

This Week's Citation Classic

CC/NUMBER 40
OCTOBER 5, 1992

Bland J M & Altman D G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307-10, 1986.

[Dept. Clinical Epidemiology and Social Medicine, St. George's Hosp. Medical Sch., London, and Div. Medical Statistics, MRC Clinical Res. Ctr., Northwick Park Hosp., Harrow, Middlesex, England]

This paper criticized the use of correlation coefficients in the comparison of two methods of medical measurement, arguing that correlation was not the same as agreement. We proposed a method based on the mean and standard deviation of the differences between the measurements by the two methods and gave a worked example. [The SC[®] indicates that this paper has been cited in more than 620 publications.]

Comparing Methods of Clinical Measurement

J. Martin Bland

Department of Public Health Sciences
St. George's Hospital Medical School
Cranmer Terrace
London SW17 0RE

and

Douglas G. Altman
Medical Statistics Laboratory
Imperial Cancer Research Fund
Lincoln's Inn Fields
London WC2A 3PX
England

Studies which compare two methods of measurement are very common in the medical literature, and they are almost always incorrectly analyzed. Our interest in this began around 1980, when each of us was approached separately by cardiologists about the use of the correlation coefficient to compare two methods of measurement. We discussed the inappropriateness of this analysis over the phone and agreed that the use of the differences between observations on the same subject was the correct approach and that a good method would be to plot these differences against the average of the two methods for the subject. The mean and standard deviation of the differences would enable us to estimate how far apart measurements by the two methods on the same subject were likely to be.

D.G. Altman mentioned the problem briefly in a series he was writing for the *British Medical Journal*.¹ The Institute of Statisticians held a conference on medical statistics in 1981, and we presented a paper on the comparison of two methods of measurements. A revised version appeared in the institute's journal in 1983.² The paper concentrated on the various incorrect methods of analysis widely used in the medical literature, and included a fairly brief

description of our recommended approach. The method was well received, but the paper, as it stood, was too mathematical for applied research workers. Several clinical colleagues suggested we write a more accessible version, including a worked example, for medical readers. J.M. Bland collected some suitable data, asking a nonrandom sample of ourselves, colleagues, friends, and relatives to blow into peak flow meters, and we wrote a paper that presented the analysis of these data in detail, plus other data sets arising from clinical collaborations, and a brief discussion of why correlation is inappropriate. The paper devoted much less space to criticizing incorrect methods and more to a detailed description of how we thought such data should be analyzed. We wanted this to be seen by many people engaged in research into medical measurement, so we sent it to the *Lancet*.

The *Lancet* paper had a much greater impact than the first in the *Statistician*.² It was widely cited by researchers advocating and developing the idea and by those using it in practice, and it was reproduced verbatim in another journal.³ The place of publication was critical. Similar ideas had been suggested in the past, but, like our 1983 paper, they had little impact. Publication in a place where the paper was seen by many workers who were developing and testing methods of measurement led to the paper being used and cited. Since much statistical analysis, reasonably enough, is done by imitating what others have done with similar data, the method was copied and cited further.

Despite the success of the paper, correlation coefficients are still widely used for the comparison of two methods of measurement, sometimes being quoted alongside our approach. Further, our method is sometimes applied inappropriately. Changing established practices can be very difficult.

As applied statisticians working in an academic environment, most of our published work is in the medical literature, yet appointment and promotions committees seem to feel that our careers should depend on publication in statistical journals. The first paper² was therefore much more attractive to us than the second, though if we had stopped there, the *Classic* might never have been born. It shows that for medical statisticians publication in the medical literature may have much greater impact than papers in the statistical literature, and we think promotions committees should be aware of this!

1. Altman D G. Statistics and ethics in medical research. V. Analysing data. *Brit. Med. J.* 281:1473-5, 1980. (Cited 30 times.)

2. Altman D G & Bland J M. Measurement in medicine: the analysis of method comparison studies. *Statistician* 32:307-17, 1983. (Cited 115 times.)

3. Bland J M & Altman D G. Statistical methods for assessing agreement between measurements. *Biochim. Clin.* 11:399-404, 1987. Received June 26, 1992