METHODOLOGY Series editors: T C Aw, A Cockcroft, R McNamee

Lung function testing: methods and reference values for forced expiratory volume (FEV₁) and transfer factor (TL)

J E Cotes, D J Chinn, J W Reed

Respiration and
Exercise Laboratory,
Department of
Physiological Sciences,
Medical School,
Newcastle upon Tyne,
NE2 4HH
J E Cotes
D J Chinn
J W Reed

Correspondence to:
Dr J E Cotes, Respiration
and Exercise Laboratory,
Department of Physiological
Sciences, Medical School,
The University, Newcastle
upon Tyne, NE2 4HH.

Accepted 20 November 1996

Background

WHICH TESTS?

Almost 50 years ago Gilson *et al* embarked on the first large scale investigation into impaired lung function and breathlessness on exertion resulting from an occupational exposure.[1] The study described the lungs in terms of their size, ventilatory power, effectiveness in distributing gas, and ability to transfer gas into the blood. The tests were chosen on account of being simple and acceptable for the subjects, requiring only a few minutes to perform, having adequate accuracy and reproducibility, and yielding relevant information which could not be obtained more conveniently in other ways. These concepts were widely accepted.[2] They pointed to a core of three general tests which should be available to occupational physicians and epidemiologists confronted by a possible respiratory hazard. In order of usefulness, the tests are ventilatory capacity (forced expiratory volume in one second and peak expiratory flow (FEV₁ and PEF)), transfer factor (TL), and the physiological responses to submaximal exercise. The tests will indicate the extent of any respiratory impairment associated with a previous or chronic chest illness or with cough, wheeze, or breathlessness on exertion, such as might be a consequence of breathing polluted air. In the presence of variable airflow obstruction, additional tests may be required.[3]

(Occup Environ Med 1997;54:457–465)

WHEN MIGHT THE TESTS BE USED?

Measurement of FEV₁ usually forms part of the pre-employment and exit medical examinations; this and the other lung function tests may also be used for surveillance and clinical management, monitoring the effects of an incident, assessing residual function and disability, and as an aid to rehabilitation.

CONTENTS

The paper describes the measurement of FEV₁ and TL, comments on some other ventilatory indices, and makes suggestions for reference values. It incorporates recommendations for standardisation of the European Respiratory Society,[4] and the British and American Thoracic Societies,[5][6] as interpreted and amplified in a recent account.[7]

Ventilatory capacity

INDICES

Indices from volume-time curve

The relation of volume expired to time of expiration (volume-time curve, also called a spirogram, fig 1) is used to obtain the FEV₁.[8] This is the maximal volume of gas that can be expired in the first second of a forced exhalation after a full inspiration. The FEV₁ is used primarily as an indicator of airflow obstruction, and in the assessment of breathlessness on exertion. The spirogram also yields forced vital capacity (FVC); this is the maximal volume that can be exhaled during the same forced expiration. The FVC is reduced in most diseases of the lung, but has limited discriminatory power. It is used to standardise FEV₁ for lung volume, in the relation FEV₁ × 100/FVC, hence FEV%. However for this application, vital capacity measured during either inspiration or a relaxed expiration (respectively IVC and EVC) provides a more appropriate denominator; this is because the requisite breathing manoeuvres, unlike forced expiration, do not give rise to dynamic compression of airways.[4][9]

As well as FEV₁ and FVC, the relation expired volume versus time can provide indices of mid-expiratory flow (fig 1), and information on the times taken for molecules of gas to leave the lungs. However, when FVC is reduced the rates of flow (hereafter referred to as flows) are difficult to interpret, and indices of transit time seem to be of limited usefulness.

Indices from flow-volume curve

Instantaneous flows, including PEF and maximal expiratory flows (MEFs) when specified volumes of gas have been expired from the lungs, are obtained from the relation of MEF to lung volume (flow-volume or v̇-V curve, fig 2). The v̇-V curve can be used to check that the forced expiratory manoeuvre has been performed correctly.[4] The MEFs when 50% and 25% of the FVC remain to be expired (MEF₅₀%FVC and MEF₂₅%FVC) can be markers of early abnormality[10][11]; however, the flows are not independent of lung volume, so their reproducibility is poor, and interpretation of deviations from the expected levels can be uncertain.[12] The PEF from a v̇-V curve
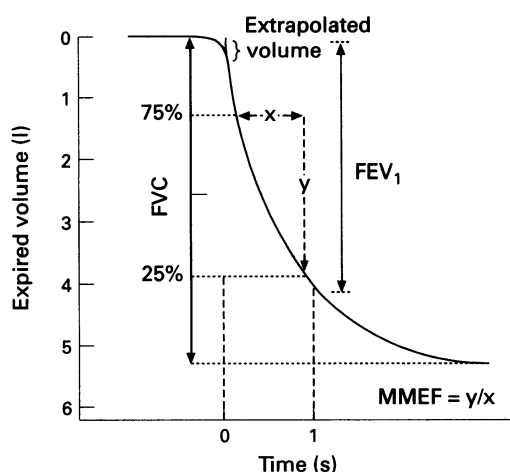
Figure 1   Idealised volume-time curve (spirogram) showing forced vital capacity (FVC), forced expiratory volume in one second (FEV$_1$), derivation of maximal mid expiratory flow (MMEF), and extrapolation procedure for estimating zero time used in the derivation of FEV$_1$. The extrapolated volume should not exceed 100 ml or 5% of FVC, whichever is greater. Adapted from Quanjer et al[4] and Cotes.[7]

provides a guide to the quality of that forced expiration (see later). However, except in some cases of increased bronchial reactivity or where there may be obstruction to upper airways, PEF contributes little additional information over and above that provided by FEV$_1$ and FVC. The PEF measured during the course of dynamic spirometry is seldom used in occupational and environmental medicine.

Peak expiratory flow from a peak flow meter
In this form the PEF is used to look for diurnal variation in airway calibre, variation over a working shift, and cyclic variation related to pattern of employment including weekends and holidays. The instrument is usually a variable orifice or turbine anemometer, which is small enough to fit in a pocket and is operated by the subject without supervision. Training should have been provided. The measurement is performed in a manner similar to that
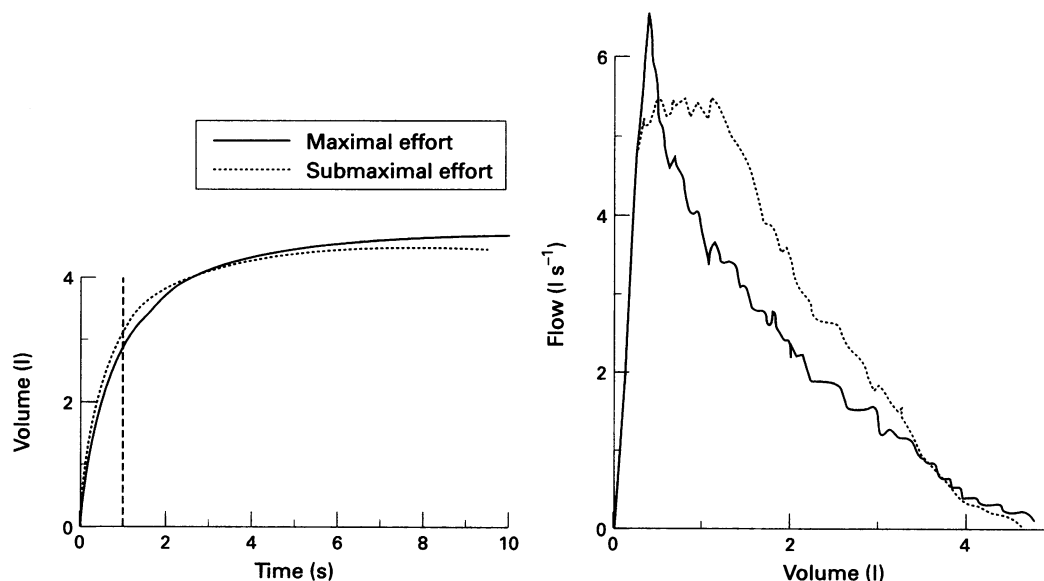
described later for dynamic spirometry, except that the forced expiration can be discontinued after mid-expiration. The calibration and use of peak flow meters are described elsewhere.[5 7 13]

EQUIPMENT
The primary measurements are either volume with respect to time, measured with a turbine anemometer, mass flow meter, or spirometer plus timer, or flow measured with a pneumotachograph, with which volume is obtained by integration of the flow signal with respect to time. The result is in the form of a numerical print out, plus a display of the associated v-V or volume-time curves.

Choice of equipment
Criteria to be taken into account include the likely numbers of sites and of measurers, the intensity and duration of use, the ease of cleaning (see later), and the extent of any epidemiological application. For longitudinal studies, and when several measurers will use the equipment, this should be robust and covered by a long term service contract. Multipurpose equipment which meets the technical specification[4-7] is often suitable, but the requirements and suitability for the other applications (which can include measurement of transfer factor, see later) should be established. The specification includes both physical characteristics (table 1) and conventions for selecting and processing the measurements. These are not always adhered to, or are interpreted by manufacturers in ways which may be inappropriate, hence expert advice should be taken.

Calibration
This is done at the start of each measurement session, with a gas syringe of capacity 1 or 3 l, which should be available at every measuring site. Volume recorders are checked directly. Recorders which derive volume from integration of a flow signal are calibrated by emptying

Figure 2   Volume-time and v-V curves for a healthy subject, showing that submaximal effort can spuriously increase the FEV$_1$ (in this instance from 2.81 l to 3.28 l). The lack of effort is apparent in the v-V curve; this shows a flattened peak which is displaced to the right. Thus the expiration was defective and the result unacceptable. From Cotes[7] with permission.

Table 1 *Minimal standards for equipment used to measure gas volumes and flow (adapted from Cotes[7])*

Volume range	0–8 l
Volume accuracy	± 3% or ± 50 ml†
Reading accuracy	25 ml
Driving pressure	<0.03 kPa
Paper speed	0.02 m s^{-1}
Recording time	14 s accurate to ± 1%
Flow range	$1–15 \text{ l s}^{-1}$
Flow accuracy	$± 4\% \text{ or } ± 0.07 \text{ l s}^{-1}$
Dynamic resistance	$< 0.05 \text{ kPa l}^{-1} \text{ s}$
Inertia	$<0.001 \text{ kPa l}^{-1} \text{ s}^2$
Maximal mouth pressure	< 0.6 kPa
Dynamic response	3–20 Hz*

*The response should be flat (within 5%) at 3 Hz for FEV_1, MMEF, and $MEF_{25\% \ FVC}$, 5 Hz for $MEF_{50\% \ FVC}$, and 20 Hz for PEF.

†Whichever is greater.

the syringe through the instrument at different time intervals, all of which should provide the same estimate of volume. Flow-volume recorders are calibrated with a flow simulator,[14] which should itself be calibrated.[15] Timers are checked against a stop watch.

Errors from equipment

Spirometers can be unreliable as a consequence of leaks, increased resistance from wear or secretions, electric batteries needing recharging, faults in the sensor, timing, or electronics, and neglect of allowance for internal temperature.[7 16 17] Flow meters can also pass unregistered gas; they may be subject to accumulation of secretions. These faults can be detected by regular calibration and maintenance of equipment.

Errors can also arise from inappropriate treatment of intermediate results (already discussed); the errors can affect new equipment, or be introduced if parts of the equipment or associated computer software are replaced or upgraded The errors are usually systematic, but their effect may be non-linear, in which case to obtain correct results the data should be reworked after revision of the computer program.

Cleanliness

All infections of the respiratory tract, open sores on the oral mucosa, and bleeding gums are potential sources of cross infection, to which immunocompromised people are particularly susceptible. However, infection from respiratory apparatus is extremely rare even among high risk patients; hence, although hygiene measures should be appropriate, there is no need for special measures such as are recommended for patients with known transmissable diseases.[5 6 18]

A mouthpiece and any valve box should be renewed or washed in antiseptic detergent and dried between patients. A bag or tubing from the mouthpiece, through which the subject both inhales and exhales, should be washed and drained after use; however, when tubing is used only for expiration, washing can be confined to the end of the measurement session, unless the session is a large one. Closed circuit breathing apparatus and measuring devices which connect directly on to the mouthpiece can require special treatment.[7]

PERSONNEL

The operator will usually be a technician or occupational nurse; he or she should have received training in the measurement of stature and lung function and in the recording of respiratory symptoms with a questionnaire. Competence in these tasks should be tested by comparison of results with those by an experienced observer.

PROCEDURE

Preparation of equipment

At the start of the session the technician performs a calibration and checks the supply of relevant stores, including clean mouthpieces and other equipment, mouth tissues, and results sheets.

Preparation of subject

With a view to subsequent interpretation of results, the name, address, sex, date of birth, and predominant ethnic group of the subject are noted. Enquiry is made about any recently inhaled medication and the time of the last dose is recorded. Also, if the information is not available elsewhere, respiratory symptoms, smoking history, any exposure to respirable dusts or vapours, and relevant previous illnesses are recorded by a trained interviewer, with a standardised questionnaire.[19] Stature is measured after removal of boots. Accuracy to within 0.5 cm is achieved by a stadiometer, with correct posture and head traction, as recommended for the international biological programme.[7 20] Neglect of these precautions can lead to material error, up to 5 cm, which in turn affects the reference values (discussed later). Body mass is measured with conventional weighing scales; the scales and the stadiometer should have been calibrated.

After these preliminaries, the subject sits in front of the spirometer, preferably having slackened any tight clothing and adopting an upright but relaxed posture; dentures which are loose should be removed. For dynamic spirometry, a nose clip is not recommended except for people with broad noses and others who may produce variable results.[7] The procedure is now demonstrated by the technician; also, it is helpful for the subject to watch a previous one.

Measurement of FEV_1 and FVC

The subject is instructed to make a really complete inspiration, which should be through the mouth if a nose clip is in place. The subject then inserts the mouthpiece into the mouth, closes the lips around it but with the jaws apart, and blows out through the mouth as hard and as completely as possible; *strong verbal encouragement should be given throughout this sequence.* After its completion, the v-V curve, and preferably also the volume-time curve should be inspected: the v-V curve should have smooth ascending and descending limbs, separated by a clear peak; the volume-time curve should progress smoothly to a steady level (figs 1, 2).

The subject should repeat the measurement, with intervals between of at least 30 seconds, and achieve three acceptable results (out of a

total of not more that eight); to this end, all instructions should have been followed correctly, the procedure should have been free from the errors listed later, and a smooth curve obtained. Definitive values for FEV$_1$ and FVC are reported at body temperature and pressure, saturated with water vapour (BTPS conditions). Each value listed is the highest from any of the three curves, subject to the provision that the highest FVC does not exceed the next highest by more than 5% or 0.1 l.[4] When the discrepancy is larger, this should be noted; however, the result should not be rejected, as its variability may itself be useful information.[21] The practice of taking all results from a best curve is not recommended.

Measurement errors

A spuriously low result can occur from:

(*a*) failure to record from full inspiration due to:

- inspiration not being complete
- tight clothing restricting the chest or abdomen
- partial expiration before the mouthpiece is inserted

(*b*) inadequate dynamic response because of:

- leakage of gas round the mouthpiece (or through the nose)
- expiration with submaximal force
- partial obstruction or crushing of the mouthpiece by lips, tongue or denture

(*c*) expiration ended prematurely (hence FVC low)

(*d*) bronchoconstriction induced before or during the measurement.

In subjects with increased bronchial reactivity, bronchoconstriction is sometimes induced by performance of the forced expiratory manoeuvre; in this circumstance the FEV$_1$ diminishes progressively with each subsequent expiration. Alternatively, bronchoconstriction can be induced voluntarily by coughing, forced breathing, omitting to take a prescribed bronchodilator medication, or inhaling tobacco smoke or other bronchoconstrictor substance before the measurement. If one of these features is suspected, the examining physician should auscultate the chest for the presence of sibilant rhonchi. Their presence in a subject who normally uses a bronchodilator aerosol is an indication for the subject to use his or her inhaler in the presence of the physician before performing the test. Alternatively the inhalation can be overseen by the respiratory technician, who should have received appropriate training for this role.

An artificially high result is uncommon, but can occur if the airways have an increased susceptibility to dynamic compression (fig 2). In this circumstance the FEV$_1$ is inversely related to the expiratory effort; maximal effort then yields a relatively low but reproducible result.

Non-cooperation

Faults seldom occur to the same extent on consecutive breaths, hence where they are introduced deliberately, the results are often unduly variable. Such non-cooperation can usually be detected by attention to the points which have been outlined, but this is not invariable. For example, among 196 applicants for industrial injuries benefit who attended our laboratory, and for whom the information was available, there were 13 in whom the measured FEV$_1$ differed from that obtained in another laboratory by more than 4 SDs (mean difference 0.87 l, range 0.4–1.66 l). In two of the 13, the lower of the two results was due to failure to use an inhaler, and in a further two, there may have been another medical explanation. Three of the subjects were recorded as not cooperating in the laboratory where the lower result was obtained. However in six subjects there was no obvious technical, physiological, or medical explanation for the lower result. In this circumstance, the possibility of malingering could not be excluded.

Reporting the result

The report should include the definitive results for FEV$_1$ and FVC, derived from the conventions already described, together with a comment on the technical quality of the measurements, and some of the factual information already listed; the items should include those used to calculate the reference values (discussed later). Reference values are usually given as means (SDs), which are the residual SDs about the regression equations (RSDs). In the case of most indices of lung function, the RSDs are nearly constant throughout adult life, hence an abnormal FEV$_1$ or FVC is one which is reduced by at least 1.64 RSD compared with the reference value. Alternatively, the probability of any given result differing from the reference value can be expressed in SD units (difference of observed from expected result÷RSD), when again a deviation >1.64 RSD is considered abnormal.[4] The percentage deviation—for example, 80% predicted—has a variable connotation, depending on the age of the subject, and should not be used.

The starting point for interpreting a result is to decide whether or not it is abnormal, as defined above. The next step is usually to look for patterns of abnormality; for example in the case of FEV$_1$, a value which is reduced relative to the FVC (hence FEV% is reduced), is evidence of an obstructive type of ventilatory defect, such as might accompany chronic bronchitis, emphysema, or asthma. If FEV$_1$ and FVC are both reduced, so that FEV% is within normal limits, the ventilatory defect is likely to be non-obstructive in type, and any of several mechanisms might be responsible.[7] The identification of patterns can sometimes be helped by expressing the deviations of the observed from the reference values on a continuous scale of SD units, or the corresponding percentiles. For example, 1.64 RSD corresponds to the 5th or 95th percentile. However, this approach can convey a spurious precision, as both the reference values may not be optimal for the population under review, and the subjects may have been identified for scrutiny on account of having possibly reduced values, and not at random, as required by statistical theory.

When a previous result is available, a second one is interpreted in relation to the first,[22] and
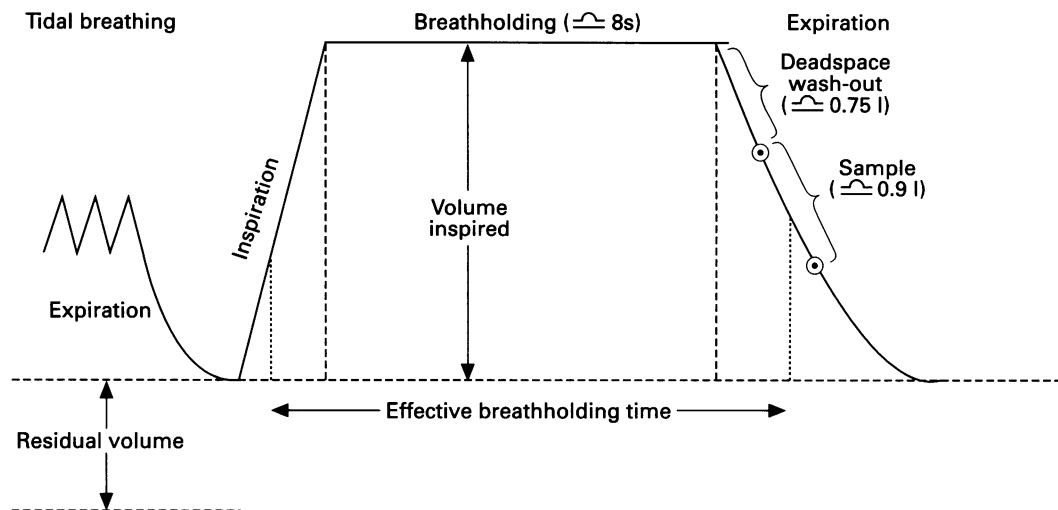
Figure 3 Spirogram showing the procedure for determining the TL (diffusing capacity) for the lung by the single breath method. The subject breathes out to residual volume, inhales a vital capacity breath of the test gas, holds the breath for eight seconds then exhales slowly; after the exhalation of 0.75 l a sample of alveolar gas is collected for analysis. From Cotes[7] with permission.

a third should help to indicate a trend; but interpretation should take into account the measurement error, which is large relative to the annual change. Use can also be made of longitudinal reference values (see later).

Transfer factor (TL)

The TL (also called diffusing capacity), describes the facility with which gas passes across the alveolar capillary membrane. It is usually measured for carbon monoxide (CO), when it is expressed for the whole lung as the quantity of CO passing per unit time per unit of pressure difference across the membrane. The quantity is affected by the volume and degree of inflation of the lung, the distribution of pulmonary blood flow, and the age, sex, anthropometric features, and haemoglobin concentration of the subject. It is reduced by smoking and by diseases which damage or thicken the alveolar capillary membrane including extrinsic allergic alveolitis from any cause, beryllium disease, pulmonary fibrosis—for example, asbestosis—and emphysema. Thus the measurement has a role in occupational respiratory medicine and epidemiology.

OUTLINE OF METHOD

For the present applications, the measurement is made by the single breath holding method of Forster et al and Cotes et al.[7] [23] [24] The subject inhales a vital capacity breath of a test gas containing carbon monoxide (usually 0.3%) and helium or other insoluble gas (concentration 2%–14%), holds it in the lungs for 10 seconds, then exhales, during which time a sample of alveolar gas is taken for analysis. The volume inspired and the time of breath holding are displayed and recorded (fig 3). The information can be sufficient for the calculation of TL, and this is done automatically when modern equipment is used.

However, results are not always compatible between different equipment manufacturers or different laboratories; this can have several causes, including the following:

Inspired oxygen concentration

In European countries the test gases are diluted in air, hence when the inert gas is to be 10% to 14% helium, the inspired oxygen concentration is about 17%, and the alveolar concentration is in the physiological range.[7] [24] In the United States the inspired concentration is set at 21%,[18] hence the alveolar oxygen concentration is higher, and the resulting TL is lower than when a more physiological concentration is used. The extent of the difference is of the order of 3%, but reference values by the two methods are similar; however, the costs differ.[25]

Breathholding time

In the calculation, the assumption is made that the decline in alveolar CO concentration with time is exponential, hence the decline is linear when plotted on a semilog scale. The slope of the line is calculated from two points; these define an initial alveolar concentration at a time near to the start of breathholding and a final concentration at the time when the alveolar gas is sampled. Some imprecision is inevitable, as gas exchange starts before all the test gas has entered the lungs, and continues during the final expiration. These sources of error can be minimised by extrapolation to define the start and end of the test inspiration,[26] together with the convention for breathholding time of Jones and Meade.[7] [27] Its use is now recommended internationally.[18] [24] The convention of Ogilvie et al can also be used for subjects capable of meeting defined criteria for the durations of inspiration and sample collection (table 2)[28]; this is likely to include most people in employment, but not those with airflow limitation, who may record unduly low results.[29] Other conventions for timing may lead to high results (described later), and are not recommended.[29] [30]

Alveolar volume during breathholding

This is usually calculated from the dilution in the lung of the inert gas present in the test gas mixture (VA,eff). Allowance should then be made for any change in gas concentration due

to absorption of CO_2 before analysis of the alveolar gas sample. However, the single breath procedure underestimates the alveolar volume in patients with chronic airflow limitation.[29] In these circumstances, alveolar volume can be obtained by adding to the volume of the test breath, the residual volume measured by an independent method—for example, the closed circuit helium dilution method (VA). Among people who are at work, the number for whom this is relevant is likely to be small.[31]

EQUIPMENT

Choice

The test is performed with compact, fully automated, digitised equipment, which can yield a concurrent estimate of ventilatory capacity (already described). The choice of equipment is likely to be influenced by cost, including that of the service contract; however, the technical specification is of overriding importance. Thus, the gas analysis should include both inspired and alveolar gas; error from possible contamination of either sample should be avoided or allowed for.[7] Three point calibration of the analysers should be straightforward. The spirogram describing the breathing manoeuvre should be available for inspection both during the measurement and subsequently, and the convention for breathholding time should be similar to that of Jones and Meade.[27] Allowance for the effect of absorbed CO_2 on the concentration of helium used to calculate alveolar volume should be realistic.[4 5]

Calibration

The standards given in tables 1 and 2 should be adhered to. For volume and flow the methods have already been summarised; for gas concentrations either the inspirate is diluted serially in the apparatus,[7] or three calibration gas mixtures chosen to span the relevant concentration ranges[32] are used instead. The interval between calibrations will depend on the stability of the respective methods of measurement and on the frequency of use: it should not be longer than every month. The helium analyser and the accuracy of the CO_2 analyser can be calibrated every three months. However, the linearity of the CO_2 analyser (+1%) should be checked before each measurement session; this is easily and quickly done by serial dilution of the inspirate gas mixture.[4 7]

Errors from equipment and cleanliness

Errors from equipment (other than calibration errors), and cleanliness have already been discussed.

PERSONNEL

Personnel have already been discussed.

PROCEDURE

Preparation of equipment

Preparation of equipment has already been discussed.

Preparation of subject

If not already available, the information already referred to should be obtained. To ensure

Table 2 Quality control when measuring transfer factor (adapted from Cotes')

Inspired CO concentration	0.25%–0.3%; if below this range the CO backtension should be allowed for.
Inspired O_2 concentration	17%–18% (or 21%, see text).
Inspired volume	> 90% of largest measured vital capacity.
Alveolar volume	Method should be stated. VA,eff should always be calculated.
Washout volume	0.75–1.0 l (0.5 l if VC ≤ 2l).
Sample volume	0.5–1.0 l collected over < 3 s into bag of volume ≤ 1l.
Times of inspiration and sample collection	Normally < 2.5 s; up to 4 s if there is airflow limitation.
Effective breathholding time	9–11 s (subject relaxes against shutter). Should include allowance for time of sample collection (Jones Meade method[27]).
Gas analysis	Gas should be conditioned by absorbing CO_2 then water vapour. Both inspirate and "alveolar" sample should be analysed, with a linearity and accuracy of ± 1%*.[24]
Definitive result	Mean of two which meet above criteria, and do not differ by more than 10%.[24]

*American Thoracic Society accept ± 3%.[18]
Condition of the subject: rested, postabsorptive, seated for at least 10 minutes, preferably adequately bronchodilated, and not having smoked or taken alcohol that day. At least four minutes should be allowed after any previous test involving breathing He or O_2.

stable conditions, the subject should be seated, rested for at least 10 minutes, and be in a postabsorptive state. The time interval since any previous test entailing inhalation of helium should have been at least four minutes, and the subject should not have smoked for at least one hour. If the subject has been exposed to CO or has smoked heavily on the day of the test, or if more than six measurements of TL are envisaged, the alveolar CO concentration should be monitored before the start of the tests and an allowance made for back tension.[7]

Measurement

The measurement procedure has been summarised already (fig 3). It is described in detail elsewhere.[5 7 18 24] Table 2 lists some points that merit particular attention, and figure 4 shows some common faults. Normally, the recommended procedure should be followed, and the tracing should be free from faults. To obtain a representative result, the alveolar volume should be within 5% or at most 10% of total lung capacity.

REPORTING THE RESULT

The TL is usually reported in standard international units (SI units) as mmol min⁻¹ kPa⁻¹; however, in the United States, diffusing capacity is reported in traditional units as ml (standard temperature and pressure dry) min⁻¹ mm Hg⁻¹. The conversion factor is 2.98 traditional units to one SI unit. A definitive result is the mean of two technically satisfactory measurements which agree to within 10%. It is for the whole lung and is usually interpreted in terms of a reference value based on age and stature, without reference to haemoglobin concentration. However, if the result is abnormal, the presence of factors which might influence the quantity of haemoglobin available for gas
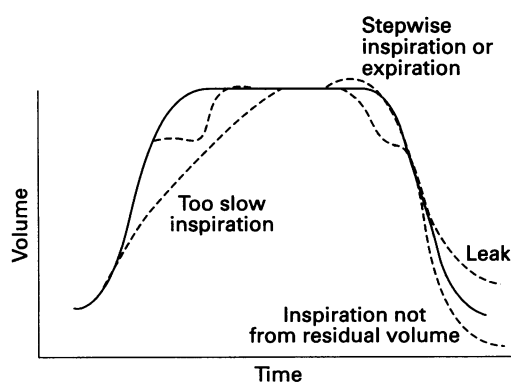
Figure 4 Faults in the respiratory manoeuvre which should lead to a measurement of TL being rejected on technical grounds. From Cotes[7] with permission.

exchange, including anaemia and heavy smoking, should be taken into account. In the case of haemoglobin concentration this is done by standardising to the mean concentrations of 14.6 g dl⁻¹ for men and 13.6 g dl⁻¹ for women.[7 18]

If alveolar volume is atypical, an additional allowance should be made. Traditionally this was done by reporting the TL per unit of alveolar volume, designated KCO or TL/VA. However, the use overcorrects for the dependency of TL on VA.[33] Because of this, if the lung volume is reduced by the subject failing to make a maximal inspiration, or by diffuse pleural fibrosis, the TL may seem to be reduced, whereas the KCO from the same breath can be normal or increased. The difficulty can be overcome by reporting the reference value for TL at the alveolar volume of the subject, with an extended reference equation, which includes alveolar volume among the reference variables.[34] Table 3 shows this for a patient with calcified pleural plaques in whom TL was apparently greatly reduced, whereas KCO was completely normal. The inconsistency was due to the conventional reference values making no allowance for the alveolar volume being reduced; when this was done by use of the new relation, the TL was found to be at the lower end of the normal range. Thus the gas exchange was neither normal nor grossly impaired; instead the patient had physiological changes consistent with early asbestosis.

The related difficulty that reference equations for TL in current use show systematic differences which are related to volume, is considered later.

Reference values

Group results, for example those obtained with a view to biological monitoring of the workplace, are likely to be interpreted statistically.[35] Individual results are usually interpreted with respect to the reference value. This is either a previous result for the person, or that predicted

for an average subject of similar age, stature, sex, and ethnic group. Also, the reference value should be appropriate to the question which is being asked. Thus if the suspected pollutant is environmental, the reference values might be for lifetime non-smokers reared and living in a clean environment. However, if the pollutant is occupational, the appropriate values might be for asymptomatic people with a similar background. Smoking habit will usually be taken into account, but this is not invariable.

Sets of reference values have been compiled.[7 36] Also, measurements from several sources have been combined to form European Coal and Steel Community reference values.[37] These have the advantage of communality to all European countries, but the disadvantage that the processes of compilation and averaging have led to systematic bias in the ratioTL/VA,[38] and possibly also FEV₁/FVC. Also, reference values become obsolete unless they are updated regularly to take into account refinements in measurement techniques and other factors.

REFINEMENTS IN THE MEASUREMENT
This has affected FEV₁, in which the currently recommended method entails extrapolation of the volume-time curve back to a hypothetical starting volume at zero time (fig 1). The extrapolation is intended to compensate for inertia in the thorax leading to maximal volume change not being achieved instantly. Compared with a direct measurement, extrapolation has the effect of increasing the measured values by on average 2.5%. A further increase of similar magnitude resulted from the convention for a definitive result being changed from the mean, to the maximum of three technically satisfactory expirations.

REFINEMENTS IN THE SELECTION OF SUBJECTS
Formerly subjects were accepted if they were free from material respiratory symptoms. Those who only wheezed occasionally, and asymptomatic smokers were included, and the effect of smoking was treated as a categorical variable (one which is present or absent). Now reference values are usually for lifetime non-smokers, preferably those without exposure to any air pollution. The average levels of reference values have increased in consequence. Exclusion of people who wheezed occasionally, has led to a further increase in the reference value for FEV₁ of about 2.1%.[35]

REFINEMENTS IN THE MODEL USED FOR ANALYSIS; RELEVANCE FOR LONGITUDINAL MEASUREMENTS
Early reference values were cross sectional, and the decline with age was treated as linear from the age of 20, by which time men as well as women had usually stopped growing in stature. A later model assumed a plateau of lung function at the age of 20-25, with a subsequent linear decline.[39] Now the decline with age is recognised to be curvilinear.[40 41] Compared with the linear cross sectional model, the curvilinear one predicts lower values for FEV₁ in older and younger subjects, and slightly higher ones for middle aged subjects.

Table 3 Lung function in a patient with calcified pleural plaques, secondary to exposure to asbestos (men aged 60, stature 1.87 m, total lung capacity 5.9 l (expected TLC 7.9 l))[34]

Observed result	Reference value	Difference (RSD units)
TL = 8.6 SI units @ VA = 5.4 l	12.2 SI units @ TLC = 7.9 l	−2.8*
	10.3 SI units @ VA = 5.4 l	−1.5
TL/VA = 1.60 SI/l	1.62 SI/l	−0.1

*Reduced compared with normal (within 1.64 RSD of reference value).

Neglect of the dependence of annual decline on age has the effect that, for a reference population studied on two occasions a few years apart, the cross sectional reference values at a given age are systematically lower on the second occasion. Clearly such values are of limited usefulness for interpreting longitudinal observations. Instead, either a curvilinear model, or longitudinal reference values should be used. Longitudinal reference values should be based on accurate observations spanning a minimal interval of five to seven years.[42] The primary observation can then be the decline per year, when the reference variables are likely to include the level of the index in question, the mean age, and the changes in stature and body mass index over the period of observation. Neglect of the body mass index in particular can lead to an occupational effect being overlooked, if the underlying cause also results in a change in body mass.[43]

IDENTIFICATION OF THE COHORT EFFECT

In many communities, the young people are on average taller than their parents, and the increased growth affects the lungs. Because of this, for a specified age and stature, men and women from recent cohorts on average have better lung function than their predecessors from earlier cohorts.[44] [45] Thus the cohort effect influences the partial regression coefficients on age in both cross sectional and longitudinal studies. The underlying factors are environmental.

ETHNIC VARIATION

On average, people coming originally from south India,[7] [46] west Africa,[47] and possibly Australasia,[48] have small lungs relative to age and stature, compared with other ethnic groups. For black people, but not for Asians, the difference is due in part to their having long legs relative to stature, hence relatively small thoracic dimensions. The ethnic factor affects mainly the vital capacity, its subdivisions, and the FEV_1. Residual volume and TL are less affected, hence $FEV_1\%$ is independent of ethnic group, unlike KCO (TL/VA) which is relatively high in groups with small lungs.[7]

The ethnic factor is inherited, so, at present, people of mixed race have intermediate values.[49] However, with increasing genetic mixing, the size of the ethnic factor will diminish, and the general level of reference values will change in consequence.

DOUBT ABOUT REFERENCE VALUES FOR TL

Results for TL are particularly susceptible to the conventions which are adopted for the measurement.[50] Formerly not all studies used the same conventions, so differences between series were inevitable. Some of the differences were identifiable—for example, that by the epidemiology standardisation project, which led to systematic underestimation of the time of breathholding and hence overestimation of TL. A more recent finding has been that mean results from different series of reference values, vary systematically with mean alveolar volume.[34] Hence, in the past, measurements of

alveolar volume were not always comparable. The error should not extend to technically advanced apparatus, built to current specifications. However, until such time as there is independent certification, this expectation can best be corroborated by biological calibration.[7]

Overview

The development of agreed international protocols for measuring FEV, FVC, lung volumes, and TL, together with a standard technique for measuring stature, have led to a situation where measurements can be comparable between laboratories. However, procedures for calibration have not kept pace with these developments, and not all laboratories use the protocols. Among those that do, the implementation is often imperfect. Thus comparability of measurements cannot be assumed.

Progress has been made towards standardisation of cross sectional reference values, with two sets in widespread use.[7] [37] Recent changes have led to an increase in the average levels of the lung function indices,[36] but the new values are not necessarily better than the old ones. Optimal reference values are those which most nearly resemble the test situation for age, cohort, ethnic group, selection factors, and technique for measurement, including instruments and protocols. International reference values are a worthwhile objective, but to be useful, the underlying measurements should be of uniformly high standard, and the interpretation should probably be based on more reference variables than is the case at present. When doubt exists, the preference should be for reference values of which the user already has experience.

1 Gilson JC, Hugh-Jones P, Oldham PD, Meade F. Lung function in coal workers' pneumoconiosis. Gas-distribution and transfer. *Special Report of the Medical Research Council (London)* 1955;290:114–24,150–201.
2 International Labour Organisation. *Respiratory function tests in pneumoconioses*. Geneva: ILO, 1966. (Occupational Safety and Health Series 6.)
3 Sterk PJ, Fabbri LM, Quanjer PhH, Cockcroft DW, O'Bryne PM, Anderson SD, et al. Airway responsiveness: standardized challenge testing with pharmacological, physiological, and sensitizing stimuli in adults. *Eur Respir J* 1993;6(suppl 16):55–86.
4 Quanjer PhH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Standardized lung function testing: lung volumes and forced ventilatory flows. 1993 update. *Eur Respir J* 1993;6(suppl 16):4–40.
5 British Thoracic Society. Guidelines for the measurement of respiratory function;recommendations of the British Thoracic Society and the Association of Respiratory Technicians and Physiologists. *Respir Med* 1994;88:165–94.
6 American Thoracic Society. Standardisation of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995;152:1107–36.
7 Cotes JE. *Lung function; assessment and application in medicine*, 5th ed. Oxford: Blackwell Sci Publ, 1993.
8 Gandevia B, Hughes-Jones P. Terminology for measurements of ventilatory capacity. *Thorax* 1957;12:290–3.
9 Tiffeneau R, Pinelli A. Regulation bronchique de la ventilation pulmonaire. *Journal Francais de Médecine et Chirugie Thoraciques* 1948;2:221–44.
10 Becklake MR, Permutt S. Evaluation of tests of lung function for screening for early detection of chronic obstructive lung disease. In: Macklem PT, Permutt S, eds. *The lung in the transition between health and disease*. New York: Marcel Dekker, 1979:345–87.
11 Davey IS, Cotes JE, Reed JW. Relationship of ventilatory capacity to hyperbaric exposure in divers. *J Appl Physiol* 1984;56:1655–8.
12 Stanescu DC, Rodenstein DO, Hoeven C, Robert A. Sensitive tests are poor predictors of the decline in forced

expiratory volume in one second in middle-aged smokers. *Am Rev Respir Dis* 1987;135:585–90.

13 Pedersen OF, Rasmussen TR, Kjaergaard SK, Miller MR, Quanjer PhH. Frequency response of variable orifice type peak flow meters: requirements and testing. *Eur Respir J* 1995;8:849–55.

14 Pedersen OF, Naeraa N, Lyager S, Hilberg C, Larsen L. A device for evaluation of flow recording equipment. *Bulletin Européen de Physiopathologie Respiratoire* 1983;19:515–20.

15 Burge PS. Calibrating the calibrators [editorial]. *Thorax* 1996;51:969.

16 Pincock AC, Miller MR. The effect of temperature on recording spirograms. *Am Rev Respir Dis* 1983;128:894–8.

17 Johnson LR, Enright PL, Voelker HT, Tashkin DP. Volume spirometers need automated internal temperature sensors. *Am J Respir Crit Care Med* 1994;150:1575–80.

18 American Thoracic Society. Single breath carbon monoxide diffusing capacity (transfer factor). Recommendations for a standard technique — 1995 update. *Am J Respir Crit Care Med.* 1995;152:2185–98.

19 Medical Research Council. *Questionnaire of respiratory symptoms.* London: MRC, 1976.

20 Weiner JS, Lourie JA, eds. *Practical human biology.* London: Academic Press, 1981.

21 Eisen E, Dockery DW, Speizer FE, Fay ME, Ferris BJ Jr. The association between health status and the performance of excessively variable spirometry tests in a population-based study in six US cities. *Am Rev Respir Dis* 1987;136:1371–6.

22 Chinn DJ, Cotes JE, Fechner M, Elliott C. Pre-employment lung function at age 16 as a guide to lung function in adult life. *Br J Ind Med* 1993;50:422–7.

23 Forster RE, Fowler WS, Bates DV, Van Lingen B. The absorption of carbon monoxide by the lungs during breathholding. *J Clin Invest* 1954;33:1135–45.

24 Cotes JE, Chinn DJ, Quanjer PhH, Roca J, Yernault J-C. Standardisation of the measurement of transfer factor (diffusing capacity). Report of working party. Standardisation of lung function tests, 1993 update. *Eur Respir J* 1993;6(suppl 16):41–53.

25 Kendrick AH, Laszlo G. Standardisation of gas mixtures for estimating carbon monoxide transfer factor. *Thorax* 1993; 48:863–5.

26 Chinn DJ, Harkawat R, Cotes JE. Standardization of single-breath transfer factor (TLCO); derivation of breathholding time. *Eur Respir J* 1992;5:492–6.

27 Jones RS, Meade F. A theoretical and experimental analysis of anomalies in the estimation of pulmonary diffusing capacity by the single breath method. *Quarterly Journal of Experimental Physiology* 1961;46:131–43.

28 Ogilvie CM, Forster RE, Blakemore WS, Morton JW. A standardized breathholding technique for the clinical measurement of the diffusing capacity of the lung for carbon monoxide. *J Clin Invest* 1957;36:1–17.

29 Chinn DJ, Askew J, Rowley L, Cotes JE. Measurement technique influences the response of transfer factor (TLCO) to salbutamol in patients with airflow limitation. *Eur Respir J* 1988;1:15–21.

30 Crapo RO. Effect of the method of measuring breathholding time on DLCO and DL/VA in normal subjects [abstract]. *Am Rev Respir Dis* 1980;121(suppl 331).

31 Van Ganse W, Comhaire F, Van der Straeten M. Alveolar volume and transfer factor determined by single breath dilution of a test gas at various apnoea times. *Scand J Resp Dis* 1970;51:82–92.

32 Chinn DJ, Naruse Y, Cotes JE. Accuracy of gas analysis in lung function laboratories. *Thorax* 1986;41:133–7.

33 Stam H, Kreuzer FJA, Versprille A. Effect of lung volume and positional changes on pulmonary diffusing capacity and its components. *J Appl Physiol* 1991;71:1477–88.

34 Chinn DJ, Cotes JE, Flowers R, Marks A-M, Reed JW. Transfer factor (diffusing capacity) standardised for alveolar volume;validation, reference values, and applications of a new linear model to replace KCO (TL/VA). *Eur Respir J* 1996;9:1269–77.

35 Chinn DJ, Cotes JE, El-Gamal FM, Wollaston JF. Respiratory health of young shipyard welders and other tradesmen studied cross-sectionally and longitudinally. *Occup Environ Health* 1995;55:33–42.

36 American Thoracic Society. Lung function testing;selection of reference values and interpretative strategies. *Am Rev Respir Dis* 1991;144:1202–18.

37 Quanjer PhH, ed. Standardized lung function testing. *Bulletin Européen de Physiopathologie Respiratoire* 1983;19:suppl 5, 1–95.

38 Love RG, Seaton A, Quanjer PhH. About the ECCS summary equations [letters]. *Eur Respir J* 1990;3:489–90.

39 Lawther PJ, Brooks AGF, Waller RE. Respiratory function measurements in a cohort of medical students: a ten-year follow-up. *Thorax* 1978;33:773–8.

40 Burrows B, Lebowitz MD, Camilli AE, Knudson RJ. Longitudinal changes in forced expiratory volume in one second in adults. Methodologic considerations and findings in healthy non-smokers. *Am Rev Respir Dis* 1986;133:974–80.

41 Dockery DW, Ware JH, Ferris BG Jr, Glicksberg DS, Fay ME, Spiro A, Speizer FE. Distribution of forced expiratory volume in one second and forced vital capacity in healthy, white, adult never-smokers in six US cities. *Am Rev Respir Dis* 1985;131:511–20.

42 Berry G. Longitudinal observations, their usefulness and limitations with special reference to the forced expiratory volume. *Bulletin de Physiopathologie Respiratoire* 1974;10: 643–55.

43 Chinn DJ, Cotes JE, Reed JW. Changes in body mass can affect interpretation of longitudinal measurements of ventilatory capacity (FEV$_1$ and FVC). *Thorax* 1996;51: 699–704.

44 Glindmeyer WH, Diem JE, Jones RN, *et al.* Noncomparability of longitudinal and cross-sectionally determined annual change in spirometry. *Am Rev Respir Dis* 1982;125:544–8.

45 Van Pelt W, Borsboom GJJM, Rijcken B, Schouten JP, Van Zomeren BC, Quanjer PhH. Discrepancies between longitudinal and cross-sectional change in ventilatory function in 12 years of follow-up. *Am J Respir Crit Care Med* 1994;149:1218–26.

46 Cotes JE, Dabbs JM, Hall AM, Lakhera SC, Saunders MJ, Malhotra MS. Lung function of healthy young men in India: contributory roles of genetic and environmental factors. *Proc R Soc Lond B Biol Sci* 1975;191:413–25.

47 White NW, Hanley JH, Lalloo UG, Becklake MR. Review and analysis of variations between spirometric values reported in 29 studies of healthy African adults. *Am J Respir Crit Care Med* 1994;150:348–55.

48 De Hamel FA, Welford B. Lung function in Maoris and Samoans working in New Zealand. *NZ Med J* 1983;96: 560–2.

49 Miller GJ, Saunders MJ, Gilson RJC, Ashcroft MT. Lung function of healthy boys and girls in Jamaica in relation to ethnic composition, test exercise performance, and habitual physical activity. *Thorax* 1977;32:486–96.

50 Morris AH, Crapo RO. Standardisation of computation of single-breath transfer factor. *Bulletin Européen de Physiopathologie Respiratoire* 1985;21:183–9.