

EDITORIAL

Performance testing new peak flow meters

O.F. Pedersen*, M.R. Miller[†], T.W. van der Mark[‡]

The increasing incidence of asthma and the subsequent need for devices to monitor peak expiratory flow (PEF) and other lung function indices has created an expanding market for monitoring devices. In this issue of the *European Respiratory Journal*, RICHTER *et al.* [1] present data from a new device. From a commercial point of view there are obvious reasons for doing this, if it can be shown that the device fulfils the generally accepted quality criteria. By presenting these data as a technical note in the *European Respiratory Journal*, a key target group will be reached in an effective way and this may help to keep users of peak flow meters and monitoring devices adequately informed about the performance and quality of new equipment.

Several important questions arise: what is a satisfactory performance, how is it best documented, and who should undertake the tests? Both the American Thoracic Society (ATS) and the European Respiratory Society (ERS) have developed guidelines [2, 3] defining performance criteria for the evaluation of spirometers. The following aspects might enter into a comprehensive evaluation: performance assessment using mechanical test rigs, clinical testing, tests of patient compliance and ease of use, hygiene and maintenance, and long-term reliability. Eventually, the responsibility for the function of the meter should be placed with the manufacturer, the distributor, the doctor and the users in such a way that ensures optimal utilization in a clinical setting. Can the manufacturer be relied upon to produce all of these test results, or should they be obtained by an independent institution?

With regard to the accuracy of measurements, there are three main elements, which can all be assessed using a test rig. The first is calibration for steady flows, which may mean that the meter's scale would not necessarily be equidistant [4] in order to give the best accuracy. The next element is the device's frequency response, which determines whether or not the meter can react quickly enough to measure the fastest components of the expiration. Rapidly occurring events, such as PEF, may be underestimated if the device's response is overdamped, or may be overestimated if the device is underdamped. The third element is the resistance of the meter, which has to be low so that it does not cause any gas compression in the lungs such that PEF is lowered significantly [5]. The calibration of a device and the determination of its internal resistance are straight-forward. The difficult part is the determination of the frequency response. The ATS [2] has approached this empirically by defining an adequate frequency response if the meter reacts satisfactorily to each one of a panel of 26 waveforms which have arbitrarily been deemed to reflect the common range of flow profiles. The ERS [3] has proposed another approach by instead considering that the

frequency response of PEF meters be tested with flow-time profiles having rise times (time from 10–90% PEF) and dwell times (duration of flow in excess of 95% or 90% PEF) that are representative of the range found in normal subjects and those with airflow limitation.

Considerable technical problems must be overcome in undertaking these tests. Using a servo-controlled calibration pump to generate flow profiles with very short rise times leads to considerable distortion of the output profile compared with the input profile [6, 7]. Thus it is difficult for the true output flow to be determined accurately. This problem with profile distortion is not the case if the calibration flow is produced by explosive decompression from a pressure tank [8]. The disadvantage now is that this output flow has to be calibrated in advance, because it is not uniquely determined in the same way that the input flow from a servo-controlled pump is derived from its dimensions and the movement of the piston. No clinical testing should be necessary, unless a meter has certain features that suggest that a patient's interaction with the meter might be different from that for other types of meter. There is a good chance that two different meters will give the same result if the calibration is correct, the frequency response is adequate and the internal resistance is not too high.

In the paper by RICHTER *et al.* [1] the meter was meticulously tested and found to give results in accordance with the ATS recommendations. It is debatable whether the arbitrary panel of 26 ATS flow waveforms is rigorous enough to cover the full range of short rise times found in both normal subjects and patients [9, 10]. The median value for 10–90% rise time to PEF in a population study was found to be 72 ms for both men and women with normal spirometry, and was 46 ms and 50 ms, respectively, for patients with airflow limitation [10]. Therefore, any pump system must be able to deliver a profile with sufficiently short rise time.

The goal for the manufacturer is to make meters that meet the agreed standards. The goal for the specialist in the field is to be sure that the standards are appropriate. A typical recent remark from a manufacturer is: "We are very well aware of the fact that the Europeans have different standards than the Americans. We now have equipment that satisfies the ATS standard, and we work on a different set up to satisfy the ERS standards". The ideal would be to obtain a single standard based on a true scientific approach, which is satisfactory to both the ATS and the ERS. Presently meters can be obtained with different scales that deviate considerably from each other. The original Wright scale, which is equidistant, gives readings that overestimate true flow by 40–80 L·min⁻¹ at 300 L·min⁻¹ [4] and underestimate by 20–50 L·min⁻¹ at 720 L·min⁻¹. A correct reading requires a nonequidistant scale that can now be obtained, but it is not always clearly indicated, which scale is on the meter.

We do not think that meters should be tested with a large panel of arbitrarily chosen profiles but we instead favour a simpler test regimen based on data from a large population. We believe that the PEF meter can be tested

*Institute of Environmental and Occupational Medicine, Aarhus University, Aarhus, Denmark. [†]Dept of Medicine, University of Birmingham, Birmingham, UK. [‡]Dept of Pulmonology, University Hospital Groningen, Groningen, The Netherlands. Correspondence: O.F. Pedersen, Institute of Environmental and Occupational Medicine, Aarhus University, Universitetsparken, Bygning 180, DK 8000 Aarhus C, Denmark. Fax: 45 89422970.

adequately using either constant flow profiles, cusp profiles [4] or profiles with long rise times in excess of 150 ms. The frequency response should then be determined using an additional profile with a rise and 90% dwell time at the lower 5th percentile of that found in patients with airflow limitation (24 ms and 10 ms, respectively [10]). This profile is then scaled to deliver a PEF at the top and bottom of the working range of the meter. If a meter was accurate with these tests then it should faithfully record PEF from over 95% of subjects.

Although a clinical comparison may not always be necessary in the development of new monitoring devices, such a comparison was done by RICHTER *et al.* [1]. Using a test rig, the authors found a mean difference between the devices tested when connected in series of -11.3 mL for the forced expiratory volume in one second (FEV₁), with the 95% limits of agreement ranging from 41.5 to 18.9 mL. However, in a similar experiment involving humans the difference in FEV₁ was 166 mL, with 95% limits of agreement ranging from -112–444 mL, an order of magnitude larger than the *in vitro* values. This important result could be due to differences between the two tests in the shape of the flow profiles, the temperature and the gas composition, or the effect of the series relationship between the two devices. These possible explanations could be further examined on a test rig.

A generally accepted design for clinical testing is not yet available. However, aspects concerning the test parameter and its dependence on the intended use of the device, the sample size, the method of analysis and ethical considerations are worthy of mention.

The parameter to be tested should preferably be compared in the setting for which the new device is designed. If it is to replace conventional spirometry, then the normal setting, procedures and patients in a particular clinic or hospital should be chosen. If the aim is to assess whether patients can use it in an unsupervised situation at home, parameters other than PEF may be important, for instance, the number of trials and number of failures. If the aim is to investigate whether the new device could detect diurnal variations in PEF, patients for such a comparison should be selected carefully.

When assessing very small differences, or proving that there is no difference, particular attention should be paid to type II errors. In these cases a conventional power of 80% and significance of 5% is insufficient and a power of 95% is recommended. Furthermore, the distances between measurements and different devices is a question of clinical judgement and accepted standards, which should ideally be agreed in advance. The defined accepted tolerance, the variation in the measurements, and the statistical power all determine the actual sample size. In general, the number of patients required in these comparisons is very large.

The preferred statistical method for assessing agreement between two methods of clinical measurement is that of BLAND and ALTMAN [11]. The reasons for this are very well described in their paper. What one needs to determine is by how much one method is likely to differ from the other, *i.e.* the limits of agreement must be determined. Conventional regression analysis indicates agreement if, and only if, the intercept of the regression line is zero and its slope is one. The degree of agreement and its limits are not given by the correlation coefficient.

Concerning ethical aspects, one must realize that all work involving studies on human subjects is, according to *European Respiratory Journal* instructions, expected to

have received approval from local Ethical Committees. It may seem a little excessive to ask the local Committee approval for standardized lung function procedures. However, performing more than one spirometric procedure can be very exhausting, particularly in older patients. Besides these considerations, an Ethical Committee would also require that safety precautions have been taken and that the new device is properly registered (CE-mark) according to European Union regulations.

The justification for more publications describing conventional and new PEF spirometric devices in the *European Respiratory Journal* must be either that some new scientific dimension is added to our knowledge, or that special features should be described, *e.g.* new principles of measurement, or new and special applications. As short technical notes these descriptions may have considerable value. Presently the confusion of scales maybe the most important problem. Comparing meters with different scales without correction [12] is misleading, as pointed out in a letter to the Editor by I. Gregg in this issue of the *European Respiratory Journal*. The challenge now is to ensure that the standards we set and the test methods we use are both valid and appropriate. Perhaps then we can agree international standards and test methods that are identical on a worldwide basis.

References

1. Richter K, Kannies F, Mark B, Jörres RA, Magnussen H. Assessment of accuracy and applicability of a new electronic peak flow meter and asthma monitor. *Eur Respir J* 1998; 12: 457–462.
2. Standardization of Spirometry 1994 Update. Official statement of the American Thoracic Society. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
3. Quanjer PH, Lebowitz MD, Gregg I, Miller MR, Pedersen OF. Peak expiratory flow: conclusions and recommendations of a working party of the European Respiratory Society. *Eur Respir J* 1997; Suppl 24: 2s–8s.
4. Miller MR, Dickinson SA, Hitchings DJ. The accuracy of portable peak flow meters. *Thorax* 1992; 47: 904–909.
5. Pedersen OF, Pedersen TF, Miller MR. Gas compression in the lungs decreases peak expiratory flow depending on resistance of peak flow meters. *J Appl Physiol* 1997; 83: 1517–1521.
6. Navajas D, Roca J, Farré R, Rotger M. Gas compression artefacts when testing peak expiratory flow meters with mechanically-driven syringes. *Eur Respir J* 1997; 10: 901–904.
7. Hankinson JL, Reynolds JS, Das MK, Viola JO. Method to produce American Thoracic Society flow-time waveforms using a mechanical pump. *Eur Respir J* 1997; 10: 690–694.
8. Pedersen OF, Naeraa N, Lyager S, Hilberg C, Larsen L. A device for evaluation of flow recording equipment. *Bull Eur Physiopath Respir* 1983; 19: 515–520.
9. Pedersen OF, Rasmussen TR, Kjaergaard, Miller MR, Quanjer PhH. Frequency response of variable orifice type peak flow meters: requirements and testing. *Eur Respir J* 1995; 8: 849–855.
10. Miller MR, Pedersen OF, Quanjer PhH. The rise and dwell time for peak expiratory flow in patients with and without airflow limitation. *Am J Respir Crit Care Med* 1998 (in press).
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i: 307–310.
12. Folgering H, van den Brink W, van Heeswijk O, van Herwaarden C. Eleven peak flow meters: a clinical evaluation. *Eur Respir J* 1998; 11: 188–193.