Understanding controlled trials

Baseline imbalance in randomised controlled trials

Chris Roberts, David J Torgerson

In a controlled trial randomisation ensures that allocation of patients to treatments is left purely to chance. The characteristics of patients that may influence outcome are distributed between treatment groups so that any difference in outcome can be assumed to be due to the intervention. However, imbalance between groups in baseline variables that may influence outcome (such as age or disease severity) can bias statistical tests, a property sometimes referred to as chance bias. Observed differences in outcome between groups in a particular trial could by chance be due to characteristics of the patients, not treatments. Some protection against chance bias is given by stratified randomisation or minimisation and by adjusting in the statistical analysis for baseline variables.

In reporting clinical trials it is recommended that prognostic variables should be described for each treatment group.[1] This may be helpful in understanding the generalisability of the study and may assure the reader that the randomisation has been properly conducted. A common practice is to check for imbalance between intervention groups by statistical tests of baseline characteristics. If the result is statistically significant, the investigator may nevertheless argue that this is not a problem as the variable is not strongly associated with outcome. Alternatively an analysis adjusting for the baseline imbalance may be presented. However, this practice of statistical testing of baseline variables to assess the effect of imbalance, although common,[2] has been criticised.[3] In carrying out such tests three questions are being confused:

- Has the randomisation been properly conducted?
- Could imbalance in baseline characteristics cause chance bias?
- Should the analysis be adjusted for baseline variables?

When randomisation has been properly conducted the null hypothesis that treatment groups come from the same population is true. In the usual framework of statistical inference rejection of the null hypothesis should lead to the conclusion that the groups are not properly randomised. Such tests may therefore be used to detect possible subversion of the allocation procedure.[4]

When there is a moderate association between the baseline characteristic and patient outcome, it has been shown that chance bias on a statistical test of outcome may be appreciable when imbalance between groups is well above the conventional 5% level for statistical significance.[5] It follows therefore that a significance test of baseline characteristic does not provide an appropriate criterion to assess the effect of imbalance on outcome or the decision to adjust for baseline variables.

As the trial size increases the absolute size of imbalance in baseline characteristics will reduce owing to reduction in sampling error. Hence the absolute magnitude of any chance bias in outcome will tend to decrease with sample size. Nevertheless, the possible chance bias on a statistical test of an outcome measure does not change with sample size,[5] so chance bias is as much of a possibility for large trials as for small. An imbalance of a given absolute size will have a greater effect on the statistical tests for larger sample sizes than for small. This means that inspection of the distribution of baseline variables between groups is also an inappropriate method on which to base the decision to adjust or not adjust a statistical test of a trial outcome.

If we accept that statistical tests and visual inspection of differences between groups are unsound methods of choosing to adjust for baseline, one proposed strategy[6] is as follows.

- At the planning stage of a study baseline variables of prognostic value should be identified on the basis of available evidence.
- These should be fitted in an analysis of covariance or equivalent technique for other data types.
- Other variables should not be added to the analysis unless information from other sources during the course of the trial suggests their inclusion.

To summarise, choice of baseline characteristics by which an analysis is adjusted should be determined by prior knowledge of an influence on outcome rather than evidence of imbalance between treatment groups in the trial. Such information should ideally be included in trial protocols and reported with details of the analysis. Baseline tests of imbalance are inappropriate unless the investigators suspect that there are problems with the randomisation.

1 Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996;313:570-1.
2 Alman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.
3 Altman DG. Comparability of randomised groups. *Statistician* 1985;34:125-36.
4 Kennedy A, Grant A. Subversion of allocation in a randomised controlled trial. *Control Clin Trial* 1997;18(suppl 3):77-8S.
5 Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 1989;8:467-75.
6 Senn S. Testing for baseline balance in clinical trials. *Stat Med* 1994;13:1715-26.

Correction

Not playing with a full DEC: why development and evaluation committee methods for appraising new drugs may be inadequate

In this article by Nick Freemantle and James Mason (29 May, pp 1480-2), the reference list did not include reference 2, the NICE discussion paper ("Faster access to modern treatment": how NICE appraisal will work. Leeds: NHS Executive, 1999). As a consequence, from reference 3 onwards, reference numbers in the text refer to the previous number in the reference list.

National Primary Care Research Development Centre, University of Manchester, Manchester M13 9PL
Chris Roberts, *senior research fellow*
National Primary Care Research and Development Centre, Centre for Health Economics, University of York, York YO1 5DD
David J Torgerson, *senior research fellow*
Correspondence to: Dr Roberts

BMJ 1999;319:185