Thorax 1984;**39**:928–932

Short term variability in FEV_1: relation to pretest activity, level of FEV_1, and smoking habits

PM TWEEDDALE, S MERCHANT, M LESLIE, F ALEXANDER, GJR McHARDY

From the Lothian Area Respiratory Function Service; and Departments of Respiratory Medicine and Medical Computing and Statistics, University of Edinburgh, Edinburgh

ABSTRACT The natural variability in forced expiratory volume in one second (FEV_1) over 20 minutes was determined in 54 fit hospital employees and 13 patients with restrictive lung disorders. Initial FEV_1 ranged from 1·1 to 6·3 l BTPS. Variability when expressed as absolute change was similar at all levels of FEV_1, so that, when expressed as percentage change, variability decreased with increasing FEV_1. Smoking habits did not appear to affect variability but activity before the test did. On the basis of these results an absolute change in FEV_1 of 190 ml would be necessary for 95% confidence that the change in FEV_1 occurred other than by chance in any one individual. This suggests that the absolute change in FEV_1 might be a more reliable criterion than percentage change when distinguishing between natural variability and a response to inhalation of bronchodilators.

The measurement of forced expiratory volume in one second (FEV_1) is of fundamental importance in respiratory medicine, not least on account of its reproducibility and worldwide usage. There are three prerequisites for valid use of the measurement: firstly, the apparatus used must be of an acceptable standard and detailed recommendations exist[1]; secondly, if predicted normal values are to be used they must be reliable and relevant, a topic considered extensively by the European Community for Coal and Steel[2]; thirdly the natural variability that is to be expected in estimates of FEV_1 must be taken into account, particularly when alterations in bronchial calibre are being assessed. This variability encompasses reproducibility and accuracy of equipment, technical expertise, and patient performance. Only when the natural variability is known can criteria be set for assessing if a significant change has taken place, for example after administration of a bronchodilator.

Although there is extensive information on variability within one set of FEV_1 measurements[3–6] and between measurements made hours or days apart,[7–9] there appears to be no documentation of the spontaneous variability in FEV_1 when measurements are made 20 minutes apart (the time allowed for response in many laboratory tests of bronchodilator action). Furthermore, controversy still exists over which criteria should be used to assess response to bronchodilators.[10 11] We therefore determined the natural variability of FEV_1 when measurements were performed 20 minutes apart (without any treatment being given) and calculated from this the change in FEV_1 that would be necessary to distinguish between spontaneous variability and a response to any treatment given.

Address for reprint requests: PM Tweeddale, Respiratory Function Laboratory, City Hospital, Edinburgh EH10 5SB.

Accepted 20 July 1984

Methods

Fifty four hospital employees (26 men and 28 women), whose ages ranged between 18 and 62 and heights from 1·45 to 1·79 m, took part. None of the subjects to their knowledge had asthma, hay fever, chronic bronchitis, or cardiac disorders, although 15 were smokers and 13 exsmokers, as defined by the Medical Research Council's Respiratory Symptom Questionnaire.[12] Most subjects were unfamiliar with spirometry. In addition, 13 patients with known restrictive ventilatory defects but no recognised obstructive component to their respiratory problems took part in the study. Their ages ranged between 20 and 70 and their heights from 1·50 to 1·75 m; two were smokers and four were exsmokers. All subjects were asked to indicate the level of their activity immediately before the test, whether seated or active. Mild activity was defined as walking within the hospital building, moderate activity as outdoor walking, stair climbing, or manual work.

FEV_1 was measured with a bellows spirometer (Vitalograph, Buckingham), which had been calibrated for volume with a precision syringe (Hans Rudolph, Kansas City). The subjects were seated and three forced expiratory manoeuvres were performed; if all three were not technically acceptable—that is, they varied by more than ±5% or by ±0·1 l, whichever was greater[1]—either one or two further manoeuvres were performed until three technically acceptable results had been obtained. The highest FEV_1 from the three acceptable results was recorded and will be referred to as FEV_1I. FEV_1 was again recorded in similar fashion after 20 minutes, this measurement being referred to as FEV_1II.

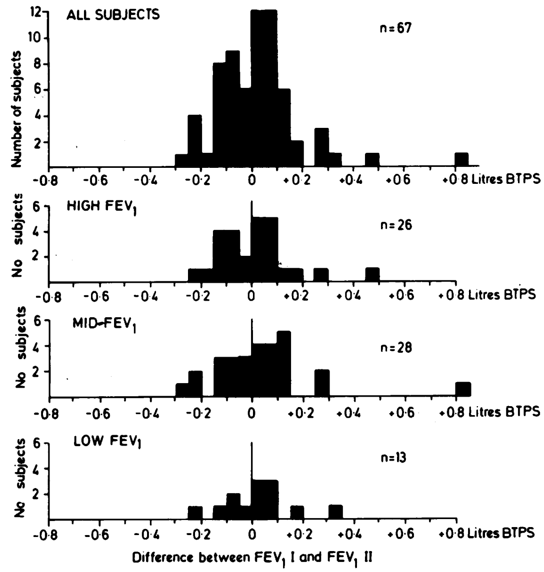
The results for the normal subjects were divided into two subgroups to permit comparison of variability at different levels of FEV_1: those with initial FEV_1I greater than 4 l (n = 26) known as the high FEV_1 group; and those with FEV_1I of 4 l or less (n = 28) known as the mid-FEV_1 group. These subgroups were compared with the patient group (n = 13) known as the low FEV_1 group.

The study received the approval of ethical committees in the north and south Lothian districts.

STATISTICAL ANALYSIS

The variability or change in FEV_1 was expressed both as absolute change (AC), where AC = FEV_1II − FEV_1I, and as percentage change (PC), where PC = (FEV_1II − FEV_1I) × 100/baseline. The baseline was defined as (FEV_1I + FEV_1II)/2.

The definition of baseline in these terms conforms to accepted statistical practice, based on the finding that the change in a variable is correlated with the initial value but is independent of the mean of the initial and final values.[13] The baseline follows the recommendation of the European Community for Coal and Steel.[2]



Histograms showing absolute changes in FEV_1 *for all subjects and three subgroups. The high* FEV_1 *group comprised normal subjects with* FEV_1*I greater than 4 l, the mid-*FEV_1 *group those with* FEV_1*I, 4 l or less, and the low* FEV_1 *group patients with restrictive disorders.* FEV_1*I—initial forced expiratory volume in one second;* FEV_1*II—forced expiratory volume in one second after 20 minutes.*

Table 1 *Mean changes in forced expiratory volume in one second* (FEV_1) *over 20 minutes*

	All subjects (n = 67)	*Patients*	*Normal subjects*	
		(low FEV_1*) (n = 13)*	*(mid* FEV_1*) (n = 28)*	*(high* FEV_1*) (n = 26)*
Range of FEV_1I recorded (l BTPS)	1·1–6·3	1·1–2·25	2·4–4·0	4·1–6·3
Mean (SD) AC (ml BTPS)	+22·5 (170·1)	+26·2 (138·1)	+36·8 (204·7)	+5·4 (146·6)
Mean (SD) PC (%)	+0·7 (5·6)	+1·6 (8·0)	+0·9 (6·2)	+0·1 (2·9)

FEV_1I—initial forced expiratory volume in one second; AC—absolute change; PC—percentage change.

Table 2 *Comparison of mean changes in* FEV_1 AC—absolute change; PC—percentage change.

	Activity before test		
	Seated (n = 31)	*Mild (n = 27)*	*Moderate (n = 9)*
Mean (SD) AC (ml BTPS)	−34·5 (102·5)	+29·6 (119·1)	+197·8 (325·5)
Mean (SD) PC (%)	−0·9 (4·5)	+1·1 (4·1)	+5·3 (9·4)

AC—absolute change; PC—percentage change.

Table 3 *Probabilities of given changes in FEV_1 occurring by chance in any one individual who has been seated or mildly active before test*

Probability of change occurring by chance (p%)	*Change in FEV_1 (ml)*	
	All subjects	*Normal subjects only*
2·5	+229	+234
5	+191	+195
10	+148	+151
20	+97	+99

One tailed analysis, where change = $t_{p \times 2}$ × sample standard deviation.
Sample standard deviation and degrees of freedom: 114·2 and 57 for all subjects; 116·2 and 46 for normal subjects alone.

Results

Normal plots of absolute change and percentage change for all 67 subjects were produced by computer, following accepted statistical practice. It was found that the variables were roughly normally distributed and that an assumption of constant variance was justifiable. Table 1 shows the mean (SD) for absolute change and percentage change for all subjects and for the FEV_1 subgroups. The figure shows the distributions of absolute change, which were roughly normal apart from one outlying result in the mid-FEV_1 group. As the mean absolute change and mean percentage change for all subjects combined and for each subgroup were not significantly different from zero we concluded that there had been no learning effect.

There was no evidence of any significant effect of smoking habit on absolute change or percentage change within the 67 subjects. When we examined activity before the test, however, we found that this did affect the values of absolute change and percentage change. Table 2 shows mean values of absolute change and percentage change for all subjects when divided into three groups on the basis of their degree of activity before the tests. One way analysis of variance showed a highly significant difference for both absolute change ($p < 0{\cdot}001$) and percentage change ($p < 0{\cdot}01$) between the groups, the change in FEV_1 increasing with increased activity ($p < 0.002$ for linear trend). This was primarily because large changes and standard deviations were associated with the nine subjects who had the highest degrees of prior activity and these included the outlying result from the mid-FEV_1 group. Even when this result was excluded from the moderate activity group, however, giving a mean (SD) absolute change for the group of +116·3 (229·6) and a mean (SD) percentage change of +2·98 (6·85) there was still a significant increase in absolute change and percentage change with increasing activity (positive linear trend $p < 0{\cdot}003$). The significant effect of activity was also found when the patient group was removed from the analysis ($p < 0{\cdot}003$ for difference between the groups and $p < 0{\cdot}002$ for linear trend, n = 54).

The seated and mild activity groups had similar estimates of variance and the mean change in FEV_1 was not significantly different ($p > 0{\cdot}05$). When the mild and moderate activity groups were compared, despite an apparently large difference in the mean changes in FEV_1, significance was not achieved ($p > 0{\cdot}05$, Wilcoxon rank sum test). This was due in part to the small number of subjects in the moderate activity group and there being a highly significant difference between estimates of variance in the two groups ($p < 0{\cdot}001$). The moderate activity group of nine subjects was therefore considered to be too heterogeneous for inclusion in the final analysis of change in FEV_1 in relation to the level of FEV_1. The remaining 58 subjects comprised 11 patients in the low FEV_1 group, 24 normal subjects in the mid-FEV_1 subgroup, and 23 in the high FEV_1 subgroup. As no learning effect had been shown, the magnitude of the changes in FEV_1, irrespective of sign, was used in the one way analysis of variance of absolute change and percentage change in all three FEV_1 subgroups. (Although the sample was reasonably large, the distribution was skewed and the results were consequently confirmed by the Kruskal-Wallis test.[14]) There was no significant difference in the mean absolute change between the three FEV_1 subgroups ($p = 0{\cdot}3$), and there was no consistent trend towards larger absolute change as the baseline increased. When the mean percentage change was compared for the three groups, however, there was a highly significant difference between subgroups ($p < 0{\cdot}001$) with the percentage change decreasing as the baseline FEV_1 increased (negative linear trend, $p = 0{\cdot}001$).

The probability of a particular absolute change in FEV_1 occurring by chance over a 20 minute interval in a subject who had been seated or mildly active was calculated from the *t* distribution (table 3).[15] A one tailed test was chosen because of its greater relevance in clinical situations. It may be seen that for practical purposes taking the changes in FEV_1 to the nearest 5 ml, a minimum change of + 190 ml would be necessary if 95% confidence was required in the change being due to reasons other than chance. The exclusion of the patient group—that is low FEV_1—makes little difference to the calculation. Had the subject been moderately active before the test a larger change would have to be taken as the criterion of a significant change.

Discussion

Interest has recently been renewed in the most reliable value of FEV_1 derived from one set of manoeuvres.[4,5] In the present study the highest FEV_1 from three technically acceptable manoeuvres was used as this was the current practice in our laboratories. This choice was based on the recommendations made by the American Thoracic Society,[1] and values obtained in this way have been used in many clinical trials.[4] Although such values have been shown to be a slightly higher and less sensitive index than the mean of a given number of blows,[5] the coefficient of variation of the largest of three manoeuvres was nevertheless only marginally higher than indices based on mean values (2·56% compared with 2·20%).[5]

Variability of an index is normally expressed in terms of the standard error of repeated measurements made on one occasion.[16] The spontaneous variation between two sets of measurements, however, is of obvious relevance when assessing the significance of a difference in FEV_1 before and after a given treatment. We therefore determined this variation over the 20 minute interval often used in our laboratory for assessment of response to bronchodilators. Because none of our subjects had an obstructive ventilatory defect, varying airway obstruction should not have been contributing to variability; the range of FEV_1 covered was wide, no learning effect was shown, and as no treatment was given a placebo effect was excluded. The differences observed during the 20 minute interval could therefore be expected to give a valid estimate of natural variability in FEV_1. This background information is necessary when assessing the significance of a change in FEV_1 after administration of a bronchodilator over a similar period.

Controversy still remains over the interpretation of bronchodilator tests. For example, should an absolute or a percentage change be used as the criterion of response?[10,11] What constitutes a therapeutically significant change as opposed to a statistically significant change? Should different criteria for response be used in normal subjects and in patients with varying respiratory disorders? Thus Cotes states that the variability of FEV_1 within one set of measurements is independent of the size of the index yet also suggests that a 10% increase in FEV_1 after bronchodilation would imply therapeutic benefit.[16] Nickerson *et al* showed that variability of FEV_1 within a single set of measurements differed in normal subjects and patients with cystic fibrosis and concluded that different criteria should be used for a significant change in these two groups—15% and 23% respectively.[8] Vale *et al*, on the other hand, did not show any difference in variability of FEV_1 within a single set of manoeuvres when studying patients with different FEV_1 levels and differing severity of airway obstruction and suggested the use of an absolute change (size unspecified) as the criterion of response in these patients.[10]

We found that variability between pairs of measurements was similar both in normal subjects and in patients with restrictive disorders. Furthermore, this variability when expressed in absolute terms was similar at all levels of FEV_1 as has been found for variability within one set of measurements.[10,16] Significant changes can therefore be defined by a single absolute value that would be valid at any FEV_1 level but cannot be defined by a single percentage value.

This study does not bear on therapeutic benefit, and it remains to be shown in patients with airway obstruction, and in particular in those with low FEV_1, whether it is valid to imply therapeutic benefit from a bronchodilator, if, say, a 10% change is observed,[16] when in absolute terms the change may be within the range of normal variability in FEV_1 and thus not significant.

Although this study did not show a relation between smoking habits and variability in FEV_1, it did, somewhat surprisingly, show that even in normal subjects moderate activity before the test affected the variability of FEV_1. This implies that unless patients are rested before the test or activity is taken into account, for example by using a larger change as the criterion of a significant response, the results may be incorrectly interpreted.

Reference

1 American Thoracic Society. ATS Statement—Snowbird Workshop on Standardization of Spirometry. *Am Rev Respir Dis* 1979;**119**:831–8.
2 European Community for Coal and Steel. Standardised lung function testing. *Bull Eur Physiopathol Respir* 1983;**19**(suppl 5):26,45–51.
3 Ferris BG. Epidemiology standardisation project. *Am Rev Respir Dis* 1978;**118**:1–120.
4 Ullah MI, Cuddy V, Saunders KB, Addis GJ. How many blows really made an FEV_1, FVC, or PEFR? *Thorax* 1983;**38**:113–8.

5 Oldham PD, Cole TJ. Estimation of the FEV_1. *Thorax* 1983;**38**:662–7.
6 Sobol BJ, Weinheimer B. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966;**21**:445–9.
7 Hruby J, Butler J. Variability of routine pulmonary function tests. *Thorax* 1975;**30**:548–53.
8 Nickerson BG, Lemen RJ, Gerdes CB, Wegmann MJ, Robertson G. Within subject variability and percent change for significance of spirometry in normal subjects and in patients with cystic fibrosis. *Am Rev Respir Dis* 1980;**122**:859–66.
9 Goldman AL. Influence of tracing selection technique on daily variation of FEV_1 and FVC. *Am Rev Respir Dis* 1979;**119**:218.
10 Vale JR, Gulsvik A, Kongerus J. Random error with the FEV_1—case for absolute values. *Lancet* 1981;ii:313.
11 Cotes JE. Absolute FEV_1 values. *Lancet* 1981;ii:423.
12 Medical Research Council. *Questionnaire on respiratory symptoms*. London: Medical Research Council, 1976.
13 Rossiter CE. Contribution to discussion. *Scandinavian Journal of Respiratory Disease* 1976;**57**:315–6.
14 Daniel WW. *Applied non-parametric statistics*. Boston: Houghton Mifflin, 1978: 200–6.
15 Snedecor G, Cochran W. *Statistical methods*. 6th ed. Ames, Iowa: Iowa State University Press, 1967.
16 Cotes JE. *Lung function. Assessment and application in medicine*. 4th ed. Oxford: Blackwell Scientific Publications, 1979:109,318.