA)

Table 2. Description of different statistical designs used in randomized controlled trials

		Intervention A	Intervention B
	Pre score	A ₁	B ₁
	Post score	A ₂	B ₂
Designs	Formula	Evaluates	Recommendation
Baseline Comparison	A ₁ vs. B ₁	Statistical significant difference between the groups at baseline.	Not Recommended: In RCT, baseline differences are because of chance.
Difference between baseline and post data	A ₁ vs. A ₂ and B ₁ vs. B ₂	Statistical significant change from baseline to follow up score, separately within each group.	Not Recommended: Only compares within a group and not between groups.
Post-intervention data comparison	A ₂ vs. B ₂	Statistical significant difference in follow up score between the groups.	Not Recommended: Does not adjust baseline imbalances.
Change scores	(A ₁ - A ₂) vs. (B ₁ - B ₂)	Statistical significant difference in change score (difference between pre and post score) data between the groups.	Not Recommended: Fails to adjust the baseline imbalances because of regression to mean. Dependent on correlation between baseline and follow up score.
Percent change	$[(A_2 - A_1) * 100 / A_1]$ vs. $[(B_2 - B_1) * 100 / B_1]$	Statistical significant difference in percent change score (percent difference between pre and post score) between the groups.	Not Recommended: Fails to adjust the baseline imbalances because of regression to mean. Dependent on correlation between baseline and follow up score.
ANCOVA	A ₂ vs. B ₂ [Adjust A ₁ and B ₁]	Statistical significant difference in post intervention data between the groups while adjusting the baseline scores.	Recommended: Adjusts baseline imbalances. Independent of correlation between baseline and follow up score. Has more statistical power. Applied only when an outcome is continuous data.

3.1. Analyzing the difference between baseline and post intervention

This method assesses the statistically significant change from baseline (i.e. Pre score versus Post score), separately in each group (21). Then it is stated that statistically significant difference was observed only in one group but not in the other, and concludes that treatments are different (21, 22). This tests the null hypothesis that the change from the baseline is zero, separately in each intervention arm (22). Conclusions derived from this method are not appropriate because of natural changes over time and regression towards the mean (23). This technique is to no purpose and illogical because our primary intention in the randomized trial is not to assess the change from the baseline (within group comparison), but to estimate whether any change is greater in one group than the other (between group comparison) (21). Bland and Altman criticized this method stating “It is conceptually wrong, statistically invalid,

and consequently highly misleading” (24). Two surveys in 1987 and 2011, reported that about 10% of trials of reputed medical journals analyzed within-group comparisons only (25, 26).

3.2. Analyzing post-intervention data (POST method)

This method ignores the baseline values in all the intervention arms and just compares the post intervention data between the groups. This method is simple to understand and less time consuming for data analysis (27). Analysis of variance or student’s t test is used to compare the follow-up scores between the groups. This technique is based on the assumption that randomization ensures baseline homogeneity of all characteristics across all arms and thus covariate adjustment is deemed unnecessary (27). However, we know that randomization does not always equally distribute all the variables in intervention arms, and simulation studies have shown that after baseline imbalance adjustment, the analysis of post intervention scores can be different (28). As it fails to adjust any baseline imbalances between the arms, this can lead to an unfair advantage to one intervention arm over another (29). This method can give biased and imprecise estimates of treatment effect.

3.3. Analyzing the change scores between post-intervention and baseline

This technique, known as change score or gain score (i.e. Pre score minus Post score), subtracts the baseline scores from the post-intervention score within each arm and then change score of each arm is compared with statistics (27). The advantage of this method is that baseline scores are accounted in the analysis (27). Although this technique takes baseline scores into account, still it fails to adjust the baseline imbalance because of regression to mean (23, 30). Regression to mean concept explains that patients with low baseline score generally tend to improve more so than patients with high baseline scores (23, 30). Statistically significant results are more likely to be seen in change score when correlation is high between baseline and follow up score, while in post-intervention comparison, significant results are seen when the correlation is low (30). However, it is important to specify the statistical design in protocol and it is inappropriate to choose statistical test based on the significant result. Application of this test is also conditional on the correlation between baseline and follow-up score, and hence not recommended.

3.4. Analyzing the percentage change between baseline and post intervention (Percent Change)

In contrast to the change score which assesses the absolute change, few researchers recommend the assessment of percentage change between the baseline and follow up scores (31, 32). Stridbeck R et al. explained the advantage of percentage change over change score with a simple example “Two obese men A and B participate in a weight loss program. Their weights at the beginning of the program are 150 Kg and 100 Kg, respectively. When they finish the program, man A, who weighs 150 Kg has lost 15 Kg, while man B has lost 10 Kg. From the example, we see that, man A lost 5 Kg more than that man B, but the percent of weight they lost are 10% in both cases” (32). Similar to change score, this method also fails to adjust the baseline imbalances because of regression to mean (23). Another disadvantage is that it is calculated using the ratio, so this method generates non-normal distributed data from normally distributed data (33). Vickers completely discourages the use of percentage change and recommends the use of ANCOVA, and when conditions where ANCOVA cannot be employed, change score or post-intervention comparison should be considered (33).

3.5. Analysis of covariance (ANCOVA)

Beholding the limitations of the above-mentioned methods’ analysis of covariance, a regression method is a better approach for statistical analysis for RCTs (27, 30). ANCOVA uses baseline scores as covariates in the analysis and it is applied when an outcome is continuous. In RCTs, ANCOVA is unaffected by baseline imbalances and has more statistical power than others in detecting treatment effects (22, 33). Vickers and Altman demonstrated through a simulation study that ANCOVA maintained its statistical power even at low correlation level between baseline and follow up scores (30). However, statistical power of change score and percent change score decreased with the dropping correlation between baseline and follow up score (30). When the correlation is high (>0.8) the efficiency of ANCOVA remains relatively similar to the change score and percentage change score (27, 30). Advantages of ANCOVA are the following; it maintains the high statistical power regardless of the correlation level, it decreases the bias and gives precise estimates by adjusting the baseline scores and also the other prognostic variables, and this regression method also helps in prediction (3, 30).

There is sufficient evidence for opting for the ANCOVA method over others in adjusting the covariates (3, 27, 29, 30). However, lack of an objective tool for assessment of baseline imbalances have made selection of the covariates for adjustment very complex (19). One should note that, adjusting the wrong covariate can lead to increase in bias and decrease in the statistical power (2). Kahan et al. demonstrated with simulation study that adjusting the

prognostic variables and also fewer suspected covariates increased the statistical power of study (29). Before considering ANCOVA, it is important to understand its two underlying assumptions: first, the covariate should be independent of outcome i.e. covariates should not be different across the groups (3, 27); second, homogeneity of regression slopes (27).

4. Conclusions

The information from this overview gives an idea of various statistical designs employed in RCTs and also discusses their strengths and limitations. Comparing baseline differences using significance test should be avoided as it is highly misleading. Do not simply estimate the statistical significance of difference between baseline and post intervention within each randomized group, rather, compare the measures between the groups. In any circumstance, ANCOVA should be the analysis of choice for adjusting the baseline imbalances and prognostic covariates in RCTs.

Acknowledgments:

NA

Conflict of Interest:

There is no conflict of interest to be declared.

Authors' contributions:

All authors contributed to this project and article equally. All authors read and approved the final manuscript.

References:

- 1) Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996; 312(7023): 71–2. doi: 10.1136/bmj.312.7023.71. PMID: 8555924, PMCID: PMC2349778.
- 2) Kent DM, Trikalinos TA, Hill MD. Are unadjusted analyses of clinical trials inappropriately biased toward the null? *Stroke*. 2009; 40(3): 672–3. doi: 10.1161/STROKEAHA.108.532051. PMID: 19164784, PMCID: PMC2693723.
- 3) 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. *BMC Med Res Methodol*. 2014; 14(1). doi: 10.1186/1471-2288-14-49. PMID: 24712304, PMCID: PMC3986434.
- 4) Gray LJ, Bath PMW, Collier T. Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke* [Internet]. 2009;40(3):888–94. doi: 10.1161/STROKEAHA.108.519207. PMID: 19164798.
- 5) Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials*. 1998; 19(3): 249–56. doi: 10.1016/S0197-2456(97)00147-5.
- 6) Hernández AV, Steyerberg EW, Habbema JDF. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol*. 2004; 57(5): 454–60. doi: 10.1016/j.jclinepi.2003.09.014. PMID: 15196615.
- 7) Random Allocation - National Library of Medicine - PubMed Health. Available from: <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0025810/>.
- 8) Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: What it is and how to deal with it. *Int J Epidemiol*. 2005; 34(1): 215–20. doi: 10.1093/ije/dyh299. PMID: 15333621.
- 9) Covariate. In: *Encyclopedia of Research Design*. Thousand Oaks California 91320 United States: SAGE Publications; 2018.
- 10) Senn S. Testing for baseline balance in clinical trials. *Stat Med*. 1994; 13(17): 1715–26. doi: 10.1002/sim.4780131703. PMID: 7997705.
- 11) Elkins MR. Assessing baseline comparability in randomised trials. *J Physiother*. 2015; 61(4): 228-30. doi: 10.1016/j.jphys.2015.07.005. PMID: 26364088.
- 12) Abel U, Koch A. The role of randomization in clinical studies: Myths and beliefs. *J Clin Epidemiol*. 1999; 52(6): 487-97. PMID: 10408986.
- 13) Peterson RL, Tran M, Koffel J, Stovitz SD. Statistical testing of baseline differences in sports medicine RCTs: a systematic evaluation. *BMJ open Sport Exerc Med*. 2017; 3(1): e000228.
- 14) Knol MJ, Groenwold RHH, Grobbee DE. P-values in baseline tables of randomised controlled trials are inappropriate but still common in high impact journals. *Eur J Prev Cardiol*. 2012; 19(2): 231–2. PMID: 22512015.

- 15) Austin PC, Manca A, Zwarenstein M, Juurlink DN, Stanbrook MB. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *J Clin Epidemiol.* 2010; 63(2): 142–53. doi: 10.1016/j.jclinepi.2009.06.002. PMID: 19716262.
- 16) Altman DG. Comparability of Randomised Groups. *Stat.* 1985; 34(1): 125.
- 17) de Boer MR, Waterlander WE, Kuijper L, Steenhuis I, Twisk J. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *Int J Behav Nutr Phys Act.* 2015; 12(1): 4. doi: 10.1186/s12966-015-0162-z. PMID: 25616598, PMCID: PMC4310023.
- 18) Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Int J Surg.* 2012; 10(1): 28–55. doi: 10.1016/j.ijsu.2011.10.001. PMID: 22036893.
- 19) Ciolino JD, Martin RH, Zhao W, Hill MD, Jauch EC, Palesch YY. Measuring continuous baseline covariate imbalances in clinical trial data. *Stat Methods Med Res.* 2015; 24(2): 255–72. doi: 10.1177/0962280211416038. PMID: 21865270, PMCID: PMC4280338.
- 20) Senn S. Seven myths of randomisation in clinical trials. *Stat Med.* 2013; 32(9): 1439–50. doi: 10.1002/sim.5713. PMID: 23255195.
- 21) Bland JM, Altman DG. Comparisons within randomised groups can be very misleading. *BMJ.* 2011; 342. doi: 10.1136/bmj.d561.
- 22) Nash R, Bunce C, Freemantle N, Doré CJ, Rogers CA. Ophthalmic statistics note 4: Analysing data from randomised controlled trials with baseline and follow-up measurements. *Br J Ophthalmol.* 2014; 98(11): 1467–9. doi: 10.1136/bjophthalmol-2014-305614. PMID: 25107901, PMCID: PMC4215292.
- 23) Bland JM, Altman DG. Regression towards the mean. *BMJ.* 1994; 308(6942): 1499. doi: 10.1136/bmj.308.6942.1499. PMID: 8019287, PMCID: PMC2540330.
- 24) Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials.* 2011; 12: 264. doi: 10.1186/1745-6215-12-264. PMID: 22192231, PMCID: PMC3286439.
- 25) Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet.* 1990; 335(8682): 149–53. doi: 10.1016/0140-6736(90)90014-V.
- 26) Vollenweider D, Boyd CM, Puhan MA. High prevalence of potential biases threatens the interpretation of trials in patients with chronic disease. *BMC Med.* 2011; 9. doi: 10.1186/1741-7015-9-73.
- 27) Zhang S, Paul J, Nantha-Aree M, Buckley N, Shahzad U, Cheng J, et al. Empirical comparison of four baseline covariate adjustment methods in analysis of continuous outcomes in randomized controlled trials. *Clin Epidemiol.* 2014; 6(1): 227–35. doi: 10.2147/CLEP.S56554. PMID: 25053894, PMCID: PMC4105274.
- 28) Overall JE, Magee KN. Directional Baseline Differences And Type I Error Probabilities In Randomized Clinical Trials. *J Biopharm Stat.* 1992; 2(2): 189–203. doi: 10.1080/10543409208835039. PMID: 1300213.
- 29) Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: An assessment of 12 outcomes from 8 studies. *Trials.* 2014; 15(1). doi: 10.1186/1745-6215-15-139.
- 30) Vickers AJ, Altman DG. Statistics Notes: Analysing controlled trials with baseline and follow up measurements. *BMJ.* 2001; 323(7321): 1123–4. doi: 10.1136/bmj.323.7321.1123. PMID: 11701584, PMCID: PMC1121605.
- 31) Törnqvist L, Vartia P, Vartia YO. How should relative changes be measured? *Am Stat.* 1985; 39(1): 43–6.
- 32) Stridbeck R, Zhang L, Han K. How to Analyze Change from Baseline: Absolute or Percentage Change? How to Analyze Change from Baseline: Absolute or Percentage Change? *Högskolan Dalarna.* 2009; 1 – 17.
- 33) 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: 1–4. doi: 10.1186/1471-2288-1-6. PMID: 11459516, PMCID: PMC34605.