Thorax 1988;**43**:265–267

Editorial

Predicted values: how should we use them?

Most tests of function used in medical practice have normal values that can be conveniently expressed as ranges pertinent for the whole population—for example, serum sodium, creatinine. The pioneering studies of lung function in the mid 19th century, however, found that spirometric indices differed not only between the sexes but also with age and height.[1] Many studies have confirmed this finding and most have found the data suitable for multiple linear regression, thus allowing predicted values to be derived for a given individual.[2] In using a particular regression equation in this way it is necessarily assumed that the reference population and the recording techniques used with that population are relevant to one's own subjects and equipment.

Although the practice of obtaining predicted values has been widely accepted for lung function data, it is not clear how an individual's value should be compared with his or her predicted value. In Britain expressing the result as a percentage of the subject's predicted value (% predicted) has been widely adopted (% predicted = recorded value × 100/predicted value) and this value is now calculated on many automated spirometers and body plethysmographs. The American Thoracic Society has recommended that confidence limits should be used to determine whether an individual's result is to be declared abnormal and that % predicted should be used to express the degree of deviation from the predicted value.[3] The use of confidence limits is logical as an individual's result should be judged in the context of the range of values found among the reference population.

Despite its widespread use the % predicted method has no basis in any scientific discipline and over 20 years ago concern was expressed in this journal about its use.[4] In expressing values as % predicted two assumptions are made: firstly, that a given % predicted for one index is comparable in terms of deviation from the predicted value to the same % predicted value for another index and, secondly, that for each index a given % predicted means the same for subjects of different age, sex, and height. When challenged neither of these assumptions turns out to be true.

If we use the European Coal and Steel Community (ECSC) prediction equations[2] for a man of 50 years, 1·7 m in height, the lower 95% confidence limits for FEV_1 × 100/FVC, FVC, residual volume (RV), and forced mid flow 25%–75% VC (FMF) are 82%, 71%, 61%, and 46% of the predicted values. Hence comparable limits in the reference population for different lung function indices are at different percentages of the predicted values. The assertion that results of lung function tests below a specified percentage of the predicted value are "abnormal" would be true only if an agreed confidence limit matched this value, which is not the case. For several indices this assertion could mean that many subjects within the "normal" reference population should be declared abnormal,[4] which would invalidate the original regression analysis. Clearly the first assumption is unfounded.

In considering the second assumption we must first look at the spread of the data within the reference population. Figure 1 shows a plot of theoretical reference data for FEV_1 based on the ECSC regression equation. The spread of those absolute data is uniform irrespective of age and such data are termed homoscedastic. In this instance the lower 95% confidence limit expressed as % predicted varies with age, being 74% of predicted at the age of 31 (subject A), and 63% of predicted at 70 (subject B). For the index FMF, which has a steeper slope with age than FEV_1, the comparable figures are 56% and 31% of the predicted values. If the absolute data were heteroscedastic in such a way that the spread of data diminished with

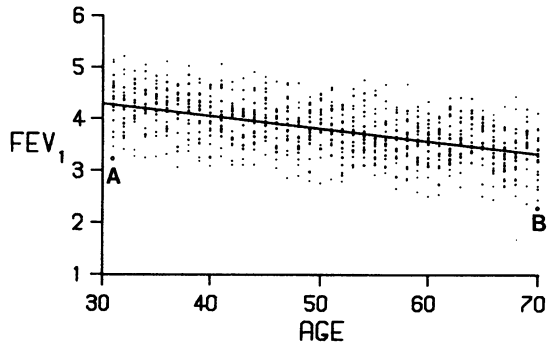

Fig 1 *Theoretical reference data for FEV_1 for men of height 1·7 m. Subjects A and B are on the lower 95% confidence limit, which is 74% and 63% respectively of their predicted values.*

age, the confidence limit might fortuitously be at a fixed % predicted irrespective of age. This has not been found for adult lung function data. Hence the second assumption concerning the % predicted method is unfounded.

If from a theoretical consideration the % predicted method is flawed, does this have a practical implication? For an index such as FEV_1, which is larger in taller subjects and also declines with age, the use of the % predicted method should lead to more elderly and shorter individuals being called abnormal than the original regression data would support. A study of workers exposed to asbestos compared confidence limits and the % predicted method for identifying subjects with lung function impairment.[5] The authors found a discordant group who were abnormal when assessed by the % predicted method but not when assessed in relation to the confidence limits. A distinguishing feature of this group was that they were older and, being older, they had experienced a longer asbestos exposure. Other workers[6] have found that a confidence limit was better than % predicted at identifying young subjects with airflow limitation, thus indicating an age bias in the % predicted method. An earlier study had shown an age and height bias in the use of % predicted to define abnormality in a reference population of over 500 subjects,[7] the authors concluding that a method using absolute residuals (residual = recorded − predicted) was superior to % predicted.

These findings mean that the use of the % predicted method may leave hidden bias in the data, which negates the whole purpose of referring to a predicted value. If this method is used to make comparisons between groups of data or to look for correlations then this hidden bias may act to defeat or enhance any true association. If in an individual instance this method is found to hold no significant bias then this is not a sufficient argument to support the continued use of % predicted, just as a demonstration of walking blindfold across a busy road without causing an incident is not a substantive argument in defence of continuing this practice.

Many find the % predicted method simple and have been comfortable with its use. Is there a suitable alternative that avoids any bias? Use of absolute residuals alone has been tested[7] but these retain the units of the index and thus numerically vary from index to index. An alternative is to use standardised residuals (SR), which are derived by dividing the absolute residual by the residual standard deviation (RSD) taken from the regression equation used.[8] This requires the use of an additional piece of information supplied with each regression equation (RSD_a in ref 2, p 49), which is a measure of the spread of the reference data. Thus a standardised residual takes into account

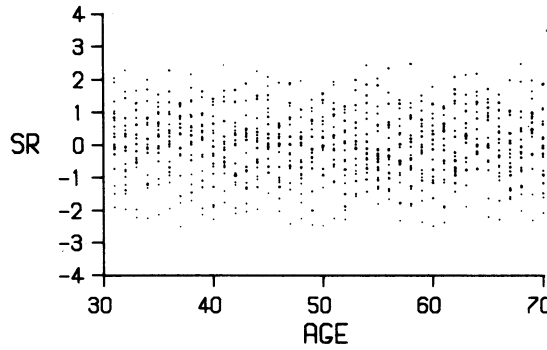

Fig 2 *Plot of the data from figure 1 expressed as standardised residuals (SR). SR = (recorded FEV_1 − predicted)/RSD, where RSD is the residual standard deviation from the regression equation used for the prediction.*

the range of values found in the reference population which % predicted does not. Figure 2 shows a plot of SR against age for the subjects who appear in figure 1. Such a plot forms part of the routine diagnostic checks on a linear regression analysis and here indicates that the spread and magnitude of SR are independent of age. The mathematical derivation of SR is simple, and SRs have the same scale and units for all indices. When data are normally distributed an SR of −1·65 is at the lower 90% confidence limit and an SR of −1·96 at the lower 95% confidence limit.

We favour the use of SR for every instance when a numerical expression of an individual's lung function is required that is free from sex, age, and height bias. This may be for determining whether an important degree of deviation from the predicted value is present or when an unbiased measure of lung function is to be related to another measurement, such as smoking exposure or bronchial reactivity. Assessment of acute changes in function in an individual, due to treatment or challenge, are often expressed as absolute or percentage change from baseline without reference to a predicted value because the subject is acting as his or her own control. The use of percentage change from baseline to define reversibility of airflow limitation has been shown to be biased, leading to more subjects with the severest airflow limitation being termed reversible than is justified.[8,9] Use of absolute change in FEV_1 to define reversibility may be a more valid approach as this better matches the short term variability of the index.[9] This issue, however, is separate from that of how to relate an individual's result to the predicted value.

With extensive standardisation of techniques for lung function testing[2,10] it is evident that attention must now be directed towards standardising the way the

results of these tests are presented.

MARTIN R MILLER
University of Birmingham Department of Medicine
Good Hope District General Hospital
Sutton Coldfield
West Midlands
ARCHIE C PINCOCK
University of Birmingham Department of Medicine
Queen Elizabeth Hospital
Birmingham

References

1 Hutchinson J. On the capacity of the lungs, and on the respiratory function, with a view to establishing a precise and easy method of detecting disease by the spirometer. *Medico-chirurgical Transactions* 1846; **29**:137–252.
2 European Coal and Steel Community recommendations. *Bull Eur Physiopathol Respir* 1983;**19**(suppl 5):1–93.
3 American Thoracic Society statement. Evaluation of impairment/disability secondary to respiratory disease. *Am Rev Respir Dis* 1982;**126**:945–51.
4 Sobol BJ, Weinheimer B. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966;**21**:445–9.
5 Oliver LC, Eisen EA, Sprince NL. A comparison of two definitions of abnormality on pulmonary outcome in epidemiologic studies. *Am Rev Respir Dis* 1986; **133**:825–9.
6 Eliasson O, Degraff AC. The use of criteria for reversibility and obstruction to define patient groups for bronchodilator trials. *Am Rev Respir Dis* 1985; **132**:858–64.
7 Harber P, Tockman M. Defining disease in epidemiologic studies of pulmonary function: percent of predicted or difference from predicted? *Bull Eur Physiopathol Respir* 1982;**18**:819–28.
8 Miller MR, Pincock AC, Grove DM. Patterns of spirogram abnormality in individual smokers. *Am Rev Respir Dis* 1985;**132**:1034–40.
9 Tweeddale PM, Alexander F, McHardy GJR. Short term variability in FEV_1 and bronchodilator responsiveness in patients with obstructive ventilatory defects. *Thorax* 1987;**42**:487–90.
10 American Thoracic Society. Standardisation of spirometry—1987 update. *Am Rev Respir Dis* 1987; **136**:1285–98.