Ophthalmic Statistics Note 4: analysing data from randomised controlled trials with baseline and follow-up measurements

Rachel Nash,[1] Catey Bunce,[2] Nick Freemantle,[3] Caroline J Doré,[4] Chris A Rogers,[1] On behalf of the Ophthalmic Statistics Group

SCENARIO

In clinical trials, continuous outcomes, such as intraocular pressure and visual acuity, are often measured both before treatment (ie, at baseline) and after treatment. Having the baseline measurement allows us to account for the initial differences between patients, which may well have arisen by chance, when comparing the outcomes of alternative treatments. While randomisation provides a good basis for comparisons between treatments and lowers the probability of baseline imbalance, imbalances can occur simply due to the play of chance, particularly when modest numbers of patients are randomised.

Whether or not baseline measurements are accounted for in the analysis may have an impact on the results of a trial. Imagine a randomised controlled trial with participants randomised to receive either treatment A or treatment B. The primary outcome for the trial is Best Corrected Visual Acuity (BCVA) in the eye with the poorer vision at recruitment, as measured by the number of letters read on an Early Treatment for Diabetic Retinopathy Study EDTRS chart at a distance of 4 m. For the purpose of this scenario, we are assuming that only data from this eye contributes to the analysis. The mean BCVA after treatment is higher in the group allocated treatment B than the group allocated treatment A. However, by chance, the mean BCVA before the start of treatment was also higher for those allocated treatment B. Therefore, although the mean values after treatment suggest that treatment B is better, there may have been as much, or indeed more, improvement with treatment A. This is something that needs to be considered when deciding how the data will be analysed. Decisions about how the data will be analysed should be made before looking at the data (so that they are not influenced by the results) and documented in a Statistical Analysis Plan.[1]

This leaves us with the question: what is the preferred way to analyse data of this nature?

METHODS OF ANALYSIS

There are three common approaches to the analysis of clinical trial data when we have both baseline and post-treatment values, namely:
1. Using a linear regression model, fitting baseline measurement as a patient-level explanatory variable. This method of analysis is known as analysis of covariance (ANCOVA).
2. Analysing post-treatment values only (ie, ignoring baseline measurements).
3. Analysing change scores (ie, the difference between the post-treatment measurement and baseline measurement for each participant).

The merits of ANCOVA and its advantages over the other two approaches are discussed in detail by Vickers and Altman.[2] Briefly, they conclude that ANCOVA is preferable because it (1) provides an estimate of the treatment effect (difference in mean BCVA between treatments A and B) that is unaffected by any baseline imbalance that may exist between the treatment groups, and (2) has a greater chance of detecting a treatment difference if it exists (ie, is a more efficient approach than the other two methods). Other statistical literature reinforces this gain in efficiency and increase in power.[3][4]

When carrying out ANCOVA, a regression model is fitted to the data of the form:

$$\text{post-treatment measurement} = a + b_1 \times \text{treatment B} + b_2 \times \text{baseline measurement} + e \quad (1)$$

The variable 'treatment B' takes the value of 0 if the participant was allocated treatment A, and 1 if the participant was allocated treatment B. Using a statistical package, we can obtain estimates of a, b_1, b_2 and e. The estimate a is a constant, and b_1 quantifies the size of the treatment effect (ie, the mean difference between treatments A and B). The baseline measurement is termed a covariate, and e is an error term. The value of the estimate b_1 depends on the baseline measurements (and the coefficient b_2), and hence we say that our estimate, b_1, 'is conditional upon' (or less formally 'has been adjusted for') the baseline values.

The post-treatment measurements or change scores (methods (2) and (3)) would typically be compared using a two-sample t test. Taking the post-treatment measurements as an example, it can be shown that a two-sample t test is equivalent to fitting a regression model of the form:

$$\text{post-treatment measurement} = a + b_1 \times \text{treatment B} + e \quad (2)$$

The estimate, b_1, corresponds to the difference in mean BCVA after treatment between the groups. The baseline measurement is not included, which is equivalent to setting the estimate b_2 in model (1) to zero.

Looking then at change scores, the model would be:

$$\text{post-treatment measurement} - \text{baseline measurement} = a + b_1 \times \text{treatment B} + e \quad (3)$$

An analysis of change scores is equivalent to setting the estimate b_2 in model (1) to 1.

EXAMPLE

Data from the Inhibition of VEGF in Age-related choroidal Neovascularisation (IVAN) randomised controlled trial comparing ranibizumab (Lucentis) and bevacizumab (Avastin) for the treatment of age-related choroidal neovascularisation[5] have been analysed using the three approaches outlined above to illustrate the differences between the methods. The mean BCVA in the study eye at baseline and at the end of the study (24 months), by drug, is shown in table 1. The mean BCVA at 24 months was slightly higher in the ranibizumab group. However, by chance, this was accompanied by a slightly higher mean baseline BCVA. The results from each analysis are shown in table 2 and discussed below.

[1]Clinical Trials and Evaluation Unit, School of Clinical Sciences, University of Bristol, Bristol Royal Infirmary, Bristol, UK; [2]NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; [3]Department of Primary Care and Population Health, University College London, London, UK; [4]UCL Clinical Trials Unit, University College London, London, UK

Correspondence to Dr Chris A Rogers, Clinical Trials and Evaluation Unit, School of Clinical Sciences, University of Bristol, Bristol Royal Infirmary, Bristol BS2 8HW, UK; Chris.Rogers@bristol.ac.uk

Table 1 Mean values at baseline and post-treatment (24 months) (number of letters on ETDRS chart)

Variable	Ranibizumab (n=268) mean (SD)	Bevacizumab (n=249) mean (SD)
BCVA at baseline	62.9 (14.6)	62.0 (15.3)
BCVA at 24 months	67.8 (17.0)	66.1 (18.4)
Change from baseline	4.9 (15.0)	4.1 (13.5)

BCVA, Best Corrected Visual Acuity.

Analysing post-treatment values only (method 2, model 2)

Looking at post-treatment values only, we do not take into account the higher baseline values in the ranibizumab group, and we therefore potentially overestimate the differences between the treatments (or had the chance imbalance been in the opposite direction, we would potentially underestimate the differences). From the results of this analysis, we would say that at the end of the trial, the average BCVA was 1.7 letters higher in the ranibizumab group compared with the bevacizumab group. The 95% CI tells us that we are 95% confident that the difference in mean BCVA is somewhere between 4.8 letters in favour of ranibizumab and 1.4 letters in favour of bevacizumab. As this CI includes zero, we would not infer a statistically significant difference in mean BCVA at the end of the trial between the two drugs.

Analysing change scores (method 3, model 3)

The analysis of change scores provides us with an estimate of the difference in the mean change from baseline between the two treatment groups. Here, we are taking account of how good the vision was at the start of the trial (through the calculation of the change score), but not of differences in starting BCVA between the two groups. From the results, we find that the mean increase in BCVA was 0.8 letters larger in the ranibizumab group, with a 95% CI from 3.3 letters in favour of ranibizumab to 1.6 letters in favour of bevacizumab. Again, the CI includes zero, so a difference between the groups is not indicated.

Analysing post-treatment values with baseline value as a covariate (ANCOVA, method 1, model 1)

This is the preferred method of analysis. Here, we are estimating the difference in the mean BCVA between the two groups, again taking account of how good the participant's vision was at the start of the trial, but relaxing the restriction on the relationship between baseline and post-treatment measurements. From the results, we would conclude that BCVA improved by an estimated 1.1 letters more, on average, in the ranibizumab group than in the bevacizumab group, with a 95% CI from 3.4 letters in favour of ranibizumab to 1.3 letters in favour of bevacizumab. As with the other two models, there is no suggestion of a difference between the groups because the CI includes zero. The estimated relationship between the baseline and post-treatment measurements for each drug is illustrated in figure 1. The mean difference between the two drugs (1.1 letters) is the vertical distance between the two parallel lines.

DISCUSSION

In this example, all three methods led to the same conclusion, namely that mean BCVA at 24 months was similar between the two drugs. However, the change score analysis which made use of both baseline and post-treatment measurements gave more precise estimates (as shown by smaller SE and narrower CI) than the analysis which just considered the post-treatment measures. Further efficiency was then gained using the more flexible ANCOVA model compared to the analysis of change scores (SE 1.21 vs 1.25 letters, 95% CI (−3.4 to 1.3) vs (−3.3 to 1.6)).

While in this example all three methods led to the same conclusion, it is possible for different models to yield estimates that might lead to different conclusions. If, for example, the more precise estimate had had a CI which excluded zero, while the less precise estimates did not, we might infer evidence of a treatment effect from one model only. Some statisticians feel so strongly about the use of ANCOVA that they describe other methods as a hallmark of second-rate analysis![6]

Another approach to the analysis which we would *not* recommend is to analyse the difference between baseline and post-treatment measurements in the two groups separately, using two paired t tests. This would test the hypothesis that the change from baseline is zero separately for each treatment group. The estimates obtained would give the mean change from baseline in each group, with a corresponding 95% CI, but we would not be able to draw a conclusion about, or quantify the difference between, the two drugs. If we were to perform two paired t tests on our data, we would conclude that there was a significant improvement in BCVA with both ranibizumab and bevacizumab, with a mean improvement of 4.9 letters (95% CI 3.1 to 6.7) and 4.1 letters (95% CI 2.4 to 5.8), respectively. Performing separate analyses within each treatment group is misleading, and either an analysis of change scores or ANCOVA are preferable.[7]

An additional consideration that has not been explored here is how to handle missing data. The methods described would exclude any participant with missing data for any of the measurements included in the analysis. While every effort should be made to prevent missing

Table 2 Analysis results

| Model	Number of patients included in the analysis*	Treatment difference†‡	95% CI	SE
		Number of ETDRS letters		
Method (2), model 2: Post-treatment values only	517	−1.7	(−4.8 to 1.4)	1.55
Method (3), model 3: Change from baseline	517	−0.8	(−3.3 to 1.6)	1.25
Method (1), model 1: Post-treatment values with baseline value as a covariate	517	−1.1	(−3.4 to 1.3)	1.21

*Although 525 patients reached the 24-month visit in the IVAN trial, BCVA data are missing for 8 of these patients.
†Difference in mean BCVA (bevacizumab—ranibizumab).
‡The results presented differ from the published results.[5] For illustrative purposes, only the drug treatment was included in the model; BCVA measured during follow-up and other factors included models used to obtain the published results have not been included.
BCVA, Best Corrected Visual Acuity.

Figure 1 Baseline and post-treatment Best Corrected Visual Acuity (BCVA) for the subset of patients with a post-treatment BCVA of more than 50 letters (n=425 patients). The estimated difference in mean BCVA between the two drugs groups from the analysis of covariance is the vertical distance between the two regression lines shown on the plot.
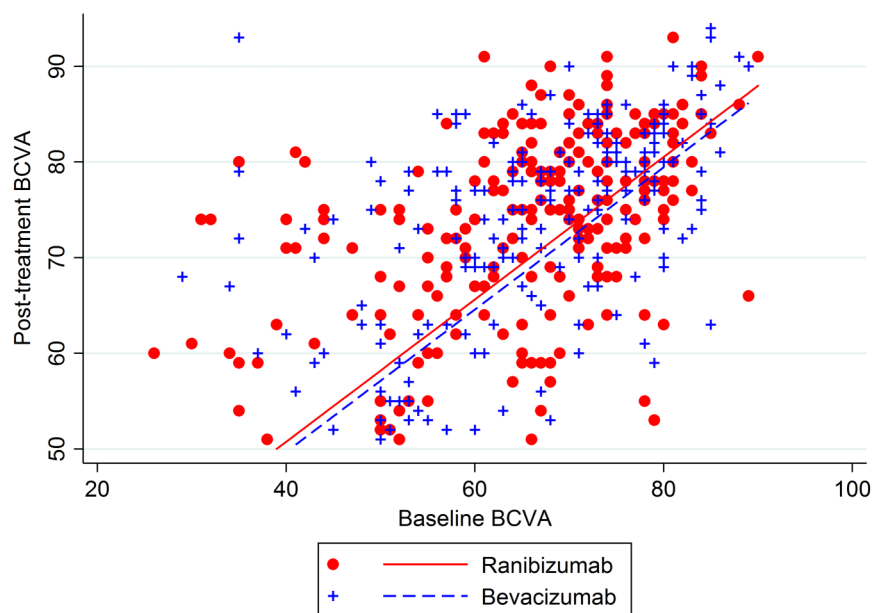
data by study design and management, missing values can and do occur. For example, in the IVAN trial, participants underwent optical coherence tomography to assess retinal thickness and other lesion morphology, but on occasions, the machine was not working and the measurements were not taken. There are different ways in which missing data can be handled and you should consult a statistician on the best way to proceed. Approaches include omitting the cases with missing data, which is an inefficient use of the data, reducing precision and power; imputing the missing values, which must be done with care[8]; and fitting a more sophisticated model where the baseline and post-treatment measurements are modelled 'jointly', which allows participants with partial missing data to be included.[9] However, it is very important to remember that none of these methods are a solution to missing data, and that every effort should be made to prevent it.

In summary, while there are different methods available for analysing trial data with baseline and post-treatment measurements, the recommended approach is ANCOVA for the reasons outlined. The choice of analysis method can impact the results of a trial, and therefore, it is important to choose the most appropriate method in advance to ensure precise and unbiased conclusions.

Collaborators The Ophthalmic Statistics Group: David Cairns, Valentina Cipriani, Jonathan Cook, David Crabb, Phillippa Cumberland, Gabriela Czanner, Paul Donachie, Andrew Elders, Marta Garcia Finana, Neil O'Leary, Krishna Patel, Toby Prevost, Ana Quartilho, Luke Saunders, Selvaraj Sivasubramaniam, Simon Skene, Irene Stratton, Joana Vasconcelos, Wen Xing, Haogang Zhu.

OPEN ACCESS

Open Access Scan to access more free content

CrossMark

REFERENCES

1 ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med 1999;18:1905–42.
2 Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ 2001;323:1123–4.
3 Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1992; 11:1685–704.
4 Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006;59:920–5.
5 Chakravarthy U, Harding SP, Rogers CA, et al. Alternative treatments to inhibit VEGF in age-related choroidal neovascularisation: 2-year findings of the IVAN randomised controlled trial. Lancet 2013;382:1258–67.
6 Senn S. Change from baseline or analysis of covariance?: Lord's Paradox and other matters. University of Glasgow, 2004.
7 Altman D. Practical statistics for medical research. London: Chapman and Hall, 1991.
8 Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338: b2393.
9 Carpenter J, Kenward M. Missing data in randomised controlled trials—a practical guide. London School of Hygiene, 2007.