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. 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; 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; 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 and between measurements made hours or days apart, 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 a response to bronchodilators. 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.

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 Symptoms Questionnaire. 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; twelve 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...
Short term variability in FEV₁

the hospital building, moderate activity as outdoor walking, stair climbing, or manual work.

FEV₁ 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—either one or two further manoeuvres were performed until three technically acceptable results had been obtained. The highest FEV₁ from the three acceptable results was recorded and will be referred to as FEV₁. FEV₁ was again recorded in similar fashion after 20 minutes, this measurement being referred to as FEV₁.

The results for the normal subjects were divided into two subgroups to permit comparison of variability at different levels of FEV₁; those with initial FEV₁ greater than 4 l (n = 26) known as the high FEV₁ group; and those with FEV₁ of 4 l or less (n = 28) known as the mid-FEV₁ group. These subgroups were compared with the patient group (n = 13) known as the low FEV₁ group.

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

Statistical Analysis

The variability or change in FEV₁ was expressed both as absolute change (AC), where AC = FEV₁,II – FEV₁,II, and as percentage change (PC), where PC = (FEV₁,II – FEV₁,II) × 100/baseline. The baseline was defined as (FEV₁,II + FEV₁,II)/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. The baseline follows the recommendation of the European Community for Coal and Steel.

Table 1 Mean changes in forced expiratory volume in one second (FEV₁) over 20 minutes

<table>
<thead>
<tr>
<th></th>
<th>All subjects (n = 67)</th>
<th>Patients (n = 13)</th>
<th>Normal subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of FEV₁</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>recorded (1 BTPS)</td>
<td>1.1–6.3</td>
<td>1.1–2.25</td>
<td>2.4–4.0</td>
</tr>
<tr>
<td>Mean (SD) AC (ml BTPS)</td>
<td>+22.5 (170–1)</td>
<td>+26.2 (138–1)</td>
<td>+36.8 (204–7)</td>
</tr>
<tr>
<td>Mean (SD) PC (%)</td>
<td>+0.7 (5–6)</td>
<td>+1.6 (8–0)</td>
<td>+0.9 (6–2)</td>
</tr>
</tbody>
</table>

FEV₁—initial forced expiratory volume in one second; AC—absolute change; PC—percentage change.

Table 2 Comparison of mean changes in FEV₁ AC—absolute change; PC—percentage change.

<table>
<thead>
<tr>
<th>Activity before test</th>
<th>Seated (n = 31)</th>
<th>Mild (n = 27)</th>
<th>Moderate (n = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD) AC (ml BTPS)</td>
<td>–34.5 (102–5)</td>
<td>+29.6 (119–1)</td>
<td>+197.8 (325–5)</td>
</tr>
<tr>
<td>Mean (SD) PC (%)</td>
<td>–0.9 (4–5)</td>
<td>+1.1 (4–1)</td>
<td>+5.3 (9–4)</td>
</tr>
</tbody>
</table>

AC—absolute change; PC—percentage change.
### 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 (\( p < 0.001 \)) and percentage change (\( p < 0.001 \)) 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 \( p < 0.003 \)). The significant effect of activity was also found when the patient group was removed from the analysis (\( p < 0.003 \) for difference between the groups and \( p < 0.002 \) 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 (\( p > 0.05 \)). 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 (\( p > 0.05 \), 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 (\( p < 0.001 \)). The moderate activity group of nine subjects was therefore considered to be 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.)

There was no significant difference in the mean absolute change between the three \( FEV_1 \) subgroups (\( p = 0.3 \)), and there was no consistent trend towards larger absolute change as the baseline \( FEV_1 \) increased. When the mean percentage change was compared for the three groups, however, there was a highly significant difference between subgroups (\( p < 0.001 \)) with the percentage change decreasing as the baseline \( FEV_1 \) increased (negative linear trend, \( p < 0.001 \)).

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). As a one tailed test was chosen because of its greater relevance in clinical situations. It may be seen that...

### 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

<table>
<thead>
<tr>
<th>Probability of change occurring by chance (p%)</th>
<th>Change in ( FEV_1 ) (ml)</th>
<th>Normal subjects only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>+229</td>
<td>+234</td>
</tr>
<tr>
<td>5</td>
<td>+191</td>
<td>+195</td>
</tr>
<tr>
<td>10</td>
<td>+148</td>
<td>+151</td>
</tr>
<tr>
<td>20</td>
<td>+97</td>
<td>+99</td>
</tr>
</tbody>
</table>

One tailed analysis, where change = \( t_p \times 2 \times \) 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.
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. 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, and values obtained in this way have been used in many clinical trials. Although such values have been shown to be a slightly higher and less sensitive index than the mean of a given number of blows, 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%).

Variability of an index is normally expressed in terms of the standard error of repeated measurements made on one occasion. 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 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? 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. 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. 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.

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. 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, 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**


10 Vale JR, Gulsvik A, Kongerus J. Random error with

Tweeddale, Merchant, Leslie, Alexander, McHardy

the FEV₁—case for absolute values. Lancet 1981;ii:313.


