Statistical Methods for Comparing Methods of Measurement

To the Editor: We read with interest the recently published article by Fortuna and coworkers[1] on the measurement of the fraction of exhaled nitric oxide (FE_{NO}) using a hand-held analyzer (NIOX-MINO, Aerocrine, Solna, Sweden). This study falls within the literature comparing methods of measurement, or studies of agreement between methods. In this case, the methods compared are quantitative. One aim of this type of study is to evaluate the reliability of new measurement devices.[2] To that end the authors compared readings taken with the NIOX-MINO device to those obtained with standard equipment considered to offer a reference, the N-6008 chemiluminescence analyzer (SIR, Madrid, Spain). They compared means with the Mann-Whitney test and calculated correlations between individual results with the Pearson method. The authors based their conclusion that the device is reliable on the finding of very good correlation between FE_{NO} measurements made with the 2 devices (r=0.92, P=.001) and they suggest a range of reference values with an estimated upper limit set at 2 SDs above the mean.

This statistical method for evaluating agreement between quantitative measures has been shown to be inadequate and the interpretation of the Pearson correlation coefficient as a measure of agreement has also been shown to be erroneous.[3,4] The lack of a difference between means only guarantees that 2 methods are based on the same value, not that all the values obtained are the same. Furthermore, in the particular case of this study, the authors compared means using the Mann-Whitney test, a nonparametric test appropriate when the means are from independent groups. Since the measurements in this study were paired, the authors should have used a nonparametric test for matched pairs, namely the Wilcoxon T test. Although other authors have also used only the Pearson correlation coefficient for the same purpose (References 2, 8, and 11 in the Table), this statistic only indicates strength of association between 2 variables (how close the data points for the 2 methods are to the regression line) and the linearity of their relationship, but not that the line necessarily bisects the set of data; even a perfect correlation cannot be considered synonymous with perfect agreement.[3,4] Nor do

Validation Studies, With the Statistical Method Used, for the Measurement of the Fraction of Exhaled Nitric Oxide

Study[a]	Pearson Correlation Coefficient	Intraclass Correlation Coefficient	Bland and Altman Plot	Bland and Altman Coefficient of Agreement
1	0.97 and 0.98	Yes	Yes	Yes
2	0.97 and 0.98	No	No	No
3		No	No	Yes
4	0.992	No	Yes	Yes
5	0.977	No	Yes	No
6, 7		Yes	Yes	Yes
8	0.99	No	No	No
9, 10	0.94 and 0.96	No	Yes	No
11	0.94-0.99[b]	No	No	No

[a]List of validation studies:

1. Alving K, Janson C, Nordvall L. Performance of a new hand-held device for exhaled nitric oxide measurement in adults and children. Respir Res. 2006;7:67 [accessed 01/07/2007]. Available from: http://www.nioxmino.con/pdffiles/Alving_Janson_Nordvall.pdf y en http://respiratory-research.com/content/7/1/67
2. Nordvall SL, Janson C, Alving K. Agreement between NIOX® (stationary device) and NIOX MINO® (hand-held device) for fractional exhaled nitric oxide (FENO) measurements in adults and children. Actas de World Allergy Congress; 2005, junio 26-29; Munich. Poster session 1, Poster No. 164 [accessed 01/07/2007]. Available from: http://www.nioxmino.com/pdffiles/WAC-poster.pdf
3. Nordvall SL, Janson C, Alving K. Agreement between NIOX and the new hand-held NIOX MINO for FENO measurements in adults and children. Proceedings of the American Thoracic Society 2005 International Conference; 2005, May 20-25; San Diego. Poster Discussion Session, Poster 801, A309 [accessed 03/07/2007]. Available from: http://www.filt.de/Hauptmenue/Akatuelles/Page10464/Abstracts_for_eNO_at_ATS_2005.pdf.
4. Sovijärvi ARA, Järvinen H, Uusi-Simola J, Nikkinen P, Piirilä PL. Accuracy and repeatability of a handheld nitric oxide analyzer in adults with respiratory symptoms. European Respiratory Society Annual Congress; 2005, September 17-21; Copenhagen. Poster 3670 [accessed 01/07/2007]. Available from: http://www.nioxmino.com/pdffiles/ers-poster.pdf
5. Vahlkvist S, Sinding M, Skamstrup K, Bisgaard H. Daily home measurements of exhaled nitric oxide an asthmatic children during natural birch pollen exposure. J Allergy Clin Immunol. 2006;117:1272-6.
6. Turner SW, McGill C, Malik G. Validation of a hand-held exhaled nitric oxide analyser for use in children. Proceedings of the American Thoracic Society 2006 International Conference; 2006, May 19-24; San Diego. Poster presentation [accessed 01/07/2007]. Available from: http://www.nioxmino.com/pdffiles/MINO%20study.pdf
7. McGill C, Malik G, Turner SW. Validation of hand-held exhaled nitric oxide analyzer for use in children. Pediatr Pulmonol. 2006;41:1053-7.
8. Torre O, Spencer A, Olivieri D, Parnes PJ, Kharitonov SA. Feasibility and repeatability of fractional exhaled nitric oxide (FENO) measurements using a hand-held NO monitoring device in asthma in general practice. Proceedings of the American Thoracic Society 2006 International Conference; 2006, May 19-24; San Diego. Volume 3, Abstracts Issue, p. A484 [accessed 01/07/2007]. Available from: http://www.nioxmino.com/pdffiles/ATS_Torre.pdf
9. Nair A, Menzies D, Lipworth BJ. Portable exhaled nitric oxide measurement: comparison with the gold-standard technique. Abstracts, British Thoracic Society Winter Meeting 2006, December 6-8, 2006, London. Thorax. 2006;61 Suppl 2:75. Proceedings of the British Thoracic Society Winter Meeting 2006; 2006, December 6-8; London. Poster P059 [accessed 01/07/2007]. Available from: http://www.nioxmino.com/pdffiles/BTS%20P059.pdf
10. Menzies D, Nair A, Lipworth BJ. Portable exhaled nitric oxide measurement: comparison with the "gold-standard" technique. Chest. 2007;131:410-4.
11. Chen W, Purohit A, Barnig C, Casset A, de Blay F. Niox® and NioxMino®: comparison of exhaled NO in grass pollen allergic adult volunteers. Allergy. 2007;62:571-2.

[b]95% confidence interval of the correlation coefficient.

the authors give confidence intervals for the coefficient. We calculated the 95% confidence interval to be 0.83 to 0.96, implying that the values for the correlation coefficient could be less than 0.90, in which case the correlation would be classified as merely good. Other studies not mentioned by the authors report Pearson correlation coefficients much closer to 1 (Table).

Among the most frequently used procedures for determining agreement between quantitative measures are *a)* analytical methods involving the calculation of various measures of agreement such as the intraclass correlation coefficient or the concordance correlation coefficient of Lin, regression models (of Deming and of Passing-Bablock), analysis of variance of repeated measures, and structural equation modeling[3]; and *b)* graphic methods such as the construction of a Bland and Altman plot[4] (in which the interval of agreement is also calculated), the folded empirical cumulative distribution curve (or "mountain plot"), and the survival-agreement plot.[2,5] Fortuna and coworkers[1] applied none of these methods. A more exhaustive review of the literature (eg, at http://www.nioxmino.com/references.html) shows that several other studies have used these methods (Table). In fact, an important observation based on the Bland and Altman plots in the literature is that the relation between paired FE_{NO} measurements between the devices compared is heteroscedastic, which is to say, the higher the FE_{NO} values, the greater the differences between measurements (References 6 and 7 in the Table).

In addition, as mentioned above, the authors suggest a cutoff value of 2 SDs above the mean to be used for reference. Even if values measured fall within the limits the authors establish, this approach to estimating reference values can only be used if the data are normally distributed.[6] However, they report no statistical test, either analytic or graphic, of normality. Other studies have performed log transformations in order to normalize data for a variable (References 5 and 10 in the Table).

Although Fortuna and colleagues[1] describe the attractive features of the new FE_{NO} measurement device (namely, that it is simple, fast, manageable, and inexpensive) and they note its usefulness for diagnosis and follow-up in asthma, they cannot assume from their analysis that the device gives measurements that agree with the reference method (reliability). Nor can they assume that the reference values they suggest are valid without providing more information about the distribution of the data.

Fortunately, a good study with insufficient analysis can always be re-analyzed. We believe that the above-mentioned information should be brought to light, so that the NIOX-MINO device can be considered validated in a Spanish population. There is a statistical package (MedCalc Software, Mariakerke, Belgium) that incorporates a procedure to compare methods directly. In this program, most of the aforementioned coefficients, regression models, and graphic methods can be implemented without the need for algebraic transformation of the data (whereas such ease of calculation is only possible for the intraclass correlation coefficient in the SPSS program). We encourage the authors to complete their analysis and publish the pertinent data for validating the device—showing its reliability—in a Spanish population.

José-Francisco Pascual-Lledó, Carmen Contreras Santos, and Carlos Martín Carrasco
Unidad de Neumología,
Hospital La Inmaculada, Huércal-Overa,
Almería, Spain

1. Fortuna AM, Feixas T, Casan P. Determinación de óxido nítrico en aire espirado (FENO) mediante un equipo portátil (NIOX-MINO® Aerocrine) en población sana. Arch Bronconeumol. 2007;43:176-9.
2. Pita Fernández S, Pertegás Díaz S. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas [accessed 01/07/2007]. Available from: http://www.fisterra.com/mbe/investiga/conc_numerica/conc_numerica.htm
3. Lluís Carrasco J, Jover LL. Métodos estadísticos para evaluar la concordancia. Med Clin (Barc). 2004;122 Supl 1:28-34.
4. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. The Statistician. 1983;32:307-17.
5. Luiz RR, Costa AJL, Kale PL, Werneck GL. Assessment of agreement of a quantitative variable: a new graphical approach. J Clin Epidemiol. 2003;56:963-7.
6. Wright EM, Royston P. Calculating reference intervals for laboratory measurements. Stat Methods Med Res. 1999;8:93-112.