Patient-Based Studies of Agreement Between Spirometers

Patient-based studies of agreement between spirometers arguably have a place in evaluating and accepting new equipment into clinical use. These studies typically trail in credibility behind more controllable simulations such as computer-driven benchmark wave forms,[1] other bench techniques, and comparisons with normal subjects. These latter methods would be expected to give more consistent input data than patients. Still, proponents argue that patient-based studies add useful technical product information. This view appears to hold, as judged by the relatively frequent appearance of patient-based studies in peer-reviewed literature. A brief review of some of the pros and cons of patient-based studies of agreement between spirometers appears worthwhile and timely with the publication of a study by Swart et al in this issue of RESPIRATORY CARE.[2]

SEE THE ORIGINAL STUDY ON PAGE 591

Swart et al report a comparison between a recently marketed desktop spirometer (Spirospec) and a laboratory spirometer (Masterlab 4.0). They studied 45 patients: 15 with obstruction, 15 with restriction, and 15 with normal spirometry. After comparing the best values for each spirometry variable from each device, the authors reported no differences for forced expiratory volume in the first second (FEV_1) or forced vital capacity (FVC) but found statistically significant differences in the best values for peak expiratory flow (PEF), peak inspiratory flow, and forced expiratory flow at 50% and 75% of the FVC (FEF_{50} and FEF_{75}). The Spirospec values were lower than the Masterlab 4.0 values for PEF and higher for FEF_{50}, FEF_{75}, and peak inspiratory flow. Swart et al provide the customary presentation of statistical data and conclude that the observed differences in flows probably have little clinical importance for patient management.

During the peer review process the authors added descriptive statistics for the differences between the best 2 efforts for spirometry indices on each instrument. With that addition they provided a means to assess the amount of patient variability with each instrument and to weigh the amount of patient variability against the difference of best values between the instruments. This makes a positive contribution to methodology for patient-based studies. Taking this analysis to the ultimate conclusion by a statistical test remains a desirable goal for patient-based studies. This task was not asked of the authors, in part because an accepted template for this type of analysis remains to be developed.

The preceding text conveys a concept that the difference of best values between instruments includes contributions from both patient variability and instrument variability. In order to confidently make inferences about agreement between instruments, patient variability should be taken into account. Patient-based studies typically report the difference of best values but seldom report a measure of patient variability.[3] This causes little concern when no difference in best values occurs, which is often the case. The conclusion may be reached that both instruments measure the variable about the same: a comfortable finding. The possibility remains, however, that patient variability obscures a true difference between the instruments.

Authors may struggle to explain a difference in best values between instruments when they find one. Some may erroneously attribute all of the difference to the instruments. In our view, without formally accounting for patient variability one cannot know how to explain the difference.

A semi-quantitative approach consists of comparing the standard deviation of the difference of best values between instruments to the standard deviations of the differences between the best and second-best efforts within each instrument. If the standard deviation between the instruments sufficiently exceeds the standard deviations within instruments, then a true difference between the instruments may be inferred. This semi-quantitative approach may aid interpretation of patient-based studies. The best statistical approach to this problem appears to be an application of analysis of variance (ANOVA).

Table 1 presents a proposed method for data entry to permit statistical testing of the above concepts. The purely illustrative table contains 4 rows for each patient. The 4 rows represent the best 2 efforts on each instrument. The categorical variables Subject, Device, and Effort represent grouping variables that allow ANOVA (with most statistics software packages). The values coded as Effort = 1 reflect the highest values on each instrument for FVC, FEV_1, PEF, and peak inspiratory flow. For mid-flows the values for Effort = 1 would usually be taken from the effort with the highest sum of FEV_1 + FVC. For Effort = 2, mid-flow values should be taken from the effort with the second-highest sum of FEV_1 + FVC. These assignments may give pause to some statisticians but in our opinion are the best approach for patient spirometry.

Table 1. Proposed Data Entry Method to Permit Statistical Testing in Comparing Spirometers

Subject	Device	Effort	FEV_1	FVC
1	1	1	2.2	3.2
1	2	1	2.1	3.1
1	1	2	1.9	2.8
1	2	2	1.8	2.9
2	1	1	2.3	3.1
2	2	1	2.2	3.2
2	1	2	2.0	2.9
2	2	2	1.9	2.8

The values are examples only: not actual patient data.
FEV_1 = forced expiratory volume in the first second.
FVC = forced vital capacity.

A key statistical test consists of ANOVA for a variable such as FEV_1 by device. Further exploration of variance by device, by effort, and by subject would be appropriate to isolate the source of variance. It should be noted that this proposed model is a generalization and may require modification for specific statistics software packages and confirmation by empirical experimentation. We hope that this model will stimulate further efforts to isolate the instrument difference. Statistical approaches such as a mixed model, repeated-measures ANOVA, or intra-class correlation (reliability tests) using the method of moments may be more appropriate, but may be difficult to perform.

Another limitation of patient-based spirometry studies is the criteria for acceptable performance of FEV_1 and FVC for clinical testing. Acceptable agreement between duplicate efforts for FEV_1 and FVC (ie, $<$ 200 mL difference) are less stringent than the required agreement between daily 3-L syringe calibration measurements (ie, difference of 50 mL or 3%, whichever is greater, for diagnostic spirometers).[1] The allowable difference is larger (100 mL or 5%) for monitoring devices. For FEV_1 and FVC it will always be a challenge for patient-based studies to show a true difference between instruments, unless the patients are held to a higher standard than the American Thoracic Society criteria.[4] As a practical matter, close agreement of the values within instruments will more likely allow a difference between instruments to emerge. Trying to achieve that may entail more spirometry efforts, which may not be advisable for many patients, and also raises additional issues about the use of human subjects.

Patient-based studies most often show differences in air flows (eg, PEF, forced expiratory flow in the middle half of the forced vital capacity [FEF_{25-75}], or peak expiratory flow) rather than volumes. This may reflect the lack of available direct flow calibration input signals in clinical laboratories. Though not practical for daily calibration, standard computer-generated waveforms are available for validating PEF.[1]

Patient-based studies should and typically do reflect a range of clinically relevant spirometry results. Clinical relevance may not always apply to studies of normal subjects or benchmark waveforms that fall outside a range of values of clinical interest. Some combinations of flow and volume results observed in patients may be relatively unique and not readily simulated by other means. Another advantage of patient-based studies is low cost. Patients presumably need to have at least one spirometry done for clinical practice, so only the extra spirometry efforts would add to costs. Still another advantage is that patient-based studies can be done by multiple investigators at multiple sites.

Patient-based studies are more likely to reproduce real-time clinical testing than more formal studies done under ideal conditions. Although this may seem like an advantage, the greater likelihood of intrasubject variability with patients actually imposes a study limitation. The presence of disease and the susceptibility to exacerbation or malaise make patient efforts more difficult to reproduce than flows from normal subjects, syringes, or benchmark waveforms.[5]

This discussion of selected pros and cons of patient-based studies of spirometer agreement has sought to recognize Swart et al for including patient variability in their methods and results. In addition we have suggested means for improving the ability to make inferences on observed differences. Hopefully, direct reporting of patient variability to isolate instrument variability will become more common and statistical approaches further refined in the future.

Thomas A Dillard MD
Juan Reyes RRT
Maribeth Johnson MSc
Randy Baker PhD RRT
Medical College of Georgia
Augusta, Georgia

REFERENCES

1. Standardization of spirometry, 1994 update. American Thoracic Society. Am J Respir Crit Care Med 1995;152(3):1107–1136.
2. 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.
3. Maree DM, Videler EA, Hallauer M, Pieper CH, Bolliger CT. Comparison of a new desktop spirometer (Diagnosa) with a laboratory spirometer. Respiration 2001;68(4):400–404.
4. Schoh RJ, Fero LJ, Shapiro H, Aslor JP, Kaelin OJ, Rollins DR, Petty TL. Performance of a new screening spirometer at a community health fair. Respir Care 2002;47(10):1150–1157.
5. Stoller JK, Orens DK, Hoisington E, McCarthy K. Bedside spirometry in a tertiary care hospital: the Cleveland Clinic experience. Respir Care 2002;47(5):578–582.

Correspondence: Thomas A Dillard MD, Section of Pulmonary and Critical Care Medicine, Medical College of Georgia, 1120 15th Street, Room BBR 5513, Augusta GA 30912-3135.