Statistics Notes

Units of analysis

Douglas G Altman, J Martin Bland

ICRF Medical
Statistics Group,
Centre for Statistics
in Medicine,
Institute of Health
Sciences, Oxford
OX3 7LF
Douglas G Altman,
head

Department of
Public Health
Sciences, St
George's Hospital
Medical School,
London SW17 0RE
J Martin Bland,
*professor of medical
statistics*

Correspondence to:
Mr Altman.

In clinical studies the focus of interest is almost always the patient. If we carry out a randomised trial to compare two treatments we are interested in comparing the outcomes of patients who received each of the treatments. In some conditions several measurements will be taken on the same patient, but the focus of interest remains the patient. Failure to recognise this fact results in multiple counting of individual patients and can seriously distort the results. We explain this error below. Its frequency in medical research is indicated by the whole chapter devoted to it in Andersen's classic compilation.[1]

The simplest case is when researchers study a part of the human anatomy which is, so to speak, in duplicate: eyes, ears, arms, etc. At the other extreme very many measurements can be taken on a single patient. Such data arise frequently in dentistry, with measurements made on each tooth, or even each face of each tooth, and in rheumatology, in which pain or mobility may be assessed for each joint of each finger. In statistical terminology the patient is the sampling unit (or unit of investigation) and thus should be the unit of analysis.

There are two related consequences of ignoring the fact that the data include multiple observations on the same individuals. Firstly, this procedure violates the widespread assumption of statistical analyses that the separate data values should be independent. Secondly, the sample size is inflated, sometimes dramatically so, which may lead to spurious statistical significance.

Inflated samples

To take a simple case, we may wish to compare the blood pressures of two groups of 30 patients. If we measured blood pressure on each arm of each patient we could double the number of observations but not the amount of information, as the two pressures from each patient will be very similar. The use of the t test to compare the two sets of 60 observations is invalid. Andersen[1] presented data from a randomised double blind crossover trial of ketoprofen and aspirin in the treatment of rheumatoid arthritis. An impressive P value of 0.00000001 was obtained from an analysis of 3944 observations, but these were obtained from only 58 patients. Such errors are not rare. In a review of 196 randomised trials of non-steroidal anti-inflammatory agents Gøtzsche found that 63% of reports used the wrong units of analysis.[2]

We previously discussed a similar fallacy arising in the use of correlation coefficients, when multiple observations from each individual produced a spurious increase in the sample size and a corresponding spurious "significant" relationship.[3] We suggested techniques to analyse such data when the focus was either the variation within subjects[4] or between subjects.[5]

There is nothing wrong in collecting such data; indeed the use of multiple observations can often improve the statistical power of a study. But such studies need to be analysed correctly. The simplest approach is to collapse all the data for an individual into a summary measure.[6] For example, we could validly analyse the mean of the two blood pressure values for each patient. Alternatively, we can use a statistical method which explicitly takes account of the multiplicity. With well designed studies we may be able to use analysis of variance. A more complex general approach is multilevel modelling,[7] which is not available in standard statistical software and may be difficult to apply and interpret.

Take account of multiplicity

The same objection applies to the use of multiple measurements made on different occasions. Here too the sampling unit is the patient, and thus the unit of analysis should also be the patient.[2] A further feature of this type of study is that in some situations the number of measurements made on a patient may itself carry prognostic information. For example, repeat measurements may be made only if there is some clinical concern—for example, fetal ultrasound measurements in pregnancy. To treat all these measurements as independent is clearly wrong, but bias is introduced too when those with more data are systematically different from those with single observations. An extreme example of this phenomenon occurs when analysing multiple hospital admissions for a potentially fatal condition.[1] Those with more than one admission must have survived the first admission.

Failure to carry out the correct analysis can lead to problems of interpretation too. Commenting on one trial, Andersen observed, "This trial resulted in the apparent conclusion that after 1 year 22% of the patients, but only 16% of the legs, have expired."[1]

Similar problems arise when we cannot sample individual patients directly but choose a sample of hospitals, wards, or general practices and then obtain data for all or a subsample of the patients within these groups. Here analysis of data for individual patients leads to the errors described above. We consider this type of study in forthcoming Statistics Notes.

1 Andersen B. *Methodological errors in medical research.* Oxford: Blackwell, 1990.
2 Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989;10:31-56.
3 Bland JM, Altman DG. Correlation, regression, and repeated data. *BMJ* 1994;308:896.
4 Bland JM, Altman DG. Calculating correlation coefficients with repeated observations. Part 1: correlation within subjects. *BMJ* 1995;310:446.
5 Bland JM, Altman DG. Calculating correlation coefficients with repeated observations. Part 2: correlation between subjects. *BMJ* 1995;310:633. [Correction *BMJ* 1996;312:572]
6 Matthews JNS, Altman DG, Campbell MJ, Royston JP. Analysis of serial measurements in medical research. *BMJ* 1990;300:230-235.
7 Goldstein H. *Multi-level statistical models.* 2nd ed. London: Edward Arnold, 1995.