B. M. Cesana
P. Antonelli
E. Gallazzi
A. Marino

Comparison of measurement methods: an endless application of wrong statistical methods

Accepted: 13 December 2010
Published online: 11 March 2011
© Copyright jointly held by Springer and ESICM 2011

An author's reply to this comment is available at: doi:10.1007/s00134-011-2198-0.

Dear Editor,

We were recently asked by a number of colleagues to perform an agreement analysis between two measurement methods using the statistical methods shown in Becker et al. [1].

At first glance, we were delighted with the correct use of the method proposed by Bland and Altman [2] and with the non-parametric Passing-Bablok regression [3], as well as with the very high number of measurements (dots in the figures). However, further study very quickly led to a change of opinion for the following reasons:

1. The authors used the correlation analysis between the two measurement methods (page 950 on the left), which has to be

considered a “totally inappropriate and incorrect method” [4];

2. The authors did not fulfil the assumption of the independence of the data by considering the 84 (approximately) measurements to have been obtained on 42 subjects instead of only the ten patients actually enrolled. Nevertheless, the referenced paper [2] shows how to deal with replicates (in equal or unequal number for each subject). In particular, it must also be mentioned that the authors did not clearly state the maximal acceptable difference between the two methods that would allow them to draw a conclusion in favour of their agreement.

We would also like to draw attention to the sentence on page 950 “correlation between paired differences of consecutive measurements...” as we believe the phrase “and their means” is missing, which would make this sentence understandable, despite no interpretations having been given. Indeed, a result that is not statistically different allows the conclusion to be drawn that the differences between the two methods do not vary in any systematic way over the range of measurement, ruling out the presence of a proportional error.

Also, the definition of “significant” for the bias (page 950, right bottom) is incorrect, since it is not derived from a statistical significance test; perhaps “clinically relevant” is the appropriate term. In addition, at the end of this page, “Accuracy was

0.52” is a sentence that is absolutely incomprehensible as written and should be justified. Finally, the repeatability estimates of the two methods should be given as is shown in Altman and Bland [2].

References

1. Becker V, Schmid RM, Umgelter A (2009) Comparison of a new device for the continuous intra-gastric measurement of intra-abdominal pressure (CiMon) with direct intra-peritoneal measurements in cirrhotic patients during paracentesis. *Intensive Care Med* 35:948–952. doi: 10.1007/s00134-009-1451-2
2. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160
3. Passing H, Bablok (1983) A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 21:709–720
4. Altman DG, Bland JM (1983) Measurement in medicine: the analysis of method comparison studies. *Statistician* 32:307–317

B. M. Cesana (✉) · P. Antonelli
Medical Statistics and Biometry Unit,
Department of Biomedical Sciences and
Biotechnologies, University of Brescia,
V. le Europa 11, 25123 Brescia, Italy
e-mail: cesana@med.unibs.it
Tel.: +39-03-03717530
Fax: +39-03-03717488

E. Gallazzi · A. Marino
Institute of Anesthesiology and Critical
Care, University of Milan, Milan, Italy