

EDITORIAL

The role of reference values in interpreting lung function tests

R.O. Crapo

The paper by GARCIA-RIO *et al.* [1] in this issue of the *European Respiratory Journal* provides good reference values for an under-represented population and an opportunity to review the role of reference values in interpreting lung function tests. The interpretation of all observed data involves comparisons with one or more types of reference data. In clinical medicine, the comparisons usually involve reference data from those with and without relevant diseases [2–4]. They may also be based on studies relating a clinical measurement to risk of disease. An elevated cholesterol level might simultaneously indicate an increased risk of cardiovascular disease and fall within the range of values obtained on individuals who met health criteria at the time of a study.

Observed data in medicine can come from many sources: clinical information from a patient interview, physical examination and laboratory data (including pulmonary function tests). Reference comparisons usually start with a comparison to data from individuals without relevant diseases ("normal" or healthy subject values). If a patient's data fall outside an appropriate reference range (are not "normal"), the next comparisons are to data from individuals with relevant diseases (disease patterns). The comparisons can be made in a variety of ways: intuitive (based on clinical experience), knowledge or uncertainty (*e.g.* with validated evidence-based criteria), formal algorithms, or reasoning based on knowledge of anatomy and physiology [3]. Whatever the question and whatever the data, the analysis of observed data always involves relating it to reference data [2–4].

A simple example would be a patient who presents to a physician with symptoms that lead the physician to order pulmonary function tests. In that process, the observed data are the patient symptoms and the physical findings which, compared to the physician's knowledge of and experience with healthy subjects, suggest the patient may not be healthy and, moreover, that the pattern suggests a lung problem. Those original steps involve several reference comparisons, leading to a beginning differential diagnosis. Should the pulmonary function test results not fit into the distribution or reference range for healthy subjects, an additional series of disease pattern comparisons will further characterise the patient. For pulmonary diseases these patterns include asthma, central airway obstruction, chronic obstructive pulmonary disease and interstitial lung disease. However, this process is rarely a formal one.

The starting point for the interpretation of lung function tests is comparing the measured values with average values from a representative sample of healthy subjects for which a reference range has been determined (for pulmonary function tests, the acceptable range is typically defined with a 95% confidence interval or a 95th percentile). The comparisons will be most accurate if patient and reference values are comparable in terms of both biological variability and analytical

compatibility [3–5]. Ideally, the patient would resemble the reference population in all aspects other than those being investigated.

Work by BECKLAKE and WHITE [6], and the American Thoracic Society interpretative statement [7] have neatly summarised the sources of variation in lung function tests. These include technical variability due to instrument and procedures, and interactions between the patient, instrument and technician. Technical sources of variation can be as small as 3% when they are carefully controlled but can be overwhelming when they are not. To reduce technical variability, respiratory societies have emphasised standardising the instruments and procedures in lung function testing. Biological sources of variability include, but are not limited to, height, weight, age, sex and ethnic origin [6, 7]. Approximately 29% of the variability remains unexplained, but is probably due to illnesses, exposures, socioeconomic factors and possibly chronobiological factors [6].

Although much less emphasis has been given to selecting and using appropriate reference values, this is every bit as important as measuring lung function accurately and precisely. The ideal reference comparison is a measured value obtained on an individual when he/she is healthy. Such a reference allows the comparison to be switched from group to self, reducing the source of variability in the comparison by about 75% (from $\pm 20\%$ to $\pm 5\%$). A measurement on a healthy identical twin might be the second-best comparison. Neither of these is practical. Measurement of baseline values is usually reserved for individuals at risk for respiratory illness or injury, including those working in occupations involving potentially harmful exposures or those scheduled for medical treatments that involve a risk of lung injury.

The imperative then becomes providing reasonable comparability between your lab's measurements and your reference values. Ideally, each laboratory would perform its own reference value studies and repeat them at some regular interval. This solution is also obviously impractical. Laboratories are left to select among published reference value sets and transfer them to their laboratory [3, 4]. The conditions for transferability are that the reference populations are adequately described and reasonably match the clientele of the laboratory. In addition, the analytical aspects (pre-analytical issues, instrument performance and procedural variables) should agree. The development of standards for pulmonary function tests allows a relatively direct method for establishing technical comparability. If the laboratory and the reference set both use established standards, the technical comparability criteria are met.

Matching populations for the sources of biological variation is more difficult. That process has been made easier in the USA by the release of the NHANES III reference data for spirometry based on a large random sample from the USA with over sampling of two large ethnic groups (Mexican Americans and African Americans) [5, 8]. The study was large, the analytical conditions matched current standards well and data for three ethnic groups were provided. Among

its limitations, NHANES III does not provide information for Asian Americans or for children aged <8 yrs.

It has been suggested that a sample of data obtained from healthy subjects be compared to the reference data to confirm the validity of the choice of reference values [3, 7]. This also appears to be impractical. We analysed the performance of the Hankinson and Knudson reference equations on our healthy subject data set and could not clearly identify spirometry differences (which were, on average, about 4%) until our sample size reached 100 [8–10]. In contrast, differences between some carbon monoxide diffusing capacity of the lung equations were found with sample sizes as small as 15–20.

In considering reference values, much attention must be given to selecting appropriate reference ranges. For most large reference sets, these confidence intervals are set statistically using either Gaussian distributions, when the distribution is clearly Gaussian. Percentiles are used when the distributions are not Gaussian. Fixed limits, such as 80% of predicted or a forced expiratory volume in one second/forced vital capacity ratio >70%, are often used for simplicity but by their very nature increase interpretative errors. Although they work relatively well for average individuals, they breakdown quickly as height and age diverge from the average.

The choice of reference values, the choice of lower limits of normal or confidence interval, and the interpretative scheme are all as important as the accuracy and precision of the original measurement. Unfortunately, attention to reference values is sometimes minimal or overlooked entirely. In a survey of pulmonary training programmes, a few laboratories were unabashed to admit they did not know what reference values they used [11]. They probably used their instrument defaults as if they had no choice or the choice wasn't important. Conversely, reference values are sometimes selected to influence the interpretation. I have personally observed a situation where reference values were picked to increase the frequency of abnormalities in individuals being screened for asbestosis.

The uncertainties in these issues emphasises the point that one can be relatively confident of comparisons with the

healthy subject reference values when a patient falls well within or well outside the boundaries, but that values near the boundaries should be approached with caution. Other information such as the clinical presentation, which will alter the prior probability of disease are vitally important in understanding the meaning of borderline values.

References

1. Garcia-Rio F, Pino JM, Dorgham A, Alonso A, Villamor J. Spirometric reference equations for European females and males aged 65–85 years. *Eur Respir J* 2004; 24: 397–405.
2. Grasbeck R. Health as seen from the laboratory. In: Reference Values in Laboratory Medicine. The current state of the art. Grasbeck R, Alstrom, eds. Chichester, Wiley, 1981; pp. 279–288.
3. Solberg HE, Grasbeck R. Reference values. *Adv Clin Chem* 1989; 27: 1–79.
4. Grasbeck R. Reference values, why and how. *Scand J Clin Lab Invest* 1990; 50: Suppl 201, 45–53.
5. Becklake MR. Concepts of normality applied to the measurement of lung function. *Am J Med* 1986; 80: 1158–1164.
6. Becklake MR, White N. Sources of variation in spirometric measurements. *Occup Med* 1993; 8: 241–264.
7. ATS Statement Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991; 144: 1202–1218.
8. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
9. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal expiratory flow volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725–734.
10. Jensen RL, Crapo RO, Flint AK, Howell HM. Problems in selecting representative reference values for spirometry. *Am J Respir Crit Care Med* 2002; 165: A200.
11. Ghio AJ, Crapo RO, Elliott CG. Reference equations used to predict pulmonary function. *Chest* 1990; 97: 400–403.