STATISTICS NOTES

Analysis of continuous data from small samples

J Martin Bland,[1] Douglas G Altman[2]

Studies with small numbers of measurements are rare in the modern *BMJ*, but they used to be common and remain plentiful in specialist clinical journals. Their analysis is often more problematic than that for large samples, and **Martin Bland** and **Douglas Altman** discuss how to tackle this

[1]Department of Health Sciences, University of York, York YO10 5DD
[2]Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Oxford OX2 6UD
Correspondence to: Professor Bland mb55@york.ac.uk

Cite this as: *BMJ* 2009;338:a3166
doi: 10.1136/bmj.a3166

Parametric methods, including *t* tests, correlation, and regression, require the assumption that the data follow a normal distribution and that variances are uniform between groups or across ranges.[1] In small samples these assumptions are particularly important, so this setting seems ideal for rank (non-parametric) methods, which make no assumptions about the distribution of the data; they use the rank order of observations rather than the measurements themselves.[1] Unfortunately, rank methods are least effective in small samples. Indeed, for very small samples, they cannot yield a significant result whatever the data. For example, when using the Mann-Witney test for comparing two samples of fewer than four observations a statistically significant difference is impossible: any data give $P>0.05$. Similarly, the Wilcoxon paired test, the sign test, and Spearman's and Kendall's rank correlation coefficients cannot produce $P<0.05$ for fewer than six observations. Methods based on the *t* distribution do not have this problem and can detect differences in samples as small as two for paired differences and three for two groups, or detect correlations in samples of three.

Example 1

We were recently asked about the data in table 1, which shows before and after measurements of pudendal nerve terminal motor latency. Should we use the Wilcoxon or the sign test? MB replied that the Wilcoxon would be acceptable, giving $P<0.05$ (actually $P=0.047$), and so would the paired *t* test, which gave $P=0.04$. The questioner also asked whether the Wilcoxon test could be used for the second group of four observations alone, for patients who had received a slightly different intervention. Here all the differences are in the same direction, but the Wilcoxon test gives $P=0.125$. It is not possible for it to give a significant difference. The paired *t* test gives $P=0.04$, a significant difference.

Example 2

On the other hand, using *t* methods when their assumptions are greatly violated can also be misleading. Table 2 shows concentration of antibody to type II group B Streptococcus in 20 volunteers before and after immunisation.[2 3] The comparison of the antibody levels was summarised in the report of this study as "$t=1.8$; $P>0.05$". The paired *t* test is not suitable for these data, because the differences clearly have a very skewed distribution. There are 8 zero differences, forming a clump at one end of the distribution, which would remain whatever transformation we used. We could consider the Wilcoxon

Table 1 | Five year follow-up of patients receiving hyperbaric oxygen therapy for faecal incontinence

	Pudendal nerve terminal motor latency (ms)		
Subgroup	Initial	Follow-up	Change
1	2.2	2.3	0.1
1	2.3	1.6	−0.7
1	2.1	2.2	0.1
1	2.4	2.3	−0.1
2	2.3	2.1	−0.2
2	2.4	1.8	−0.6
2	2.4	1.9	−0.5
2	2.6	1.6	−1.0

Table 2 | Concentration of antibody to type II group B Streptococcus in 20 volunteers before and after immunisation (Baker et al,[2] reported by Altman[3])

Antibody to type II group B Streptococcus (μg/ml)		
Before	After	Change
0.4	0.4	0.0
0.4	0.5	0.1
0.4	0.5	0.1
0.4	0.9	0.5
0.5	0.5	0.0
0.5	0.5	0.0
0.5	0.5	0.0
0.5	0.5	0.0
0.5	0.5	0.0
0.6	0.6	0.0
0.6	12.2	11.6
0.7	1.1	0.4
0.7	1.2	0.5
0.8	0.8	0.0
0.9	1.2	0.3
0.9	1.9	1.0
1.0	0.9	−0.1
1.0	2.0	1.0
1.6	8.1	6.5
2.0	3.7	1.7

paired sample test, but this method assumes that the differences have a symmetrical distribution, which they do not. The sign test is preferred here; it tests the null hypothesis that non-zero differences are equally likely to be positive or negative, using the binomial distribution. We have 1 negative and 11 positive differences, which gives P=0.006. Hence the original authors failed to detect a difference because they used an inappropriate analysis.

We have often come across the idea that we should not use *t* distribution methods for small samples but should instead use rank based methods. The statement is sometimes that we should not use *t* methods at all for samples of fewer than six observations.[4] But, as we noted, rank based methods cannot produce anything useful for such small samples.

Draw on general experience

The aversion to parametric methods for small samples may arise from the inability to assess the distribution shape when there are so few observations. How can we tell whether data follow a normal distribution if we have only a few observations? The answer is that we have not only the data to be analysed, but usually also experience of other sets of measurements of the same thing. In addition, general experience tells us that body size measurements are usually approximately normal, as are the logarithms of many blood concentrations and the square roots of counts.

We thank Jonathan Cowley for the data in table 1.

Competing interests: None declared.

Provenance and peer review: Commissioned, not externally peer reviewed.

1 Altman DG, Bland JM. Parametric *v* non-parametric methods for data analysis. *BMJ* 2009;338:a3167.
2 Baker CJ, Kasper DL, Edwards MS, Schiffman G. Influence of preimmunization antibody levels on the specificity of the immune response to related polysaccharide antigens. *N Engl J Med* 1980;303:173-8.
3 Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1991: 224-5.
4 Siegel S. *Nonparametric statistics for the behavioral sciences.* 1st ed. Tokyo: McGraw-Hill Kogakusha, 1956:32.

Research methods and reporting

Research methods and reporting is for "how to" articles—those that discuss the nuts and bolts of doing and writing up research, are actionable and readable, and will warrant appraisal by the *BMJ*'s research team and statisticians. These articles should be aimed at a general medical audience that includes doctors of all disciplines and other health professionals working in and outside the UK. You should not assume that readers will know about organisations or practices that are specific to a single discipline.

We welcome articles on all kinds of medical and health services research methods that will be relevant and useful to *BMJ* readers, whether that research is quantitative or qualitative, clinical or not. This includes articles that propose and explain practical and theoretical developments in research methodology and for those on improving the clarity and transparency of reports about research studies, protocols, and results.

This section is for the "how?" of research, while the "what, why, when, and who cares?" will usually belong elsewhere. Studies evaluating ways to conduct and report research should go to the *BMJ*'s Research section; articles debating research concepts, frameworks, and translation into practice and policy should go to Analysis, Editorials, or Features; and those expressing personal opinions should go to Personal View.

Articles for Research methods and reporting should include:

- Up to 2000 words set out under informative subheadings. For some submissions, this might be published in full on bmj.com with a shorter version in the print *BMJ*
- A separate introduction ("standfirst") of 100-150 words spelling out what the article is about and emphasising its importance
- Explicit evidence to support key statements and a brief explanation of the strength of the evidence (published trials, systematic reviews, observational studies, expert opinion, etc)
- No more than 20 references, in Vancouver style, presenting the evidence on which the key statements in the paper are made
- Up to three tables, boxes, or illustrations (clinical photographs, imaging, line drawings, figures—we welcome colour) to enhance the text and add to or substantiate key points made in the body of the article
- A summary box with up to four short single sentences, in the form of bullet points, highlighting the article's main points
- A box of linked information such as websites for those who want to pursue the subject in more depth (this is optional)
- Web extras: we may be able to publish on bmj.com some additional boxes, figures, references (in a separate reference list numbered w1,w2,w3, etc, and marked as such in the main text of the article)
- Suggestions for linked podcasts or video clips, as appropriate
- A statement of sources and selection criteria. As well as the standard statements of funding, competing interests, and contributorship, please provide at the end of the paper a 100-150 word paragraph (excluded from the word count) explaining the paper's provenance. This should include the relevant experience or expertise of the author(s) and the sources of information used to prepare the paper. It should also give details of each author's role in producing the article and name one as guarantor.