Letters

Evaluation of a New Desktop Spirometer: Problems With Statistical Methods Used

The study by Swart et al in the June 2003 issue of RESPIRATORY CARE was well done, and the editorial in that same issue by Dillard et al properly addressed some of the problems with patient-based comparisons of new devices.[1,2] However, a few issues remain regarding the statistical analysis of the Swart et al data.

First, the graphs in Figure 1 indicate the statistical analysis to be the coefficient of determination (r^2) and not Pearson's product moment correlation, as reported in the article (as correlation coefficient). This, no doubt, was a typographical error.

Second, the use of Pearson's r should always be accompanied by an alpha value (p), which gives the probability that a correlation of a given size could occur by chance if the 2 variables were not correlated.[3] Although the findings were strongly correlated, this omission casts uncertainty on the results.

Third, although it was appropriate to use the paired Student's t test to determine statistically significant differences, the article did not specify whether a 2-tailed or 1-tailed test was used.

Fourth, by my count and not including tests that should have been done for the Pearson's r, there were 24 comparisons made via t test. Since the Type I error rate increases with the number of comparisons, a correction for multiple comparisons should have been made to compensate. For instance, it would have been more appropriate to use a more conservative alpha value ($p < 0.01$) or apply the Bonferroni approach.[4]

Fifth, it is good practice to report the actual p value (eg, $p = 0.03$) and not simply the alpha level for statistical significance (eg, $p < 0.05$).

Sixth, the authors stated that they found statistical significance in the difference of 4 variables but then stated that the differences were too small to be of clinical relevance. This is typical of a study that does not determine power and sample size *before* the study begins. Proper power and sample size calculations help eliminate differences that are large enough to be statistically significant but too small to be clinically relevant. Interestingly, the authors conclude in the discussion section that their sample size was too small.

Seventh, the article's section on statistical analysis does not mention regression analysis, but the equations for this method are found in the description of Figure 1, leaving one to wonder where they came from.

Individually, these omissions and errors are small. However, in sum they could threaten the validity of the paper and should be addressed to ensure that the integrity of the article and its conclusions remain intact.

Norman Tiffin RRT MSA
Bright's Grove, Ontario,
Canada

REFERENCES

1. Swart F, Schuurmans MM, Heydenreich JC, Pieper CH, Bolliger CT. Comparison of a new desktop spirometer (Spirospec) with a laboratory spirometer in a respiratory out-patient clinic. Respir Care 2003;48(6):591–595.
2. Dillard TA, Reyes J, Johnson M, Baker R. Patient-based studies of agreement between spirometers. Respir Care 2003;48(6):587–588.
3. Lang TA, Secic M. How to report statistics in medicine. Philadelphia: American College of Physicians; 1997:102.
4. Petrie A, Sabin C. Medical statistics at a glance. Malden, Massachusetts: Blackwell Science; 2000:45.

The authors respond:

Norman Tiffin points out some statistical shortcomings of our report, and we agree with most of the comments he made.

1. The correlation coefficient in the Figure 1 should indeed have been abbreviated as r^2.

2. All p values of the regression models were < 0.001. Whether one should mention these p values routinely in cases with such high correlation coefficients is debatable, as it is self-evident that they are highly significant given the sample size.

3. Two-tailed tests were used, as the tested machine could have over- or under-estimated the true value.

4 and 5. These are valid comments; those methods were not respected in our report. However, the decision as to which p value should be used is controversial. If the administration of the test would have dire consequences, a smaller p value is indicated, which is not the case in this report.

6. This point is well taken; however, we maintain that regardless of sample size the absolute differences observed are not clinically important in the variables that were significantly different. Power or sample size calculations are certainly an essential component of a well-planned study.

7. Regression analysis was performed, despite the fact that it was not explicitly mentioned in the statistical analysis section. The lack of that information does not really affect the statistical veracity of the article; it rather draws attention to common shortcuts taken in statistical reporting.

In summary Mr Tiffin's criticisms are very useful and we are grateful he raised them.

Macé M Schuurmans MD
Clarissa H Pieper MD
Chris T Bolliger MD
University of Stellenbosch
Tygerberg, Cape Town
South Africa

CORRECTION

Table 2 in "Comparison of a New Desktop Spirometer (Spirospec) with a Laboratory Spirometer in a Respiratory Out-Patient Clinic" by Swart et al (Respir Care 2003;48[6]:591–595) is in error:

The correct peak expiratory flow (PEF) values (mean $\pm$ SD) in the restrictive subgroup are 6.2 ± 2.1 L/s and 6.7 ± 2.3 L/s for Spirospec and Masterlab 4.0, respectively.