

Editorial

Standardised lung function testing

Quality control is an essential ingredient of all clinical laboratory work and it has been difficult to achieve in clinical respiratory physiology. Tests of respiratory function call for the active cooperation of an untrained individual who has to perform quite complex manoeuvres on request. To reach the position where different laboratories would produce the same result on the same patient, a number of conditions have to be met. First, the respiratory manoeuvres requested must be identical. Secondly, the apparatus must respond in the same way to the same stimulus. Thirdly, the results have to be calculated in a standard way and expressed in conventional units. Finally, the reference or normal values have to be agreed between the laboratories, if estimates of the likelihood of normality are to be provided.

The European Coal and Steel Community (ECSC) has for a long time sponsored work into pulmonary function as part of long term studies of occupational lung disease and chronic bronchitis. As a result, a working party was set up which was concerned with the standardisation of pulmonary function tests. Recommendations on the practice of spirometry were published in 1971.¹ Since then further tests have gained wide acceptance in epidemiology and the SI system of units has been introduced. It became apparent that there was a need for harmonisation of techniques and of nomenclature if comparable studies were to be made in different centres. The working party recognised that harmonisation was not an end in itself and that it would probably be effective only for well established routine tests. They therefore set out to review current techniques, concerning themselves mainly with those tests which are applied frequently, departing as little as possible from accepted practice in Europe and the United States. July 1981 was the date set for the publication of the first draft. This was circulated widely and suggestions from a number of experts were incorporated into the final version. The result is a lengthy report entitled "Standardized Lung Function Testing," which has been published as a supplement of the *European*

Bulletin (Clinical Respiratory Physiology).² The authors, who are all notable experts in this field, have tackled all of the problems of standardisation and produced a series of recommendations which are required reading for everyone who is concerned with the organisation of a clinical respiratory laboratory. The recommendations are concerned firstly with the performance of the tests and secondly with the provision of reference values.

As the lack of accepted standards has been a major problem for epidemiologists concerned with lung disease, the National Heart and Lung Institute (NHLI) in the United States also embarked on an epidemiology standardisation project, eventually published in full in 1978.³ This document was concerned with the preparation of standard questionnaires and the reporting of chest radiographs as well as with pulmonary function, but a succinct chapter was included which dealt with spirometry, diffusing capacity and lung volumes. This study is interesting to read because the authors have taken great pains to discuss the reasons why they have chosen one course rather than another and because new work was undertaken to solve some of the problems that emerged during the preparation of the document. Fortunately, the ECSC working party have generally made the same recommendations, though they have taken rather less trouble to explain their conclusions. The task of reading the sections which deal with individual tests is somewhat daunting. To make their recommendations comprehensible, the authors have provided considerable detail about how they think the standard tests should be carried out. Even so, the work is not a complete manual of lung function testing, and the summary provided unfortunately does not tally in every detail with the contents of the principal chapters. The purpose of this review is to examine the consequences of following the recommendations in the ECSC document.

Static lung volumes**VITAL CAPACITY**

Vital capacity is defined as the volume change at the mouth between the positions of full inspiration and full expiration. The Working Party have proposed that expired vital capacity (EVC) and inspired vital

Address for reprint requests: Dr G Laszlo, Respiratory Department, Bristol Royal Infirmary, Bristol, BS2 8HW.

capacity (IVC) should be measured separately and specified for certain situations.

Inspired vital capacity is greater than expired vital capacity in patients with chronic airflow obstruction and it is therefore recommended (1) for the calculation of FEV_1/VC , the Tiffeneau index, and (2) for the calculation of residual volume and total lung capacity: $RV = TLC - IVC$

It is true that because of the difference between forced vital capacity EVC and IVC in patients with chronic airflow obstruction, the ratio FEV_1/VC is a better discriminator than FEV_1/FVC and is correlated with the severity of emphysema in patients with chronic airflow obstruction.⁴ EVC can be measured on portable equipment, while IVC requires a more elaborate spirometer. In normal subjects and patients without airflow obstruction, FVC is usually greater than EVC and identical to IVC.

LUNG VOLUMES: CHOICE OF METHOD

The section on lung volumes reflects the change in opinion during the preparation of successive drafts which has arisen as a result of the demonstration of serious problems with the use of body plethysmography in patients with airflow obstruction.^{5,6} Body plethysmography is recommended as the method of choice in patients with chronic airflow obstruction but with a warning of the possible overestimate which may occur in the presence of airway closure when this method is employed.

Radiographic methods give similar results to body plethysmography and closed circuit inert gas dilution methods in subjects with normal intrapulmonary gas mixing.^{7,8} In a series of 19 healthy controls, a small positive difference was accounted for entirely by those who had smoked cigarettes.⁹

The average difference between residual volume measured by closed circuit helium dilution and by the single breath method with 10 seconds of breath holding is -0.11 ± 0.7 (2 SD) in patients with lung diseases not characterised by airflow obstruction in whom FEV_1/VC is normal (Laszlo, unpublished observations). In 80% of patients with chronic bronchitis and emphysema, however, the difference is greater than 0.7 l.¹⁰ There is a corresponding difference between TLC measured radiographically and by closed circuit helium dilution using an equilibration time of up to 10 minutes. The magnitude of the latter difference is dependent on the severity of the airflow obstruction and on the topographical distribution of emphysema, peripheral lesions apparently filling with helium more slowly than central emphysema.^{4,9} The Working Party concluded that more research was needed into use of forced breathing during closed circuit

helium dilution in the estimation of lung volumes and into possible modifications to plethysmographic methods to achieve accurate results in patients with chronic airflow obstruction. It seems likely that a technique which involves panting near TLC at controlled low frequencies will be shown to produce a good estimate of thoracic gas volume even in the presence of quite severe airflow obstruction.^{5,11} Meanwhile, the best starting point at present for calculating the subdivisions of lung volume in these patients is probably the radiographic measurement of TLC,^{8,12} provided that there is no thoracic deformity.

TECHNIQUE

For body plethysmography the mean of three estimations of FRC is recommended. Then: $TLC = FRC + \text{inspiratory capacity (IC)}$ (largest of 3) and $RV = TLC - IVC$. For closed circuit helium dilution, only one measurement is recommended. Unfortunately, there are a number of sources of potential error in this section. Firstly a rebreathing time of up to 10 minutes is suggested, which may be cut short if the helium reading is stable for two minutes. This conforms with common practice elsewhere, it is stated that 20 minutes is required for full equilibration when there is severe chronic airflow obstruction, but experience suggests that this is rarely satisfactory because leakage is common with long rebreathing times and there is a theoretical possibility that helium may be swallowed.

Secondly it is suggested that attempts to stabilise the trace should be abandoned after one minute and the effect of drift corrected by a formula. This is acceptable in subjects with rapid intrapulmonary mixing, if the katharometer reading can be corrected for the error caused by the change of inspired oxygen concentration. The slow component of helium equilibration is normally due to unwanted effects such as tissue uptake. The method of correction recommended completely removes the slowly ventilated compartment and no data are given to show how this gives a result for lung volume which is different from the erroneously low value obtained by the single breath 10 second breath holding method in patients with chronic airflow obstruction.

For the measurement of the subdivisions of lung volume, the report suggests two permissible sequences, starting from FRC. The first is to breathe to RV, followed by a full inspiration to TLC. This method permits the calculation of all the subdivisions of lung volume in one breath: $RV = FRC - ERV$ and $TLC = RV + IVC$. This is, however, given as the alternative method for obtaining TLC, the preferred method being to breathe first to TLC then $TLC = FRC + IC$ as for the body plethysmo-

graph. The recommended method requires two separate breaths for a single estimation of lung volumes. This seems inelegant, and no very good reasons are given for adopting it, except that the same sequence of breaths is then employed as in body plethysmography.

Open circuit procedures and other methods of measuring lung volumes are not recommended.

Dynamic lung volumes

FORCED EXPIRATION

FEV₁ and FVC are the starting point for the estimation of lung function. Other timed subdivisions of the flow-time curve are not recommended either by the ECSC working party or by the NHLI.

The properties of spirometers used for survey work are critical and are discussed in detail. The recommendations were made with the best types of traditional instruments in mind. Indeed, some water filled spirometers are the only instruments which match exactly the specifications stipulated. The provision of electronic alternatives of equal accuracy and stability is still providing a challenge to designers and manufacturers.

It is proposed that spirometers should have an inertia of 0.01 kPa l⁻¹ s² and a dynamic resistance of 0.1 kPa l⁻¹ s. The spirometer most commonly used in clinics and occupational health centres in the United Kingdom, the Vitalograph, does not always meet the criterion for inertia with important consequences which are discussed below.

The need to record temperature in water filled spirometers, at a distance from soda lime, is mentioned. There is no discussion of the tendency of some bellows spirometers to alter their performance with changing ambient temperature. Here the application of conventional correction factors can introduce substantial errors.¹³ A satisfactory means of converting FEV₁ to BTPS units using these spirometers has not yet been determined.

The ability to record 8 litres is specified as a requirement for spirometers. Clearly this is not necessary for routine practice but is essential if subjects with exceptionally large lungs are not to be excluded from collection of normal data, with a consequent distortion of the statistics. Similarly, in studies including patients with mild or moderate airflow obstruction, the ability to record a forced expiration for 14 seconds is necessary. This has to be emphasised in view of the increasing number of different automated systems becoming available, some of which define the end of a forced expiration too early.

The end of the test is defined as the time when the spirometer starts to move at a rate of less than 0.025

litres/s. A forced expiration of this length is usually tolerable although some individuals will find it unpleasant or will show deteriorating performance if the procedure is repeated. Repeated forced expirations of this duration are not suitable for prolonged procedures such as challenge tests or dose-response curves. There has been no general agreement about how to define the end of a forced expiration when forced expired time is prolonged beyond the normal 5 seconds. The working party's proposal is helpful and could generally be adopted.

To define the start of forced expiration after full inspiration, backward extrapolation of the fastest portion of the curve is preferred to the practice of discarding the convex first portion of the curve, usually 0.1 litres. The latter convention, however, has been convenient in ordinary clinical spirometry as the kymographs of bellows spirometers are conveniently activated at the start of expiration by the movement of the volume recording arm. The difference between the two methods is unimportant.

CALCULATION OF INDICES FROM FORCED EXPIRATORY TESTS

Preferred indices for the maximum expiratory flow volume (MEFV) and FVC (volume/time) curves are FEV₁, FVC, MEF_{50%FVC}, FEF_{25-75%FVC}, and FEV₁/FVC. Subscripts of the MEFV indices refer to the fraction or percentage of the forced vital capacity remaining to be expired. The additional use of FEV₁/IVC is recommended. MEF_{60%TLC} and MEF_{25%FVC} require better reference values.

The envelope method of analysis is recommended, whereby the curves are superimposed graphically at TLC and the highest value of each index is taken. The alternative method given is easier to use if the method is to be computerised. MEF at 75%, 50% and 25% FVC are taken to be the highest value obtained from each of the three largest FVC tests, FVC varying by not more than 5%. For results before and after helium, FVC should not differ by more than 0.2 litres.

The poor reproducibility of flow-volume indices and their limited value in the absence of good reference data are recognised. It is assumed that the measurements will be made during dynamic spirometry. For physiological studies such as the construction of isovolume pressure-flow curves it is necessary to measure changes of thoracic volume in a body plethysmograph, when different curves are obtained in the presence of very minor degrees of airflow obstruction.

CHANGE

The document contains no discussion about how bronchodilator studies or challenge tests should be

performed. It is suggested that change between two measurements x_1 and x_2 should be reported as a fraction of the mean of the two measurements (and not of the initial value):

$$x_2 - x_1 / 0.5 (x_2 + x_1)$$

Because of the problem of regression to the mean¹⁴ this is theoretically the correct way of expressing the difference between two serial results. It would be impossible to analyse dose-response curves, challenge studies and diurnal variation in this way and no evidence is presented to show that this suggestion has any practical value.

Because changes or improvements in FVC are generally caused by changes in RV, rather than TLC, it is suggested that MEFV curves after being superimposed at TLC should be recorded as the change in MEFV at 50% of initial FVC. This recommendation seems sensible (though it would have been more consistent to have chosen MEF at 50% of mean FVC.)

HOW MANY BLOWS?

The NHLI document discussed this in detail and, while the ECSC working party had access to more data, they reached the same conclusion. FEV₁ and FVC should each be recorded as the best of three technically satisfactory attempts¹⁵ not exceeding the next best by more than 0.3 litres. Two practice attempts are allowed but after eight attempts the procedure should always be abandoned. Falling values should be noted as indicating the onset of bronchoconstriction or fatigue.

This approach assumes that all the errors are in the direction of under-reading. While this may be true for instruments of high inertia such as the Vitalograph, readings on critically balanced instruments such as the Garw bellows spirometer tested by Ullah *et al*¹⁶ are normally distributed, which suggests that, after the learning effect has been eliminated, over-reading should be as common as under-reading. These considerations lead to the choice of the mean of the best three of five readings. The working party considered this and thought the improvement in reproducibility not to be cost effective.

What constitutes a technically satisfactory test is not defined. In practice, it means that the subject appears to have taken a full inspiration, held it until the start of a maximal forced expiration and then to have exhaled fully and forcefully without leaks or closure of the glottis. It follows that the operator must see the trace as soon as it has been generated. Instruments that generate numbers alone do not meet this strict criterion, although experienced operators and patients regularly obtain good results

from such devices as peak flow meters in wards and in domiciliary studies as well as in the laboratory.

Other tests

The chapter on transfer factor (diffusing capacity) contains little new material and in general follows the advice of Cotes.¹⁷ Correction for haemoglobin according to a formula is recommended. KCO is defined as TL (STPD)/VA (BTPS).

The sections dealing with lung elasticity and airways resistance spell out useful criteria for the characteristics of the instruments employed but offer little advice about how best to present the results. Presumably it was recognised that these techniques are not routine and that each application requires a slightly different approach.

Reference values

The entire literature on reference values has been reviewed critically, an immense undertaking. Summary equations, based on the most acceptable data, were obtained by calculating sample results for hypothetical individuals of each combination of height and age, without weighting for the number of subjects in each study.

The working party deplored the recent trend whereby smokers are excluded from normal populations, because it has led to the publication of reference values which are regarded as unrealistically large for populations which include a high proportion of smokers and therefore reduced the usefulness of reference values as indicators of disease. The possibility of producing useful separate reference values for different groups of smokers and for non-smokers is not discussed explicitly. It is evident from the detailed tables given that this would not at present be feasible, though it would be desirable. In environmental research, the differences from normality being sought are generally too small to permit the use of published normal data and controls are necessary for each study. In clinical physiology, reference values are used to measure the likelihood that an individual patient departs from normality, or to normalise data for body size. For each application, a knowledge of the effects of smoking habit on the test under consideration is helpful. To include smokers in normal reference data widens the scatter but correspondingly a deviant value is more likely to mean that disease is present.

Quite large variations were found between different populations within the Community, which are only partly explained by the inclusion of varying numbers of smokers. Ethnic differences and social

Reference values: differences between Cotes¹⁷ and the European Coal and Steel Community document²

| Male | 1-55 | | 1-95 | | Female | | 1-45 | | 1-8 | |
|--|-------|-------|-------|-------|--------|-------|-------|-------|-----|----|
| | 25 | 70 | 25 | 70 | 25 | 70 | 25 | 70 | 25 | 70 |
| Height (m) | | | | | | | | | | |
| Age (y) | 25 | 70 | 25 | 70 | 25 | 70 | 25 | 70 | 25 | 70 |
| FEV ₁ (l)* | +0.02 | +0.12 | +0.3 | +0.4 | -0.1 | +0.05 | +0.04 | +2.6 | | |
| FVC (l)* | +0.01 | -0.16 | +0.22 | +0.06 | -0.24 | -0.16 | -0.36 | -0.20 | | |
| MEF ₅₀ (ls ⁻¹)† | -0.9 | -0.9 | -0.4 | -0.4 | -0.3 | -0.3 | -0.3 | -0.3 | | |
| TLCO†† | +0.2 | +0.2 | +0.3 | +0.3 | -0.02 | +0.01 | +0.2 | +0.2 | | |
| KCO** | +0.4 | +0.4 | +0.4 | +0.4 | +0.25 | +0.38 | +0.25 | +0.38 | | |

MEF₅₀—maximum expiratory flow when 50% of FVC has been expired; TLCO—single breath CO transfer factor; KCO—transfer coefficient (TLCO/VA BTSP).

*litres BTSP; †litres BTSP/per second; ††SI units: mmol min⁻¹ kPa⁻¹; ** SI units: mmol min⁻¹ kPa⁻¹ l⁻¹ (conversion factor from SI units to traditional units × 3).

background, which may have an influence on lung growth, are difficult to control. The importance of standardising for time of day and year, to overcome circadian and seasonal variation, is just becoming appreciated and has an important bearing on serial data in individuals as well as on population studies.

A plea is made for information about healthy subjects in more homogeneous populations within the Community and for better information about subjects between the ages of 15 and 30. Deficiencies in the reporting of normal data abound in the literature and the report calls for proper methods of collection of data in future using rigorous selection criteria, employing standard questionnaires and adhering to the guidelines laid down for the procedures as well as providing accurate information about the exact procedures used.

The reference data presented with the report have been prepared carefully and are likely to remain the best available for many years. The consequences of changing from the formulae in the fourth edition of Cotes¹⁷ to the ECSC figures are summarised in the table. For example, there is a 10% change in KCO in women and in VC in short women but in general the changes are relatively slight.

Conclusion

Standardisation is important in epidemiology and in research. It is sometimes suggested that it is less important in clinical practice. It is true that the accuracy required of a blood gas analyser in the management of respiratory failure is less than that needed to measure alveolar-arterial differences. In practice lung function tests are used by clinicians (1) to detect departures from normality, (2) to measure the severity of disease and (3) to identify variability or change in a patient's condition. All of these purposes call for good equipment and standard techniques, particularly the third. Most tests have an intrinsic variability of about 5% and yet in many

conditions it is necessary to identify changes of 10%. This is the average improvement of FEV₁ seen after low doses of bronchodilators in asthma and of VC when stage II sarcoidosis goes into remission. This degree of reproducibility calls for a high standard of manufacture and maintenance of equipment if comparable results are to be obtained on machines in different clinics, wards and laboratories. The robustness of one bellows wedge spirometer has made it almost the universal choice in clinics and health centres. This instrument has imposed its own standards, because laboratories have been reluctant to make occasional measurements on superior equipment with loss of comparability and perhaps only a marginal increase in accuracy.

To make use of published information in clinical practice it is necessary to know how it was obtained. Editors should ask for accurate descriptions of methods used, including formulae. It is equally important for authors to insist that their methods sections are not cut excessively when their work appears in general journals or in journals which purport to cater for clinicians.

The ECSC recommendations are in general sensible and deserve wide discussion by those responsible for the provision of lung function services. The alternative approaches are listed rather than discussed; this will make debate a little more difficult than it should be. Nonetheless, the summary provides useful guidelines which, if widely accepted, should be particularly valuable for reference when buying equipment and setting up new projects.

G LASZLO
Bristol Royal Infirmary
Bristol

The assistance of AH Kendrick in the preparation of this article is gratefully acknowledged.

References

- 1 Cara M, Hentz P. *Aide-memoire of spiographic practice for examining ventilatory function*. 2nd ed. Luxembourg: Commission of the European Communities, Industrial Health and Medicine Series 11, 1971.
- 2 Quanjer PH, ed. Standardized lung function testing. *Bull Europ Physiopath Resp* 1983;19 (suppl. 5):1-95.
- 3 Ferris BG (principal investigator). Epidemiology standardization project. *Am Rev Respir Dis* 1978;118 (no 6, part 2):55-88.
- 4 Goddard PR, Nicholson EM, Laszlo G, Watt I. Computed tomography in pulmonary emphysema. *Clin Rad* 1982;33:379-87.
- 5 Brown R, Ingram RH jr, McFadden ER. Problems in the plethysmographic assessment of changes in total lung capacity in asthma. *Am Rev Respir Dis* 1978;44:469-73.
- 6 Brown R, Scharf S, Ingram RH. Non-homogeneous alveolar pressure swings in the presence of airway closure. *J Appl Physiol* 1980;49:398-402.
- 7 Tierney DF, Nadel JA. Concurrent measurements of functional residual capacity by three methods. *J Appl Physiol* 1962;57:18-28.
- 8 Loyd HM, String TI, DuBois AB. Radiographic and plethysmographic determination of total lung capacity. *Radiology* 1966;86:7-14.
- 9 Nicholson EM. The measurement of pulmonary emphysema in life. MD thesis, 1980; University of Bristol.
- 10 Todd JR. Lung volumes and pattern of breathing in patients with chronic airways obstruction. MSc thesis, 1981; University of Bristol.
- 11 Shore SA, Huk O, Mannix S, Martin JG. Effect of panting frequency on the plethysmographic determination of thoracic gas volume in chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1983;128:54-9.
- 12 Pierce RJ, Brown DJ, Holmes M, Cumming G, Denison DM. Estimation of lung volumes from chest radiographs using shape information. *Thorax* 1979;34:726-34.
- 13 Perks WH, Sopwith T, Brown D, Jones CH, Green M. Effects of temperature on Vitalograph readings. *Thorax* 1983;38:592-4.
- 14 Rossiter CF. Discussion in *Scand J Resp Dis* 1976;113:587-600.
- 15 Fletcher CM, Peto R, Tinker C, Speizer FE. *The natural history of chronic bronchitis and emphysema*. Oxford: Oxford University Press, 1976.
- 16 Ullah MI, Cuddihy V, Saunders KB, Addis GJ. How many blows really make an FEV₁, FVC or PEF_R? *Thorax* 1983;38:113-8.
- 17 Cotes JE. *Lung function: assessment and application in medicine*. 4th edition. Oxford: Blackwell Scientific Publications, 1979.