

## Perspectives and Editorials:

## Letter to the Editor

### **Use of Misleading Statistics in Method Comparison Analyses**

To the Editor:

**1** A couple of articles about methods comparison have been published recently in the journal (Anzar et al, 2009; Lin et al, 2009). I am afraid to say that the authors fail to apply adequate statistics; therefore, little can be concluded about the agreement or repeatability of the methods tested. Unfortunately, similar mistakes affect many articles in the area of reproductive biology. The arguments developed below can be found in Bland and Altman (1986, 2003) and references therein.

The use of paired *t* tests and multiple comparison tests in agreement studies is not valid. “Not significantly different” does not equal “agree.” It is easy to realize that the greater the measurement error, the more likely it is that the mean values do not differ significantly. When comparing a very accurate and precise method with an inaccurate and imprecise method, it is perfectly possible that differences were not significant, because of the wide confidence intervals of the imprecise method. Likewise, regression analyses cannot be used for method agreement as showed in the aforementioned studies. Two methods may strongly agree despite the regression equation having intercept  $> 0$  and slope  $> 1$ , the contrary being also true. Furthermore, that situation is normal and even expected. Nevertheless, with a different use, regression can be useful.

A very simple and powerful method for assessing method agreement and repeatability was devised more than 20 years ago (Bland and Altman, 1986) by plotting the differences against the mean of both measurements

(difference plot). An example of use in reproductive sciences can be found in Nagy et al (2003). Of course, other methods might be more adequate, depending on the context (eg, receiver operating characteristic [ROC] curves).

Journals sometimes dedicate reviews or editorials to statistics. I would like to suggest that the *Journal of Andrology* lend attention to method comparison statistics, because it would be a very needed contribution to the area.

*Felipe Martínez-Pastor*  
*ITRA-ULE, INDEGSAL*  
*University of León*  
*Campus de Vegazana*  
*24071 León, Spain*

### **References**

- Anzar M, Kroetsch T, Buhr MM. Comparison of different methods for assessment of sperm concentration and membrane integrity with bull semen. *J Androl.* 2009;30:661–668.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1:307–310.
- Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* 2003;22:85–93.
- Lin CC, Huang WJ, Chen KK. Measurement of testicular volume in smaller testes: how accurate is the conventional orchidometer? *J Androl.* 2009;30:685–689.
- Nagy S, Jansen J, Topper EK, Gadella BM. A triple-stain flow cytometric method to assess plasma- and acrosome-membrane integrity of cryopreserved bovine sperm immediately after thawing in presence of egg-yolk particles. *Biol Reprod.* 2003;68:1828–1835.