Thorax 1983;**38**:113–118

How many blows really make an FEV_1, FVC, or PEFR?

MAHMOOD I ULLAH, VINCENT CUDDIHY, KENNETH B SAUNDERS, GEORGE J ADDIS

From the Department of Medicine, Middlesex Hospital, London; the University Department of Materia Medica, Stobhill General Hospital, Glasgow; and the Chest Unit, Heathfield Hospital, Ayr

ABSTRACT We have collected peak expiratory flow rates, one-second forced expiratory volumes, and forced vital capacities in sets of 10 or 20 values at one-minute intervals from 30 normal, 49 asthmatic, and 26 bronchitic subjects. Analysis shows that the derivatives are compatible with a normal distribution of the values in the sets, so that the true value is best represented by the arithmetic mean of all valid attempts. One-third of all subjects showed skewness in one or more indices but these were equally divided between positive and negative directions. There is no sign of the dominant negative skewness that would result if the true value was indeed a maximum, which could be approached or equalled but never exceeded. There is no sign that repetition worsens performance. Seventy-two subjects showed no regression in any index and those of the remainder who deteriorated were balanced by equal numbers in all categories who improved. There is a significant tendency for both the highest and the lowest values to occur in the earlier part of any series. Probability theory suggests that this is a statistical phenomenon. The best estimate of the true value of these indices is probably the mean of as many observations as can be conveniently obtained and the data can be treated statistically as if they were a sample from a normally distributed population.

In 1846 Hutchinson reported his invention of the spirometer and introduced the concept of vital capacity to designate the volume of a forced maximal expiration from full inspiration, reporting the values obtained from 3000 subjects.[1]

The maximum forced expiratory manoeuvre expressed as the peak expiratory flow rate (PEFR) and forced expiratory volume in one second (FEV_1), as well as the forced vital capacity (FVC), are widely used as indirect measurements of airway diameter to measure the effect of drugs and in epidemiological research. Hutchinson advised that for forced vital capacity, after a single training attempt, the mean of three satisfactory attempts should be taken as most representative of the true value. In 1949 Gilson and Hugh-Jones[2] showed that for one normal subject performing these manoeuvres in batches several times a day over about three weeks the maximum value was produced at the 251st attempt out of 400 and that the results conformed to a normal frequency distribution. This valuable observation was not extended to asthmatic or bronchitic patients and perhaps because of this its implications for measurement have been ignored. Subsequently the Medical Research Council advised taking the mean of the third, fourth, and fifth of a series of technically satisfactory expirations after rejecting the first two[3]; but Freedman and Prowse thought that the mean of the second and third of three attempts within 10% would do.[4] Hughes and Empey,[5] Hruby and Butler,[6] and Nathan *et al*[7] advise taking the best of three values, as did the American Thoracic Society[8]; while Cotes[9] and Tager *et al*[10] recommend the mean of the best three of five. For the FVC Cotes advises the mean of the last two of six attempts.[9] Clearly therefore there is no real agreement on how these values should be determined. The mathematical and clinical aspects were well discussed by Oldham.[11] Most drug studies use the highest value from three attempts[12–18] and, usually in bronchodilator trials, the highest value in the control period is compared with the highest value from three similar attempts after the drug has been administered. A post-drug measurement higher than the control shows

Address for reprint requests: Dr MI Ullah, Heathfield Hospital, Heathfield Road, Ayr KA8 9DZ.

bronchodilatation, and if the post-drug measurement is equal to or lower than the best control value then the drug is said not to have had any effect. More than three forced expiratory manoeuvres are rarely carried out because it is widely believed that fatigue or repetition itself may result in deterioration of function.[4 7 10 12] None of these studies reports observation of more than five repetitions of the maximal forced expiration and in none was the question of deterioration with repetition actually examined. In previous studies, however, we did not observe significant deterioration when using a greater number of measurements.[19 20]

This paper reports the results of an investigation of the problem of whether the indices of maximum forced expiration really do show a deterioration with repetition or a significant negative skewness, or whether the results can be regarded as having the characteristics of a normal frequency distribution.

Methods

We chose 75 patients (table) who were all familiar with the apparatus in the pulmonary function laboratory, having previously attended at least once for routine assessment. They were included if their FEV_1 had previously shown an acute improvement of at least 10% after administration of isoprenaline aerosol. Twenty-two patients had atopic asthma, defined as periodic attacks of wheezing associated with positive skin-prick test reactions and sputum or blood eosinophilia; and 27 had non-atopic asthma. Twenty-six patients had chronic bronchitis by clinical criteria.[2] Patients receiving corticosteroid treatment for asthma were not excluded provided that the dose had remained unchanged for several weeks. All bronchodilator treatment was stopped at least 12 hours before they attended the laboratory at midday. Smoking and drinking tea and coffee were not allowed for at least four hours before or during the assessment. Ventilatory function was determined by measuring PEFR (Wright peak flow meter) and FEV_1 and FVC (recorded on a bellows-type digital spirometer: McDermott, Garw Electronics). Measurements of PEFR and of FEV_1 and FVC were made alternately at one-minute intervals during a period of 20 minutes. In 13 of the patients (all non-atopic) following the same protocol over this same period, 20 measurements each of PEFR, FEV_1 and FVC were obtained with a Lilly pneumotachograph driving an electronic spirometer (Mercury Electronics, Glasgow), which gave simultaneously the three measurements from one forced expiratory manoeuvre. Informed consent was obtained from all the patients and the approval of the hospitals' ethical committees was obtained. Thirty normal, healthy volunteers carried out the protocol after instruction and practice in the technique.

Data on patients and normal subjects producing maximal forced expirations

	Patients	*Normal subjects*
Number	75	30
Male	40	16
Female	35	14
Age range (y)	19–67	18–59
Atopic	22	
Non-atopic	27	
No with chronic bronchitis	26	
FEV_1 (% predicted) range	25–116	75–121

Statistical analysis used standard computerised methods, including calculation of means and moments, skewness and kurtosis, Student's paired t tests, regression analysis, and χ^2 and Kolomogorov-Smirnov tests.

Skewness is a measure of the symmetry of the distribution and is calculated by the formula coefficient of skewness:

$$\frac{(x - \bar{x})^3}{n.SD^3}$$

and for a normal distribution this value should lie in the range of ±0·5. Values below this are negatively skewed with a long tail of low values and above this are positively skewed with a tail of high values.

Kurtosis is a measure of the height of a frequency distribution related to its width and is calculated by the general formula coefficient of kurtosis,

$$\frac{(x - \bar{x})^4}{n.SD^4}.$$

For a normal frequency distribution this value should be around 3. Flatter curves give lower values, and the more peaked the curve the higher the value.

Results

Figure 1 shows the distribution of the values of relative skewness obtained in the 62 patients for their series of 10 maximal forced expirations. The great majority of the values are within two standard deviations of the mean and indicate symmetrical distributions. Chi-square and Kolomogorov-Smirnov tests confirm that these PEFR, FEV_1, and FVC relative skewness values do not differ significantly from normal. In each case the mean values are close to zero. There is no sign of the dominant negative skewness which would result if fatigue and so on

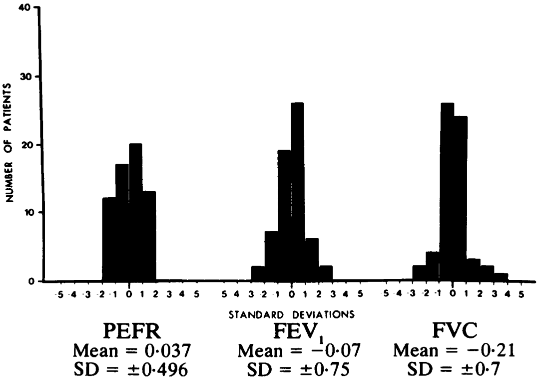

Fig 1 *Distribution of skewness values for the groups of 10 maximal forced expirations from each of 62 patients.*

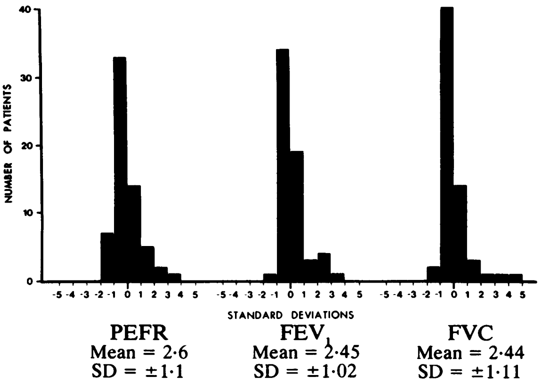

Fig 2 *Distribution of kurtosis values for the groups of 10 maximal forced expirations from each of 62 patients.*

consistently caused results less than the highest. There is no evidence that this occurred even in individual patients since the same set of forced expirations could yield FVC values which were skewed and PEFR and FEV_1 values which were not.

The mean kurtosis values for PEFR, FEV_1, and FVC are each close to the value for an ideal normal curve. Figure 2 shows the distribution of kurtosis values. A χ^2 test on the cumulative frequencies shows that the distributions are not normal ($p < 0·02$ in each case). They show positive skewness, with each modal value near to but lower than its corresponding mean and a long tail of low frequencies above the mean. This indicates that the most frequent distribution was more peaked than the ideal normal distribution. These distributions are, however, still consistent with the derivation of individual kurtosis values from normal frequency distributions and reflect the smallness of the coefficient of variation found. Figure 3 shows the distribution of the regression coefficients (r) for PEFR, FEV_1, and FVC with repetition. These are symmetrical about means which approximate to zero and χ^2 testing confirms that they are not different from normal. Only one value lies outside two standard deviations from the respective means, −0·95 for one subject's FVC values. The PEFR and FEV_1 values from the same forced expiration did not show this reduction with repetition and there is no evidence here that

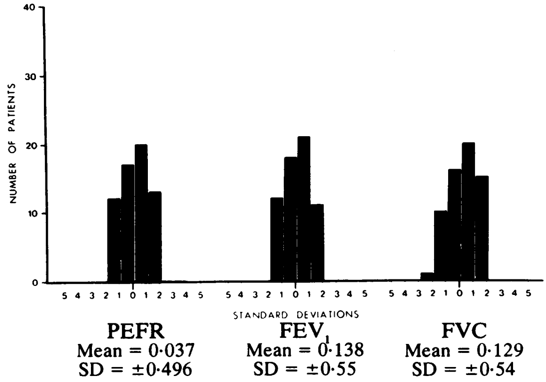

Fig 3 *Frequency distribution of regression (r) values with repetition: distribution of r values expressed as standard deviations of the group mean with repetition from 10 maximal forced expirations in each of 62 patients.*

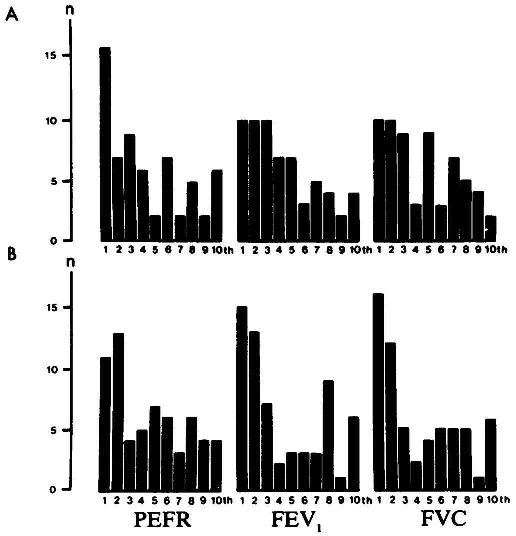

Fig 4 *Distribution of order of occurrence of highest (A) and lowest (B) values in 10 successive attempts in 62 patients.*

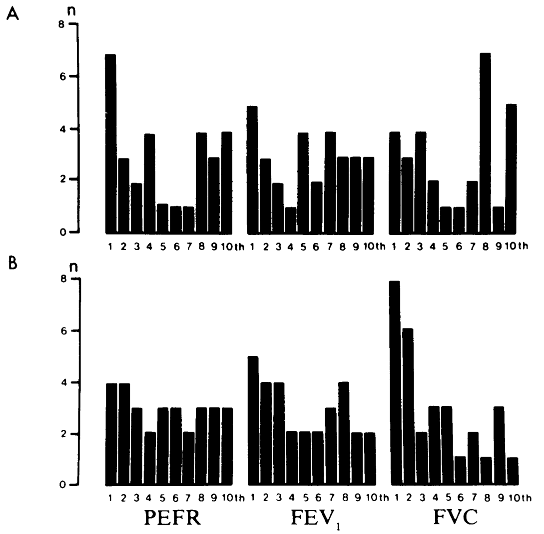

Fig 5 *Distribution of order of occurrence of highest (A) and lowest (B) values in 10 successive attempts at maximal forced expirations in 30 normal subjects.*

the group showed any consistent change on repeating their efforts.

Figure 4 shows the frequency distribution of the order of occurrence of the highest and lowest values in the 10 attempts in 62 patients. Figure 5 similarly shows the highest and lowest values in the controls. There was a significant effect shown in the reducing frequency of occurrence of the lowest FVC value with repeated attempts in the control subjects ($r = 0{\cdot}75$, $p = 0{\cdot}01$) but no correlation of any sort in any of the other results shown in figures 4 and 5. Some of the patients producing their highest value in the first three attempts also produced their lowest value in the same three attempts. Only a third of the patients produced their highest value in the first three attempts.

Of the 13 patients who carried out 20 forced expiratory manoeuvres at one-minute intervals using the electronic spirometer, only two showed significant (positive) correlations with repetition in all three indices. One showed a significant negative correlation in PEFR and FEV_1. The rest showed no change in any index.

No significant differences were noted when the mean of the first five readings was compared with the mean of the second five readings in 62 patients for all three indices. Similarly, in the 13 patients using the electronic spirometer there was no significant difference between the mean of the first 10 readings and the mean of the second 10 readings. Furthermore, in the 62 patients the variance of the first five readings was the same as the variance of the second five readings. The results for the variance of the first 10 readings and the second 10 readings in the patients using the electronic spirometer were the same. The mean coefficient of variation for PEFR, FEV_1, and FVC was very similar: 5·3, 4·8, and 3·8 respectively (based on the means and standard deviations of each measurement on 62 patients).

Discussion

These findings show that repeating maximal forced expirations at one-minute intervals as often as 20 times gives results which have a normal frequency distribution and a coefficient of variation as small as can be expected in any biological measurement.

The self-evident truth that subjects cannot exceed their maximum value, but may do less well, clearly is not attacked by these findings; but the assumption derived from it, that repeated attempts must give results which are negatively skewed, is now untenable. We have shown that the highest value in the first few attempts is not necessarily the highest value achieved—in fact, half of the patients achieved the highest values in the fourth and subsequent attempts in the series of 10.

If only the patients' best values are to be used then more than three measurements are required. The question then arises of how many attempts are needed to establish the best estimate of the true value, and hence the most powerful discriminant in detecting drug effects. From the subject's and the experimentalist's point of view one would be the most desirable number. From the statistical viewpoint the more the better.

The choice then is determined by the statistical test which is most appropriate. We have previously used Student's *t* test[19] and the Mann Whitney U test.[20] There is no agreement about the minimum number acceptable for these tests but we have found in practice that seven satisfactorily performed manoeuvres give a small coefficient of variation and provide sufficient numbers for significant results with even quite small changes in mean values. We could argue from probability theory that with repeated attempts there is a diminishing likelihood that the highest value will occur later in a series like this. But our results are most compatible with the theory that maximal forced expirations are normally distributed, randomly occurring, stochastically independent variables. The probability that any particular effort will be the greatest of a series is not influenced by the previous efforts in that series and it is therefore numerically equal to the reciprocal of the total number in the series.

Freedman and Prowse found that 85% of experienced patients achieved the highest value with the first forced expiratory manoeuvre out of five attempts for the FEV_1.[4] Our results with patients who are all "experienced" show that the highest FEV_1 occurred in only 16% of patients in the first attempt. Undoubtedly in the study by Freedman and Prowse the results could have been affected by the fact that the attempts in their series of five were repeated at approximately 15-second intervals. Gayrard *et al*[21] and Orehek *et al*[22] have suggested that after a maximum expiratory manoeuvre in asthmatic patients there is an immediate increase in airways resistance. They used only two maximal expiratory manoeuvres but did not state the interval between the two manoeuvres or show how long the increase in airways resistance persisted. In our study subjects carried out up to 20 maximum expiratory manoeuvres but at one-minute intervals, which may have given the airways time to recover.

In most bronchodilator drug studies all results for a group of individuals are combined in an attempt to provide sufficient data to detect significant changes. Single measurements of PEFR, FEV_1, FVC, etc, in an individual do not permit evaluation of significant intra-individual effects or changes. We have shown elsewhere that values obtained by analysis of repeated attempts can be used to construct individual dose-response curves for bronchodilators[23] and can also be used as the effect measurement in pharmacokinetic modelling studies.[24] We have also shown previously that individual responses may differ appreciably from the group response and by acceptance of the group results some subjects may suffer.[20] Review of published reports show that most investigators in pharmacological research use the highest value of three forced expiratory manoeuvres on the assumption, now shown to be invalid, that the distribution is negatively skewed. It is worth considering how the customary "selective" methods of obtaining data may have influenced results. Discarding observational data obviously increases the likelihood of unconscious observer bias and discarding low values will tend to give a higher estimate of the true value, but if the procedure is applied consistently it will not create bias in comparisons. Even if the Medical Research Council recommendation[3] is followed the mean result expressed as a single measurement does not permit statistical assessment of intra-individual change. These methods all diminish sensitivity, the "best of three" or other single-value estimates being the least sensitive. Comparisons based on these methods will detect only the most powerful drug effects and therefore tend to develop type II statistical errors—that is, fail to find a difference which really exists.

We have shown that for each of the three indices PEFR, FEV_1, and FVC a very small non-significant number of patients did show deterioration with repetition but that a greater, but also non-significant, number showed an improvement in performance with repetition. Most patients were found in the no-change category. This is in keeping with the symmetrical distribution of the regression, and with chance as the explanation of such increase and decrease as we have observed.

Our finding that there was no significant difference between the means of the first and the second five values, or between the means of the first and the second 10 values in the group using the electronic spirometer, shows that the mean of several readings is more reproducible and therefore more typical of the patient's real performance. The fact that there is no difference in the variance within individuals between the first and the second half of the control period suggests that the initial measurements are all acceptable and should not be discarded. The mean value and standard deviation therefore may be more easily obtained and thereby permit precise assessment of intra-individual change. The measurement of peak expiratory flow rate is said not to be particularly reproducible and therefore not a satisfactory substitute for the FEV_1.[9] Our results show, however, that the mean coefficient of variation for each index was similar at around 5%, so that all three indices seem equally acceptable.

We advise that the practice of using the highest value from three forced maximum expiratory manoeuvres should be abandoned and that the mean of as many attempts as can be obtained conveniently should be used. Repeated attempts are particularly valuable in pharmacological research, rendering it possible to evaluate within-individual changes.

References

1 Hutchinson J. On the capacity of the lungs and on respiratory functions. *Med Clin Trans* 1846;**29**:137–252.

2 Gilson JC, Hugh-Jones P. The measurement of the total lung volume and breathing capacity. *Clin Sci* 1949;**7**:185–216.

3 Medical Research Council. Definition and classification of chronic bronchitis for clinical and epidemiological purposes. *Lancet* 1965;i:775–80.

4 Freedman S, Prowse K. How many blows make an FEV_1? *Lancet* 1966;ii:618–9.

5 Hughes DTD, Empey DW. Ten years' experience in running a pulmonary function laboratory. *Br Med J* 1972;iv:470–2.

6 Hruby J, Butler J. Variability of routine pulmonary function tests. *Thorax* 1975;**30**:548–53.

[7] Nathan SP, Lebowitz MD, Knudson RJ. Spirometric testing: number of tests required and selection of data. *Chest* 1979;**76**:384–8.

[8] American Thoracic Society Statement—Snowbird workshop on standardisation of spirometry. *Am Rev Respir Dis* 1979;**119**:831–8.

[9] Cotes JE. *Lung function*. 4th ed. Oxford: Blackwell, 1979:10, 97, 103.

[10] Tager I, Speizer FE, Rosner B, Prang G. A comparison between the largest and three last of five forced expiratory manoeuvres in a population study. *Am Rev Respir Dis* 1976;**114**:1201–3.

[11] Oldham PD. *Measurement in medicine*. London: English Universities Press, 1968.

[12] Choo-Kang YFJ, Simpson WT, Grant IWB. Controlled comparison of the bronchodilator effects of three beta-adrenergic stimulant drugs. *Br Med J* 1969;ii:287–9.

[13] Duncan D, Paterson IC, Harris D, Crompton GK. Comparison of the bonchodilator effects of salbutamol inhaled as a dry powder and by conventional pressurised aerosol. *Br J Clin Pharmacol* 1977;**4**:669–71.

[14] Hetzel MR, Clark TJH. Comparison of salbutamol rotahaler with conventional pressurised aerosol. *Clin Allergy* 1977;**7**:563–8.

[15] Madsen BW, Tandon MK, Paterson JW. Cross-over study of the efficacy of four β_2-sympathomimetic bronchodilatory aerosols. *Br J Clin Pharmacol* 1979;**8**:75–82.

[16] Pierce RJ, Allen CJ, Campbell AH. A comparative study of atropine methonitrate, salbutamol and their combination in airways obstruction. *Thorax* 1979;**34**:45–50.

[17] Allen CJ, Campbell AH. Dose response of ipratropium bromide assessed by two methods. *Thorax* 1980;**35**:137–9.

[18] Simonsson BG, Jacobs FM, Nadel JA. Role of autonomic nervous system and the cough reflex in the increased responsiveness of airways in patients with obstructive airways disease. *J Clin Invest* 1967;**46**:1812–8.

[19] Addis GJ, Thorp JM. Effects of oxprenolol on the airways of normal and bronchitic subjects. *Europ J Clin Pharmacol* 1976;**9**:259–63.

[20] Ullah MI, Newman GB, Saunders KB. Influence of age on response to ipratropium and salbutamol in asthma. *Thorax* 1981;**36**:523–9.

[21] Gayrard P. Orehek J, Grimaud C, Charpin J. Bronchoconstrictor effects of a deep inspiration in patients with asthma. *Am Rev Respir Dis* 1975;**111**:433–9.

[22] Orehek J, Gayrard P, Grimaud C, Charpin J. Effect of maximal respiratory manoeuvres on bronchial sensitivity of asthmatic patients as compared to normal people. *Br Med J* 1975;i:123–5.

[23] Barclay J, Whiting B, Meredith PA, Addis GJ. Theophylline-salbutamol interaction: Bronchodilator response to salbutamol at maximally effective plasma theophylline concentrations. *Br J Clin Pharmacol* 1981;**11**:203–8.

[24] Whiting B, Kelman AW, Barclay J, Addis GJ. Modelling theophylline response in individual patients with chronic bronchitis. *Br J Clin Pharmacol* 1981;**12**:481–7.