

METHODOLOGY

Open Access



# The correlation between baseline score and post-intervention score, and its implications for statistical analysis

Lei Clifton<sup>1</sup>  and David A. Clifton<sup>2\*</sup>

## Abstract

**Background:** When using a continuous outcome measure in a randomised controlled trial (RCT), the baseline score should be measured in addition to the post-intervention score, and it should be analysed using the appropriate statistical analysis.

**Methods:** We derive the correlation between the change score and baseline score and show that there is always a correlation (usually negative) between the change score and baseline score. We discuss the following correlations and provide the mathematical derivations in the Appendix:

- Correlation between change score and baseline score
- Correlation between change score and post score
- Correlation between change score and average score.

The setting here is a parallel, two-arm RCT, but the method discussed in this paper is applicable for any studies or trials that have a continuous outcome measure; it is not restricted to RCTs.

**Results:** We show that using the change score as the outcome measure does not address the problem of regression to the mean, nor does it take account of the baseline imbalance. Whether the outcome is change score or post score, one should always adjust for baseline using analysis of covariance (ANCOVA); otherwise, the estimated treat effect may be biased. We show that these correlations also apply when comparing two measurement methods using Bland-Altman plots.

**Conclusions:** The correlation between baseline and post-intervention scores can be derived using the variance sum law. We can then use the derived correlation to calculate the required sample size in the design stage. Baseline imbalance may occur in RCTs, and ANCOVA should be used to adjust for baseline in the analysis stage.

**Keywords:** Correlation, Baseline, Change score, Post-intervention, Treatment, Statistical analysis, Sample size, Independent, Means, Standard error (SE), Standard deviation (SD), Regression to the mean (RTM), Outcome, Analysis of covariance (ANCOVA), Randomised controlled trial (RCT), Bland-Altman plot

\* Correspondence: [davidc@robots.ox.ac.uk](mailto:davidc@robots.ox.ac.uk)

<sup>2</sup>Institute of Biomedical Engineering (IBME), Department of Engineering Science, University of Oxford, Oxford, UK

Full list of author information is available at the end of the article



## Background

When using a continuous outcome measure in a randomised controlled trial (RCT), the baseline score should be measured in addition to the post-intervention score. In a previous paper [9], we have shown how to derive  $r$ , the correlation between baseline score and post score. In this paper, we derive correlations between different variables in the Appendix, assuming  $r$  is known. There are two options for the outcome measure: change score or post score. We examine the validity of using change score as the outcome measure in the “Method” section, and then further discuss the applications of our methods outlined in the “Discussion” section.

## Method

### The choice of outcome measure: post score vs. change score

The methods outlined in this paper are suitable for any continuous measure, and therefore we have used a generic notation. Suppose the primary continuous outcome measure is  $Y$ , with  $Y_0$  and  $Y_1$  denoting the value of  $Y$  at baseline and post-intervention, respectively. Let  $r$  denote the correlation coefficient between  $Y_0$  and  $Y_1$ .

We will call  $Y_1$  “post score”,  $Y_0$  “baseline score”, and  $(Y_1 - Y_0)$  “change score”. We note that  $Y_1$  is also called “follow up score” in [17]. The authors assumed there is no interaction between baseline and intervention group, and we make the same assumption in this paper.

In analysis of covariance (ANCOVA), we estimate parameters  $a$ ,  $b$ , and  $c$  in the following regression question:

$$\text{post score} = a + b \times \text{baseline} + c \times \text{group}$$

where “group” stands for “intervention group”. One is usually most interested in the estimate of  $c$ , the treatment effect, in an RCT. Substituting  $Y_1$  for “post score”,  $Y_0$  for “baseline score”, and  $G$  for “group”, we have the following regression equation for a standard ANCOVA that uses post score as outcome:

$$Y_1 = a + bY_0 + cG \quad (1)$$

Rearranging Eq. 1, we have

$$Y_1 - Y_0 = a + (b-1)Y_0 + cG \quad (2)$$

Equation 2 is ANCOVA using change score ( $Y_1 - Y_0$ ) as outcome and adjusting for baseline  $Y_0$ . Compared with the standard ANCOVA in Eq. 1, where post score  $Y_1$  is the outcome, nothing has changed except that the regression coefficient for  $Y_0$  has decreased by 1. The significance level and the width of the confidence intervals for all estimated regression coefficients remain the same as those in a standard ANCOVA. Further mathematical details can be found in [14].

Rearranging Eq. 1 in a different way, we have

$$Y_1 - bY_0 = a + cG \quad (3)$$

Equation 3 shows that using change score as outcome without adjusting for baseline is only equivalent to a standard ANCOVA when  $b = 1$ . In practice, the estimated  $b$  in an ANCOVA is rarely equal to 1; hence, it is only a special case of ANCOVA.

### Regression to the mean (RTM) and ANCOVA

RTM is a well-known statistical phenomenon, first discovered by Galton in [10]. RTM has been discussed by a number of authors, e.g. [3, 6, 7, 15], etc. In this paper, we consider RTM in the context of baseline measures.

If an extreme measure is observed at baseline, then its value is likely to be less extreme in the post-intervention measure, even if the intervention has no effect. In the RCT example in [17], the treatment effect of acupuncture was measured by a 100-point rating score, where lower scores indicate poorer outcomes. Suppose that the baseline scores of the control group reflect the scores of the general population, and that acupuncture has no treatment effect. If, by chance, the baseline scores of the intervention group are lower than the scores of the general population, their post scores will still be higher than their baseline scores, due to RTM. We consider two options of outcome measure:

1. Post score: If post score is positively correlated with the baseline score (which is usually the case in clinical practice), acupuncture will appear to have a negative effect, even though it has no effect; i.e. the treatment effect of acupuncture will be underestimated.
2. Change score: Acupuncture will appear to have a positive effect, even though it has no effect; i.e. the treatment effect of acupuncture will be overestimated.

In both of the preceding scenarios, the appropriate statistical analysis is ANCOVA adjusting for baseline scores. The first scenario corresponds to Eq. 1, with its left-hand side showing the post score as the outcome measure. The second scenario corresponds to Eq. 2, with its left-hand side showing the change score as the outcome measure.

Using change score as the outcome measure does not address the problem of RTM, nor does it take account of the baseline imbalance. Even if change score is deemed to be the appropriate outcome measure after careful consideration, ANCOVA should still be used to adjust for baseline scores, as shown by Eq. 2.

Using change score as outcome does not adjust for baseline imbalance; instead, any imbalance will be reversed due to RTM [4]. Equation 2 shows that when

change score is the chosen outcome, one should still adjust for baseline using ANCOVA. In such a case, ANCOVA is the valid statistical analysis.

ANCOVA has the advantages of being unaffected by baseline imbalance [17], and it has greater statistical power than other methods [16]. An RCT reduces RTM at the design stage, but one should still use ANCOVA to adjust for baseline in the analysis stage [3].

#### The validity of using change score as outcome measure

Let  $r$  denote the correlation between post score  $Y_1$  and baseline score  $Y_0$ . Let  $s_{Y_0}^2$  denote the sample variance of baseline score  $Y_0$ ,  $s_{Y_1}^2$  denote the sample variance of post score  $Y_1$ , and  $s_{(Y_1-Y_0)}^2$  denote the sample variance of the change score  $(Y_1 - Y_0)$ . Let  $s_{Y_0}$ ,  $s_{Y_1}$ , and  $s_{(Y_1-Y_0)}$  denote their corresponding standard deviations (SD).

The Appendix shows that the correlation between change score  $(Y_1 - Y_0)$  and baseline score  $Y_0$  is

$$\text{corr}(Y_1 - Y_0, Y_0) = \frac{r s_{Y_1} - s_{Y_0}}{\sqrt{s_{Y_0}^2 + s_{Y_1}^2 - 2 r s_{Y_0} s_{Y_1}}} \quad (4)$$

Equation 4 shows that  $\text{corr}(Y_1 - Y_0, Y_0)$  will be positive if  $r > s_{Y_1}/s_{Y_0}$ , and vice versa. In the special case of  $r = 0$  (i.e.  $Y_1$  and  $Y_0$  are not correlated), there still will be negative correlation between change score  $(Y_1 - Y_0)$  and baseline score  $Y_0$ . If the post score and baseline have similar variance, then  $\text{corr}(Y_1 - Y_0, Y_0)$  will usually be negative because  $r \leq 1$ . Most importantly, Eq. 4 shows that there is always a correlation between the change score and the baseline score; therefore, one should use ANCOVA to adjust for the baseline score.

Equation 4 was also applied when comparing methods of measurement [8], where  $Y_1$  and  $Y_0$  were replaced by test measure and standard measure, respectively. The authors of [8] show that plotting difference against standard method is misleading, because there will be a negative correlation even if the two methods are not correlated. The authors also conclude that plotting difference against the average is more useful in almost all medical measures.

We now consider the variance of the change score  $(Y_1 - Y_0)$ . The variance sum law states that the variance of the change score is

$$s_{(Y_1-Y_0)}^2 = s_{Y_0}^2 + s_{Y_1}^2 - 2 r s_{Y_0} s_{Y_1} \quad (5)$$

Equation 5 shows that if  $r$  is small,  $s_{(Y_1-Y_0)}^2$  will be greater than  $s_{Y_1}^2$ ; i.e. using change score will add variance compared with using post score as the outcome measure, and therefore will be less likely to show a significant result. Conversely, the post score will be more likely to show a significant result if  $r$  is high. However, the choice

of the outcome measure should not be driven by the likelihood of a significant result; instead, it should be pre-specified in the trial protocol [17].

Using change score as outcome has undesirable implications. For example, if there is a hard lower or upper limit on the score, it may lead to “floor” or “ceiling” effects in change score. If transformation of the original scores is used during data analysis, it is not guaranteed that the transformation applies to the change score. Different transformations can reorder change scores across patients. By contrast, using post scores is always valid and never misleading [12].

The change score can be a reasonable outcome when the correlation between baseline and post scores is high (e.g.  $r > 0.8$ ) in stable chronic conditions such as obesity [17]. In this instance, ANCOVA is still the preferred general approach.

In the Appendix, we show that the change score is always correlated with the baseline score (Eq. 9) and with the post score (Eq. 10). This is purely a statistical artefact, and it exists regardless of whether the treatment is effective or not. The method of using change score as outcome measure is prone to incorrect interpretations of such correlations.

In summary, one should be cautious about using change score as the outcome measure. If justification exists for using change score as the outcome measure, one should still adjust for baseline using ANCOVA. This will increase statistical power and avoid the pitfall of RTM.

## Discussion

### Potential imbalance of baseline in RCTs

In practice, given the finite sample size and random nature of RCTs, any important prognostic factors and baseline score may not be balanced between arms. RCTs of small or moderate sample sizes are particularly prone to such imbalances.

The balance of specific prognostic factors can be achieved during randomisation. The most commonly used randomisation methods are stratified permuted blocks [1] and minimisation [2]. Both methods allow randomisation to be stratified according to prognostic factors, such as gender, disease severity, age group, etc., which ensures that these characteristics are balanced between the treatment and control arms. One should adjust for stratification or minimisation factors during the data analysis stage [13].

However, the randomisation methods outlined above do not deal with the potential imbalance in baseline scores between arms. It is therefore of particular importance to measure the baseline scores *before* randomisation and then use ANCOVA to adjust for baseline in the data analysis stage, as shown in this paper.

**The correlation between change score and baseline score**

The correlation between change score  $(Y_1 - Y_0)$  and baseline score  $Y_0$  has previously been observed in the context of initial blood pressure and its fall with treatment [11]. We provide a detailed mathematical derivation in the Appendix. Equation 9 shows that there is always a correlation between the change score and the baseline, regardless of any treatment effects. This correlation will be negative if  $r$  is small; that is, if the change score is the chosen outcome measure (for instance, of blood pressure), we will observe a fall in the blood pressure against the baseline blood pressure. An incorrect interpretation of such an observed decrease in the change score would be to conclude that the treatment is more effective for patients whose initial blood pressure is high.

**Deriving correlation between baseline score  $Y_0$  and post score  $Y_1$**

In this paper, we have assumed that the value of  $r$ , the correlation between baseline score  $Y_0$  and post score  $Y_1$ , is known. However, the value of  $r$  is usually not readily available in the design stage of an RCT. In a previous paper [9], we have shown how to derive  $r$  using the variance sum law based on a published paper, and then use the derived value of  $r$  to calculate sample size using different methods.

Once we have derived  $r$ , we can derive correlations between different variables using equations derived in the Appendix. In the ideal situation when the raw data are available, one can fully investigate correlations between different variables using the equations provided in this paper.

**Bland-Altman plots**

The same mathematical principles derived in the Appendix can be applied to both choosing outcome measures in an RCT and assessing agreement between two measurement methods [5, 8]. A Bland-Altman plot shows the difference of the two measures on the  $y$ -axis and their average on the  $x$ -axis.

When assessing the agreement between two measurement methods, one should use a Bland-Altman plot showing the difference of the two measures against their average. The correlation  $r$  between the two measures does not assess their agreement [5]. We note that if the ranges of the two measures are different, their variances will be different; therefore, there will be a trend on the Bland-Altman plot, caused by the correlation shown in Eq. 13. Therefore, one should examine the variances of the two measures before using Bland-Altman plots.

Similarly, in the context of outcome measures in an RCT, one can plot the change score against the average of the baseline and post scores, as in a Bland-Altman plot. Equation 13 shows that the correlation between the change score and their average will be zero if the baseline score and post score have equal variance.

**Limitations**

The methods described in this paper only consider continuous variables or outcome measures. They are not applicable to binary variables.

**Appendix**

We denote the variance of variables  $X$  and  $Y$  by  $\sigma_X^2$  and  $\sigma_Y^2$ , respectively, and their correlation  $\rho$ . We use  $corr()$ ,  $cov()$ ,  $var()$  as the generic notation for correlation, covariance, and variance, respectively.

**Correlation between  $(Y - X)$  and  $X$**

In this section, we derive  $corr(Y - X, X)$ , the correlation between variables  $(Y - X)$  and  $X$ .

By definition, we have

$$\begin{aligned} corr(Y - X, X) &= \frac{cov(Y - X, X)}{\sqrt{var(Y - X) \cdot var(X)}} \\ &= \frac{cov(Y, X) - \sigma_X^2}{\sqrt{var(Y - X) \cdot \sigma_X^2}} \end{aligned} \tag{6}$$

By the variance sum law, we have

$$var(Y - X) = \sigma_Y^2 - 2\rho\sigma_X\sigma_Y + \sigma_X^2 \tag{7}$$

By definition, we have

$$cov(Y, X) = corr(Y, X)\sigma_X\sigma_Y = \rho\sigma_X\sigma_Y \tag{8}$$

Substituting Eqs. 8 and 7 into Eq. 6, we have

$$\begin{aligned} corr(Y - X, X) &= \frac{\rho\sigma_X\sigma_Y - \sigma_X^2}{\sigma_X\sqrt{(\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2)}} \\ &= \frac{\rho\sigma_Y - \sigma_X}{\sqrt{\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2}} \end{aligned} \tag{9}$$

**Correlation between  $(Y - X)$  and  $Y$**

Following the same method for deriving  $corr(Y - X, X)$  in the previous section, it can be shown that  $corr(Y - X, Y)$ , the correlation between variables  $(Y - X)$  and  $Y$ , is

$$corr(Y - X, Y) = \frac{\sigma_Y - \rho\sigma_X}{\sqrt{\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2}} \tag{10}$$

Equations 9 and 10 have been discussed in the context of comparing two measurement methods

[8]. In the special case of equal variance between the two methods (i.e.  $\sigma_X = \sigma_Y$ ), they reduce to

$$\begin{aligned} \text{corr}(Y-X, X) &= -\sqrt{\frac{1-\rho}{2}} \\ \text{corr}(Y-X, Y) &= \sqrt{\frac{1-\rho}{2}} \end{aligned}$$

**Correlation between  $(X - Y)$  and  $(X + Y)$**

In this section, we derive the correlation between variables  $(X - Y)$  and  $(X + Y)$ .

By definition, we have

$$\begin{aligned} \text{corr}(X-Y, X+Y) &= \frac{\text{cov}(X-Y, X+Y)}{\sqrt{\text{var}(X-Y) \cdot \text{var}(X+Y)}} \\ &= \frac{\sigma_X^2 + \text{cov}(X, Y) - \text{cov}(Y, X) - \sigma_Y^2}{\sqrt{\text{var}(X-Y) \cdot \text{var}(X+Y)}} \\ &= \frac{\sigma_X^2 - \sigma_Y^2}{\sqrt{\text{var}(X-Y) \cdot \text{var}(X+Y)}} \end{aligned} \tag{11}$$

By the variance sum law, we have

$$\begin{aligned} \text{var}(X-Y) &= \sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2 \\ \text{var}(X+Y) &= \sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2 \end{aligned}$$

Therefore,

$$\begin{aligned} \text{var}(X-Y) \cdot \text{var}(X+Y) &= (\sigma_X^2 - 2\rho\sigma_X\sigma_Y + \sigma_Y^2) \\ (\sigma_X^2 + 2\rho\sigma_X\sigma_Y + \sigma_Y^2) &= (\sigma_X^2 + \sigma_Y^2)^2 - (2\rho\sigma_X\sigma_Y)^2 \end{aligned} \tag{12}$$

Substitute Eq. 12 into Eq. 11, we have

$$\text{corr}(X-Y, X+Y) = \frac{\sigma_X^2 - \sigma_Y^2}{\sqrt{(\sigma_X^2 + \sigma_Y^2)^2 - (2\rho\sigma_X\sigma_Y)^2}} \tag{13}$$

Equation 13 shows that:

- If  $X$  and  $Y$  have equal variance (i.e.  $\sigma_X^2 = \sigma_Y^2$ ), the correlation between  $(X - Y)$  and  $(X + Y)$  will be zero. This is true whether or not  $X$  and  $Y$  are correlated.
- If  $X$  and  $Y$  are uncorrelated (i.e.  $\rho = 0$ ), then the correlation between  $(X - Y)$  and  $(X + Y)$  will be

$$\text{corr}(X-Y, X+Y) = \frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2} \tag{14}$$

It can be shown that

$$\text{corr}(X-Y, X+Y) = \text{corr}\left(X-Y, \frac{X+Y}{2}\right) \tag{15}$$

A Bland-Altman plot shows the difference (i.e.  $X - Y$ ) of two measures against their average (i.e.  $\frac{X+Y}{2}$ ). The right-hand side of Eq. 15 is what a Bland-Altman plot shows; therefore, what we have derived in this section also applies to Bland-Altman plots.

**Acknowledgements**

The research presented in this paper builds on original contributions from Professor Doug Altman, although he was not directly involved in the work described by this paper. This work was undertaken during the years LC worked for Professor Altman, who passed away during the review period of this paper. The authors dedicate this paper to the memory of Professor Altman, with our deepest admiration, affection, and respect.

**Funding**

Not applicable.

**Availability of data and materials**

Not required.

**Authors' contributions**

LC conceived the research idea and led the writing of the paper. DC contributed to writing the paper. Both authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not required.

**Consent for publication**

Both authors have given consent for publications. No other personal data are used in this methodological paper.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Centre for Statistics in Medicine (CSM), NDORMS, University of Oxford, Oxford, UK. <sup>2</sup>Institute of Biomedical Engineering (IBME), Department of Engineering Science, University of Oxford, Oxford, UK.

Received: 16 May 2018 Accepted: 6 December 2018

Published online: 11 January 2019

**References**

1. Altman DG, Bland JM. How to randomise. *BMJ*. 1999;319(7211):703–4.
2. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005; 330(7495):843.
3. 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.
4. Bland JM. Regression towards the mean or Why was Terminator III such a disappointment? 2004. <https://www-users.york.ac.uk/~mb55/talks/regmean.htm>. Accessed 23 Dec 2018.
5. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–10.
6. Bland JM, Altman DG. Some examples of regression towards the mean. *BMJ*. 1994a;309(6957):780.
7. Bland JM, Altman DG. Statistic notes: regression towards the mean. *BMJ*. 1994b;308(6942):1499.
8. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346(8982): 1085–7.

9. Clifton L, Birks J, Clifton DA. Comparing different ways of calculating sample size for two independent means: a worked example. *Contemp Clin Trials Commun.* 2018;100309.
10. Galton F. Regression towards mediocrity in hereditary stature. *J Anthropol Inst G B Irel.* 1886;15:246–63.
11. Gill JS, Beevers DG, Zezulka AV, Davies P. Relation between initial blood pressure and its fall with treatment. *Lancet.* 1985;325(8428):567–9.
12. Harrell F. How should change be measured? 2017. <http://biostat.mc.vanderbilt.edu/wiki/Main/MeasureChange>. Accessed 23 Dec 2018.
13. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med.* 2012;31(4):328–40.
14. Laird N. Further comparative analyses of pretest-posttest research designs. *Am Stat.* 1983;37(4):329–30.
15. Pocock SJ, Bakris G, Bhatt DL, Brar S, Fahy M, Gersh BJ. Regression to the mean in SYMPLICITY HTN-3: implications for design and reporting of future trials. *J Am Coll Cardiol.* 2016;68(18):2016–25.
16. 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):6.
17. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ.* 2001;323(7321):1123.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

