

Baseline data in clinical trials

David C Burgess, Val J Gebski and Anthony C Keech

ALTHOUGH REPORTING BASELINE DATA seems simple, it is crucial information for readers in judging the validity of a trial. Knowing the baseline characteristics of the trial participants allows readers to assess how closely these match patients seen in their own clinical practice, and therefore how generalisable the results of the trial will be (so-called external validity). Baseline characteristics also allow the success of randomisation to be assessed. In studies where important baseline factors appear well balanced, it is likely that any differences in outcome between the intervention and control groups are a real effect of treatment (one component of internal validity). For these reasons, the reporting of baseline demographic and clinical characteristics of each group is a requirement of the CONSORT statement.¹ The item and its descriptor as they appear in the CONSORT checklist are shown in Box 1, and a checklist for baseline data is provided in Box 2.

Content

Baseline data should adequately describe the population in the trial. This means including demographic variables, known factors that influence the outcome (including medications being taken by participants), factors that are likely to modify any benefit of treatment, and those that may predict adverse reactions. These factors are called potential “confounders”, because, if they are imbalanced between the treatment groups at baseline, they may result in an apparent treatment effect when none exists, or mask an effect that does exist. Baseline data should also include any factors (especially known potential confounders) that have been used as strata for randomisation. Stratified randomisation, described in detail earlier,² is used when a baseline characteristic, such as tumour stage, is known to affect outcome risk; the characteristic is therefore included in the randomisation algorithm to minimise imbalances between treatment groups. This is particularly useful in small studies.

If the study population contains subgroups of particular interest, the characteristics defining these subgroups, and numbers or proportion in each group, should be stated. For example, in a long-term trial of a new medication for preventing heart attack, diabetes mellitus would be a potential confounder (as people with diabetes have a much higher

1: CONSORT checklist of items to report when reporting a trial¹

Section and topic	Item no.	Descriptor
Baseline data	15	Baseline demographic and clinical characteristics of each group

risk of heart attack than similar people without diabetes). Those with diabetes in this study would also be an interesting subgroup in whom the effects of the intervention might be different. Similarly, concurrent therapy with aspirin (which would substantially reduce the risk of heart attack) could confound the trial results if there was an imbalance between trial groups in the proportions of patients taking aspirin; aspirin therapy might also influence the likelihood of adverse reactions to study therapy. Baseline factors can be determined from interviews, physical examination, laboratory measures or imaging studies.

Measurement

Baseline data are measured as close as possible to the time that participants are randomly allocated to study groups, and in all cases, should be measured before the allocated treatment commences (information collected after the commencement of trial treatment may have been altered by the treatment itself, and is generally not regarded as baseline data). Ideally, baseline data should be collected on all patients screened for eligibility, as this would provide further information about the generalisability of the trial population. However, this is not always practicable or affordable, so some variables (eg, tissue biopsy, measurement of genetic markers, expensive imaging tests) are measured only in actual participants randomly allocated to a trial group.

For factors that are not constant, the conditions under which the baseline data are collected should be stated in the methods section of the study report. For example, it should be clear whether blood pressure recordings were measured sitting or supine, or after a specified rest period; also whether a single reading, the average of several readings, the highest of two, or the first of two or more, was used.

Baseline data as entry criteria

In some circumstances, threshold levels of one or more baseline variables will form part of the entry criteria for the study. In this case, if an extreme value of a baseline factor, such as high blood pressure, is required to qualify a person for entry into a study, potential participants whose value on the day of screening is more extreme (higher) than their usual level will be more likely to qualify for entry. A second

NHMRC Clinical Trials Centre, University of Sydney, Camperdown, NSW.

David C Burgess, BMed, Cardiology Research Fellow; Val J Gebski, BA, MStat, Principal Research Fellow; Anthony C Keech, MSc(Epid), FRACP, Deputy Director.

Reprints will not be available from the authors. Correspondence: Associate Professor Anthony C Keech, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown, NSW 1450. tony@ctc.usyd.edu.au

2: Checklist for baseline data

Measurement

Consider all important baseline variables to be measured and how they are to be measured before treatment starts:

- Demographic characteristics (age, sex, height, weight, etc)
- Known factors that predict the outcome (potential confounders)
- Factors that predict or alter the risk of adverse reactions
- Stratification factors
- Pre-specified subgroups.

Reporting

Tabulate relevant summary measures (eg, mean and standard deviation).

Include all important baseline characteristics while keeping the table readable.

Wherever possible, avoid displaying *P* values.

Analysis

In the results section, discuss the similarity of the two groups, highlighting any clinically important differences that may influence the outcome.

Discussion

Discuss the effect of the baseline data balance on the internal validity of the study and the comparability of the study population to patients seen in wider clinical practice.

baseline reading of the average blood pressure for this group will be lower and more accurately reflect their usual blood pressure; this is known as regression towards the mean.³ For this reason, remeasuring factors required for entry is desirable, to establish a more realistic group average value of the characteristic at baseline.

Presentation

The baseline characteristics are usually presented in the first table in a report. Care should be taken to include the necessary descriptive information without overwhelming readers with unnecessary details. For example, in the recent AFFIRM trial comparing rate control with rhythm control of atrial fibrillation, the published first table has 16 baseline characteristics, each with a mean and percentage value for the overall group, and for both treatment groups separately, together with *P* values.⁴ The resulting table of 107 values and four footnotes may make it difficult for some readers to extract the key information.⁵ A simpler presentation appears in the FRISC II study of invasive compared with non-invasive treatments for unstable coronary artery disease.⁶ This presents more baseline characteristics (20), but by minimising detail (omitting overall group and *P* values), allows a more rapid comparison of the characteristics between groups.

Comparability between groups

If randomisation has been performed correctly, the groups should be similar in baseline characteristics, except for the play of chance. Stratification in the randomisation process further restricts the extent of chance imbalances.² For

continuous variables (such as blood pressure, age, cholesterol level), the similarity of the treatment groups should be assessed by comparing relevant summary measures (mean and standard deviation, or median and range). For categorical factors (such as sex, disease stage), the numbers and proportions in each category level should be shown for each treatment group. The more similar the treatment groups, the more credible are the trial results as reflecting a true result of treatment, especially if unadjusted analyses are presented.^{5,7}

Use of *P* values to assess randomisation

Use of statistical tests to compare the balance and/or values of baseline characteristics between the study groups and the presentation of *P* values are not uncommon. However, many authors assert that this is inappropriate.^{3,5,8-10} If randomisation has been performed correctly, chance is the only explanation for any observed difference between groups at the outset of the study, in which case statistical tests become superfluous. Consequently, only if it is suspected that the randomisation process has failed or was flawed, can performing significance tests on the baseline data be readily justified.⁸ It is worth remembering that, if 20 baseline characteristics are presented from a trial using simple randomisation, it is more likely than not that at least one characteristic will show a significant imbalance between groups at two-sided $P < 0.05$ by chance alone (actual likelihood, 64%).

In any case, providing *P* values is not a substitute for carefully describing, in the results section, any imbalances between study groups that may be clinically important. For example, in a trial of a thrombolytic drug, a 1% baseline difference in history of previous intracranial haemorrhage may not be statistically significant, but could still affect haemorrhagic stroke rates after treatment (an outcome of the study), and hence could be regarded as potentially clinically significant. If there are imbalances that are considered important to the final study results, they should be accounted for by an adjusted analysis of the data, not simply noted with a *P* value in the first table.⁷

Other uses of baseline data

A longer-term benefit of collecting comprehensive baseline data is that, after outcome data become available, it allows the estimation of risk of the outcome in the control group, related to various baseline characteristics. This effectively uses the control group as an epidemiological cohort study, providing contemporary information about predictors of disease outcomes.

In summary, careful planning and collection of baseline data enables performance of a high-quality trial and allows readers to clearly see the internal and external validity of the study.

Competing interests

None identified.

References

1. Moher D, Schulz K, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials. *Lancet* 2001; 357: 1191-1194.
2. Beller EM, Gebbski VJ, Keech AC. Randomisation in clinical trials. *Med J Aust* 2002; 177: 565-567.
3. Friedman L, Furberg C, DeMets D. Fundamentals of clinical trials. 3rd ed. New York: Springer, 1998.
4. Wise DG, Waldo AL, DiMarco JP, et al for AFFIRM Investigators. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med* 2002; 347: 1825-1833.
5. Pocock SJ, Assmann SE, Enos LE, Casten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Stat Med* 2002; 21: 2917-2930.
6. Wallen L, Swahn E, Kontny F, et al for FRISC II Investigators. Invasive compared with non-invasive treatment in unstable coronary-artery disease: FRISC II prospective randomised multicentre study. *Lancet* 1999; 354: 708-715.
7. Gebbski V, Keech AC. Statistical methods in clinical trials. *Med J Aust* 2003; 178: 182-184.
8. Matthews J. An introduction to randomised controlled clinical trials. London: Arnold, 2000.
9. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-152.
10. Senn S. Testing for baseline balance in clinical trials. *Stat Med* 1994; 13: 1715-1726.

(Received 6 Jun 2003, accepted 9 Jun 2003)

□

snapshot

Bowel wall “thumbprinting” in pseudomembranous colitis

A 39-YEAR-OLD WOMAN with AIDS (CD4 count, 30 cells/ μ L) had a 4-day history of nausea, vomiting and profuse watery diarrhoea. The patient was afebrile and had a distended abdomen with diffuse guarding without rebound tenderness. Abdominal x-rays (Box 1) and computed tomography scans (Box 2) were performed. The white blood cell count was within normal limits and stool cultures were negative. Colonoscopy revealed yellow plaques throughout the colon. The patient improved clinically after taking oral metronidazole.

Bowel wall “thumbprinting” (the appearance of “thumbprint”-shaped projections) is a radiological sign of thickening of the colonic wall. It occurs secondary to submucosal haemorrhage and oedema from capillary leakage.¹ It can result from any process that leads to oedema of the bowel wall, including pseudomembranous colitis (as shown here), ischaemic colitis, non-infective inflammatory bowel disease, other infective bowel diseases, submucosal/intramural haemorrhage and other conditions.² The mucosal damage and inflammation seen in pseudomembranous colitis are caused by *Clostridium difficile* toxin.³

Moishe Liberman,* Chris Labos,† Jeff Wiseman‡

*Surgical Resident, Department of Surgery, McGill University Health Centre, Montreal, QC, Canada; †Medical Student McGill University Medical School, Montreal, QC, Canada
‡Physician, Department of Medicine, Royal Victoria Hospital Montreal, QC, Canada

Moishe.liberman@mail.mcgill.ca

1. Grainger RG, Allison D, Adam A, et al (editors). Grainger and Allison's diagnostic radiology: a textbook of medical imaging. 4th ed. New York: Churchill Livingstone, 2001: 1132.
2. Gore RM, Levine MS. Textbook of gastrointestinal radiology. 2nd ed. Philadelphia: WB Saunders, 2000: 1163.
3. Kelly CP, Pothoulakis C, LaMont JT. *Clostridium difficile* colitis. *N Engl J Med* 1994; 330: 257-262.

□

