The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study

ABSTRACT

Percentage change from baseline should not be used in statistical analysis. Trialists wishing to report this statistic should use another method, such as ANCOVA, and convert the results to a percentage change by using mean baseline scores.

INTRODUCTION

Background Many randomized trials involve measuring a continuous outcome at baseline and after treatment. Typical examples include trials of pravastatin for hypercholesterolemia, exercise and diet for obesity in osteoarthritis patients and acupuncture for pain in athletes with shoulder injuries. In each trial, the outcome measure used to determine the effectiveness of treatment - cholesterol, body weight or shoulder pain - was measured both before treatment had started and after it was complete. In the case of a single post treatment outcome assessment, there are four possibilities for how such data can be entered into the statistical analysis of such trials. One can use the baseline score solely to ensure baseline comparability and enter only the post-treatment score into analysis (I will describe this method as "POST"). Alternatively, one can analyze the change from baseline, either by looking at absolute differences ("CHANGE") or a percentage change from baseline ("FRACTION"). The most sophisticated method is to construct a regression model which adjusts the post-treatment score by the baseline score ("ANCOVA"). Figure 1 describes each of these methods in mathematical terms. Figure 2 gives examples of the results of each method described in ordinary language. Some trials assess outcome several times after treatment, a design known as "repeated measures." Each of the four methods described above can be used to analyze such trials by using a summary statistic such as a mean or an "area-under-curve". There are several more complex methods of analyzing such data including repeated measures analysis of variance and generalized linear estimation. These methods are of particular value when the post-treatment scores have a predictable course over time (e.g. quality of life in late stage cancer patients) or when it is important to assess interactions between treatment and time (e.g. long-term symptomatic medication). This paper will concentrate on the simpler case where time in not an important independent variable. The choice of which method to use can be determined by analysis of the statistical properties of each. An important criteria for a good statistical method is that it should reduce the rate of false negatives (β). The β of a statistical test is usually expressed in terms of statistical power ($1-\beta$). Power is normally fixed, typically at 0.8 or 0.9, and the required amount of data (e.g. number of evaluable patients) is calculated. A method that requires relatively fewer data to provide a certain level of statistical power is described as efficient. The characteristics of the four methods - POST, CHANGE, FRACTION and ANCOVA - have been studied by statisticians for some time. In this paper, I aim to provide statistical data that can guide clinical research yet is readily comprehensible by non-statisticians. Accordingly, I will compare the methods using a hypothetical trial and express results in terms of statistical power.

CONCLUSION

Reporting a percentage change from baseline gives the results of a

randomized trial in clinically relevant terms immediately accessible to patients and clinicians alike. This is presumably why researchers investigating issues such as the effects of medication on hot flashes, or of different chemotherapy regimes on quality of life, report this statistic. However, percentage change from baseline is statistically inefficient. Perhaps counterintuitively, it does not correct for imbalance between groups at baseline. It may also create a non-normally distributed statistic from normally distributed data. Percentage change from baseline should therefore not be used in statistical analysis. Trialists wishing to report percentage change should first use another method, preferably ANCOVA, to test significance and calculate confidence intervals. They should then convert results to percentage change by using mean baseline and post-treatment scores. For an example of this approach, see Crouse et al.. The findings presented here reconfirm previously reported data suggesting that ANCOVA is the method of choice for analyzing the results of trials with baseline and post treatment measurement. In cases where ANCOVA cannot be used, such as with small samples or where the assumptions underlying ANCOVA modeling do not hold, CHANGE or POST are acceptable alternatives, especially baseline variables are comparable between groups (perhaps ensured by stratification) and if correlation between baseline and post-treatment scores are either high (for CHANGE) or low (for POST). The use of FRACTION should be avoided.