

The flow-volume loop: reproducibility of air and helium-based tests in normal subjects

J B MACDONALD AND T J COLE

From City Hospital, Nottingham and MRC Dunn Nutrition Unit, Cambridge

ABSTRACT A statistical technique new to lung function testing is described and has been applied to a study of flow-volume loop tests in normal subjects. Maximal expiratory flow-volume loops were performed on air, then breathing an 80% helium : 20% oxygen mixture for three deep breaths (3B) then three minutes (3M) twice daily for five consecutive days on 25 asymptomatic subjects (eight smokers and 17 non-smokers). The tests were performed on an Ohio differentiating spirometer and recorded on a storage oscilloscope. The standard lung function tests—forced vital capacity (FVC), forced expiratory volume in one second (FEV_1) and peak expiratory flow rate (PEFR)—were the most reproducible. Maximal expiratory flow rate at 50% vital capacity (MEF_{50}) and at 75% vital capacity (MEF_{75}) were passably reproducible. MEF_{50} and MEF_{75} were as reliable breathing helium as breathing air, with 3B as reliable as 3M. However ratios of MEF helium/MEF air were very unreliable indeed, for both MEF_{50} and MEF_{75} and 3B and 3M. Thus percentage improvement in MEF after breathing helium appears to be an unreliable test in normal subjects. Helium isoflow volume was also very poorly reproducible with, on average, more variation within a single subject than between subjects. These conclusions apply to normal subjects and will require further work to assess them in specific diseases. However the very poor performance of isoflow volume and flow ratios in normal subjects suggests that it may be relatively difficult to distinguish normal subjects from patients with small airways obstruction using these tests. An application of canonical variate analysis to the data is described. This statistical technique assesses which indices or combinations of indices vary independently and hence are measuring qualitatively different parameters. Almost all the data variation was encompassed by variations in four parameters. Thus there are probably only four features in the flow-volume curve which can be used to discriminate between individuals. Repeating this analysis in patients with a specific disease should enable the best tests for use in that particular disease to be defined precisely. Canonical variate analysis should also be a useful method of assessing a new lung function test, since it assesses a test's reliability and more importantly whether it contributes any new information not given by older tests.

Flow-volume loops are being used increasingly in testing lung function, particularly in assessing small airways obstruction. There has been considerable discussion of which indices to use, and whether loops obtained breathing an 80% helium : 20% oxygen mixture (He) contribute extra information. McFadden and Linden¹ first proposed maximal expired flow at 50% of vital capacity (MEF_{50}) and Mead *et al*² suggested flow at 75% vital capacity (MEF_{75}). Indices based on comparing air and He loops have recently become

popular. Hutcheon *et al*,³ Dosman *et al*,⁴ and Gelb *et al*⁵ have suggested helium isoflow volume (isoV)—the volume at which the flows with air and helium become equal, expressed as a percentage of forced vital capacity (FVC). These authors have shown that isoV is probably the most sensitive test of small airways obstruction currently available. Dosman *et al*⁴ have also suggested that the percentage increase in MEF_{50} breathing He may be a more specific index of small airways calibre than isoV, although rather less sensitive.

However these tests must be shown to be repeatable as well as sensitive if they are to be widely

Address for reprint requests: Dr J B Macdonald, Ayrshire Central Hospital, Irvine, Ayrshire KA12 8SS.

used. We present here a method of assessing how much practical contribution each test makes, and report the results of a pilot study using normal subjects.

Subjects and methods

Twenty-five subjects were studied (12 men and 13 women), comprising 17 lifelong non-smokers and eight smokers with no respiratory symptoms, with a mean age of 28.6 years (range 19–44 years). Each subject had maximal expiratory flow volume (MEFV) and volume-time curves measured simultaneously on air, then MEFV curves after three vital capacity manoeuvres (3B) breathing the helium/oxygen mixture then after breathing the mixture for three minutes (3M). After a further 10 minutes breathing air an end-expiratory sample was taken for helium analysis and the procedure repeated. This sequence was performed each day for five consecutive days, giving 10 sets of flow-volume loops per subject. The curves were obtained using an Ohio 840 dry spirometer (Ohio Medical Products, Madison, Wis) displayed on a Tektronix 510 3N storage oscilloscope (Tektronix Inc, Beaverton, Oregon), and recorded by polaroid photography. On each set of curves the following parameters were measured: maximum expiratory peak flow rate (PEFR), forced expiratory volume in one second (FEV₁), forced vital capacity (FVC), MEF₅₀ and MEF₇₅ breathing air, MEF₅₀ and MEF₇₅ after both 3B and 3M of the helium mixture, and helium isoflow volume (isoV) obtained by comparison of the air and 3M curves, matched at residual volume. Subjects were given a short session to accustom them to the apparatus before starting the study, and they refrained from smoking for at least three hours before the daily session. Subjects who had any acute episode, such as a cold, likely to affect their performance during the study were excluded.

STATISTICAL METHOD

A good index of lung function should satisfy two criteria: it should be reproducible when carried out repeatedly on the same subject, but it should at the same time be sensitive to differences in respiratory function between subjects. These criteria can be seen as a low "noise" (variation within subjects) and a high "signal" (variation between subjects), concisely summarised by the ratio of signal to noise.

For a single lung function index, the signal-to-noise ratio is here defined as the ratio of variances between and within subjects— b^2/w^2 where b and w

are the between and within subject standard deviations. The individual measurements in the present paper have been transformed to natural logarithms before analysis on the grounds that the variability of each index tends to be larger in subjects with a larger mean. The resulting log standard deviations are closely similar to coefficients of variation. The procedure also converts ratios of indices to differences in log indices. Thus in addition to the signal-to-noise ($s-n$) ratios for each of the 10 tests studied, we can also calculate the $s-n$ ratios for the percentage improvement in MEF on breathing helium. This parameter is essentially MEF helium/MEF air, and when the variables are logged this becomes (MEF helium–MEF air). This index has a "signal" variance equal to the sum of the signal variances for MEF helium and MEF air, minus twice their signal covariance. The signal covariance is the product of the MEF helium and air standard deviations and their signal correlation. The "noise" variance is derived by a similar method.

If a number of indices of lung function are measured, each index has its own $s-n$ ratio, which represents its useful information. However the different indices are all intercorrelated, both between and within subjects, so that the information they contain is to some extent shared. Consider for example just two indices, x and x' , with standard deviations between and within subjects of b, w, b' , and w' respectively. Further let their correlations between and within subjects be r_b and r_w . If the two indices are added together to make a new index ($x+x'$), its $s-n$ ratio is given by

$$\frac{(b^2 + b'^2 + 2r_b \cdot b \cdot b')}{(w^2 + w'^2 + 2r_w \cdot w \cdot w')}$$

This may be better or worse than either of the individual $s-n$ ratios. Now let the combination of the two indices be weighted in an index of the form ($x+kx'$), so that b' and w' in the equation are replaced by kb' and kw' . It is possible to find a value for k which maximises the $s-n$ ratio of this combination, and which is better than either index alone. The constant k may be positive or negative; in particular, if it is set to -1 the index becomes ($x-x'$). This is the log of the ratio of the original indices, since each index has been logged. Thus ratios of indices are easily interpreted in the analysis.

The statistical technique used to estimate k is termed canonical variate analysis. It extends to as many variables as desired, and provides the largest possible $s-n$ ratio for a given set of indices, plus the appropriate weightings for each index to

achieve this ratio. However it also produces a series of subsidiary indices, all uncorrelated with each other, which again have s-n ratios which are as large as possible, but subject to their being uncorrelated. As a result of this constraint, the s-n ratios for successive new indices are progressively smaller, and the last few often hold very little information at all. There are as many new indices produced as there are original variables, so the analysis indicates how many of them provide useful information. Further details of the theory of the technique are given in the appendix.

Results

The log standard deviations between and within subjects for the 10 indices are shown in table 1 with the corresponding s-n ratios. The "between-subject" standard deviations ranged from 0.406 for isovolume down to 0.205, half as much, for FEV₁. Among the "within subject" standard deviations isovolume was again the highest at 0.487, while

the lowest was FVC with 0.0316, 15 times smaller. However the most important parameter was the s-n ratio. Hence the FVC (s-n ratio 49.8) was by far the best test in the group studied, with FEV₁ about half and PEFR one-third as good. The MEFs were only passably reproducible, with MEF_{7.5} performing consistently slightly better than MEF₅₀ in both air and helium. Isovolume had the lowest s-n ratio -0.7. Maximal expiratory flow rate values breathing helium (3B or 3M) appeared as reliable as those breathing air.

Table 2 shows the intercorrelations between indices, with the "between" correlations above the diagonal and the "within" below. The between correlations were not, surprisingly, larger than those within in nearly all cases, although many within correlations were significant. This is a measure of small changes in performance of the subjects during the study, but these are obviously far smaller than between subjects. These correlations enabled us to calculate the s-n ratios for the four helium/air flow rates (table 3). All four

Table 1 *Signal-to-noise ratios for flow-volume loop tests*

	PEFR	FEV ₁	FVC	MEF ₅₀ air	MEF ₅₀ 3B	MEF ₅₀ 3M	MEF _{7.5} air	MEF _{7.5} 3B	MEF _{7.5} 3M	IsoV
Log standard deviations between subjects	0.252	0.205	0.223	0.220	0.234	0.230	0.390	0.388	0.389	0.406
Log standard deviations within subjects	0.0653	0.0407	0.0316	0.0861	0.105	0.103	0.151	0.141	0.151	0.487
Signal-to-noise ratios	14.9	25.5	49.8	6.5	5.0	4.9	6.7	7.5	6.7	0.7

Table 2 *Within and between subject correlation matrices. All variables transformed to logs*

	<i>Between subject</i>										
	PEFR	FEV ₁	FVC	MEF ₅₀ air	MEF ₅₀ 3B	MEF ₅₀ 3M	MEF _{7.5} air	MEF _{7.5} 3B	MEF _{7.5} 3M	IsoV	
	1	0.693	0.701	0.536	0.570	0.490	0.220	0.187	0.121	0.444	PEFR
		1	0.956	0.703	0.733	0.678	0.590	0.609	0.565	0.152	FEV ₁
			1	0.543	0.616	0.534	0.345	0.379	0.333	0.335	FVC
PEFR	1			1	0.948	0.949	0.718	0.711	0.680	-0.029	MEF ₅₀ air
FEV ₁	0.291	1			1	0.983	0.600	0.598	0.568	0.137	MEF ₅₀ 3B
FVC	0.313	0.459	1			1	0.652	0.657	0.636	0.058	MEF ₅₀ 3M
MEF ₅₀ air	0.260	0.500	0.083	1			1	0.981	0.967	-0.506	MEF _{7.5} air
MEF ₅₀ 3B	0.085	0.111	-0.08	0.292	1			1	0.993	-0.514	MEF _{7.5} 3B
MEF ₅₀ 3M	0.125	-0.054	-0.125	0.151	0.466	1				-0.553	MEF _{7.5} 3M
MEF _{7.5} air	-0.016	0.127	-0.215	0.332	0.174	0.086	1			1	IsoV
MEF _{7.5} 3B	-0.019	0.066	-0.185	0.171	0.448	0.327	0.290	1			
MEF _{7.5} 3M	-0.012	0.014	-0.288	0.152	0.327	0.503	0.258	0.522	1		
IsoV	0.074	0.065	0.078	0.119	-0.137	-0.115	0.185	-0.264	-0.272	1	
	PEFR	FEV ₁	FVC	MEF ₅₀ air	MEF ₅₀ 3B	MEF ₅₀ 3M	MEF _{7.5} air	MEF _{7.5} 3B	MEF _{7.5} 3M	IsoV	
<i>Within subject</i>											

For r ≥ 0.215 p < 0.001
r > 0.171 p < 0.01

Table 3 Signal-to-noise ratios for helium-based proportional tests

	MEF ₅₀		MEF ₇₅	
	3B air	3M air	3B air	3M air
Signal-to-noise	0.422	0.343	0.190	0.296

ratios were very low indeed.

Canonical variate analyses carried out on all 10 indices gave the results shown in tables 4 and 5. In table 4 the first index had an s-n ratio of 68.3, appreciably greater than for any test alone: the best was 49.8 for FVC (table 1). The next three indices all have useful information with ratios greater than one, while the last six indices were all uninformative, their summed s-n ratios being only 1.49.

Table 5 gives the weightings for each of the original variables in making up the first four new indices. The weightings for each variable have been scaled to make the largest unity. The first index is almost entirely made up of FVC, with only tiny contributions from the other variables. The second index is essentially PEFr. The third and fourth variables are less obvious, but the third is essentially FVC-PEFr, which because of the log transformation is actually the log ratio FVC/PEFr. The fourth index involves nearly all the original variables, with PEFr, FEV₁, and MEF₇₅ positive and MEF₅₀ and FVC negative. The contribution of isovolume to all four indices is tiny.

Despite its popularity in clinical practice, FEV₁ also had remarkably small weightings in the four indices, probably because of its high correlation with FVC. Repeating the whole analysis with FEV₁

omitted shows that it contained almost no unique information. The sum of the first four s-n ratios with FEV₁ missing, at 90.6, was only 2.2 less than with FEV₁ included. Inspection of the coefficients shows that in each of the four indices the FEV₁ contribution was taken up by the FVC. Thus if the number of tests is to be reduced to simplify interpretation, FEV₁ could be excluded. Similarly isovolume made virtually no contribution. The 3B and 3M weightings were usually of the same sign and magnitude so 3B could be omitted. Repeating the analysis on the six remaining variables (table 6) showed that very little information had been lost. The accumulated s-n ratios for the first four indices fell from 92.8 to 87.5, but the weightings are easier to interpret. Indices 1 and 2 are unchanged but 3 is now a comparison of FVC to the flow rates while 4 contrasts MEF₅₀ with PEFr and MEF₇₅.

A large part of the discriminatory power of FVC is the result of its strong correlation with stature. The size of this "anatomical" component can be judged by standardising FVC for height and repeating the analysis. Standardisation can be achieved by expressing FVC as FVC/height² (Cole 1975, 1977), and the results are shown in table 7. As expected the first s-n ratio was substantially reduced from 65.5 to 38.3. However the other three s-n ratios were hardly affected and two of them actually increased.

End-expiratory samples were taken just before the day's second test. They showed a mean of 0.02% helium (range 0-0.08%). This very small residual concentration is unlikely to have any detectable effect. Ranking tests on the results confirmed this by showing no systematic difference between daily first and second tests for any variable.

Table 4 Signal-to-noise ratios of the 10 indices derived from canonical variate analysis in descending order

	Variable									
	1	2	3	4	5	6	7	8	9	10
Signal-to-noise ratio	68.3	11.7	7.4	5.4	0.69	0.37	0.17	0.15	0.07	0.04

Table 5 Coefficients of each of the 10 variables in the first four discriminants, all scaled so that the largest coefficient is unity

	PEFr	FEV ₁	FVC	MEF ₅₀ air	MEF ₅₀ 3B	MEF ₅₀ 3M	MEF ₇₅ air	MEF ₇₅ 3B	MEF ₇₅ 3M	IsoV
1	0.00	0.19	1.00	-0.07	0.02	0.04	0.06	0.02	0.06	0.00
2	1.00	-0.35	0.05	-0.11	0.28	0.03	-0.19	-0.26	-0.23	0.01
3	-0.76	0.00	1.00	-0.21	0.01	-0.10	-0.13	-0.11	0.02	0.02
4	1.00	0.58	-0.49	-0.99	-0.47	-0.70	0.17	0.23	0.25	0.01

Table 6 As table 5 but reduced to the six most important variables

	Signal-to-noise ratio	PEFR	FVC	MEF ₅₀ air	MEF ₅₀ 3M	MEF ₇₅ air	MEF ₇₅ 3M
1	65.5	0.00	1.00	-0.02	0.03	0.06	0.07
2	10.4	1.00	-0.35	-0.15	0.06	-0.15	-0.26
3	7.0	-0.62	1.00	-0.31	-0.16	-0.16	-0.07
4	4.6	-0.93	0.22	0.98	1.00	-0.33	-0.54

Table 7 As table 6 but with the effect of height removed

	Signal-to-noise ratio	PEFR	FVC	MEF ₅₀ air	MEF ₅₀ 3M	MEF ₇₅ air	MEF ₇₅ 3M
1	38.3	0.01	1.00	-0.02	0.02	0.09	0.09
2	12.9	1.00	-0.65	0.06	0.13	-0.05	-0.15
3	7.6	0.15	1.00	-0.56	-0.26	-0.15	-0.11
4	3.9	-1.00	1.00	0.78	0.98	-0.37	-0.58

Discussion

The aim of the study was to discover which of 10 lung function tests are most useful in discriminating between individuals in a group of normal subjects. Since the subjects are normal the discrimination has to rely on anatomical rather than pathological differences in lung function. However subclinical pathology may be present which would aid the discrimination. Thus our conclusions cannot necessarily be extrapolated to patients with respiratory disease. Nonetheless, it is difficult to explain the four new indices found purely on the basis of anatomical variation. The first index remains highly discriminatory with an s-n ratio of nearly 40, even after standardising FVC for height. The other three indices are unaffected by height standardisation, so are unlikely to be dependent on lung size.

Forced vital capacity and PEFR appear to be excellent discriminators since the first two new indices had a combined s-n ratio of 80 (table 4). Even after removing four variables and standardising FVC for height, the combined s-n ratio was still 51.2 (table 7). The excellent discriminating power of these well-tried tests is confirmed by the high s-n ratios of the individual tests shown in table 1. The value of the MEFs only appears in the third and fourth indices, and in table 1 they appear to be only moderately reproducible. This is mainly because of their larger noise components. Surprisingly the performance of MEF₇₅ was as good as MEF₅₀—perhaps even a little better. This reinforces the suggestion of Mead *et al*² of MEF₇₅ as a useful test. The role of the helium-based tests was more uncertain in our group. Helium isoflow volume was probably too unreliable to be useful: variation within an individual subject was, on

average, greater than variation between subjects. Zeck *et al*⁷ have also drawn this conclusion for normal subjects. Isoflow volume also made virtually no contribution to any of our first indices, suggesting that it contains little unique information.

The other helium-based tests are MEF₅₀ and MEF₇₅ after three breaths (3B) or three minutes (3M) breathing helium. Table 1 shows that both 3B and 3M were as reproducible as their parent air-based tests. However tables 5–7 show that the weightings given to the MEFs on air and helium were usually of the same sign and similar magnitude. This implies that 3B and 3M usually contain qualitatively similar information to the air-based tests. Thus duplicate air tests might well provide as much extra information as helium tests. Dosman *et al*⁴ have proposed the percentage increase in MEF₅₀ on breathing helium as an index of small airways disease. This is effectively MEF helium/MEF air. Table 3 shows that all these flow ratios were poorly reproducible. This would be expected if MEFs on helium and air contained similar information since in the ratio the information in the two MEFs cancels out. In our group 3B appeared as reliable as 3M and qualitatively similar. Thus three deep breaths of helium should be sufficient.

Another technique for assessing the relative importance of each test is by a stepwise procedure, by first finding the most important single variable then finding successive variables whose addition increases the information by the greatest amount. Starting with the best test, FVC (s-n ratio 49.8), the best test to add is surprisingly MEF₇₅ which raises the combined s-n ratio by 12.2. Since it is worth only 6.7 units by itself (table 1), this implies that MEF₇₅ standardised for FVC could be a use-

ful index, simply as the ratio MEF_{75}/FVC . The next most useful index to add is PEF_R (adding 8.6 units) then MEF_{50} (a further 3.1 units). These four variables give a combined s-n ratio of 73.7, about four-fifths of the figure for all 10 variables.

Canonical variate analysis is not new but does not appear to have been used previously for lung function tests. It has produced interesting results in this pilot study of normal subjects. It suggests, for instance, that there may just be four features of the flow-volume curve that can be used to discriminate between normal individuals. Repeating the analysis in patients with a specific disease will enable the best tests for use in that particular disease to be defined. With computer assistance present tests could be refined. For instance only three flow indices are commonly used from the flow-volume curve—PEFR, MEF_{50} , and MEF_{75} . Are these the best? Would MEF_{25} or MEF_{90} contain extra information in for example small airways disease? Would helium-based tests be better discriminators if a dense gas such as sulphur hexafluoride were chosen rather than air for comparison with helium, as suggested by Brooks *et al*?⁸ At present new lung function tests are developed and adopted in a very empirical way. We are confident that canonical variate analysis could not only rationalise the development of new indices but test whether they contribute any qualitatively new information.

We wish to thank Mrs M McDermott, MRC Pneumoconiosis Unit, Penarth for her constructive criticism.

Appendix

CANONICAL VARIATE ANALYSIS⁹

Assume the analysis is done using n lung function variables. Let B be the between subjects (signal) covariance matrix and W the within subjects (noise) covariance matrix, and let l be the vector of weightings for each variable which maximises the signal-to-noise ratio V . Then in matrix notation $V=l'B/l'W$ needs to be maximised. This is equivalent to finding the eigenvalues λ of the equation $(B-\lambda W)l=0$, where $|B-\lambda W|=0$. In general there are n distinct eigenvalues, each with an associated eigenvector l . The largest eigenvalue is the maximised s-n ratio V , and the appropriate weightings for each variable are contained in the corresponding eigenvector l . Successively smaller eigenvalues and eigenvectors correspond to combinations of variables which are uncorrelated with previous combinations, but which maximise the s-n ratio subject to this constraint.

A geometric explanation for this procedure is

roughly as follows, taking n to be three for simplicity. Imagine a three-dimensional graph, where each axis represents a variable. The scales are adjusted to take account of the size of the variable's within subject variation (noise), so that "noisy" variables have their scale expanded. Each subject's data are averaged, and plotted as a single point in this three-dimensional graph. Thus there are 25 points (subjects) suspended in mid-air, and they can be viewed from any direction. Assume the scatter of points is roughly ellipsoidal (like a rugby ball) so that the scatter is greatest in one particular direction (the major axis of the ball). The analysis determines this direction and the magnitude of the scatter is the s-n ratio. The direction is defined by the weighting of the variables.

Having identified the line of greatest scatter, the analysis as a next stage confines itself to a direction at right angles to this, which thus lies in a plane. The second eigenvalue and eigenvector define the maximised s-n ratio and its direction in this plane. The remaining scatter is at right angles to both the first and second directions, and its direction is thus already determined since there are only three dimensions. This direction is given by the third eigenvector.

References

- 1 McFadden ER, Linden DA. A reduction in maximum mid-expiratory flow rate: a spirographic manifestation of small airway disease. *Am J Med* 1972; **52**:725-37.
- 2 Mead J, Turner JM, Macklem PT, Little JB. Significance of the relationship between lung recoil and maximum expiratory flow. *J Appl Physiol* 1967; **22**:95-108.
- 3 Hutcheon M, Griffin P, Levison H, Zamel N. Volume of isoflow: a new test in detection of mild abnormalities of lung mechanics. *Am Rev Respir Dis* 1974; **110**:458-65.
- 4 Dosman J, Bode F, Urbanetti J, Martin R, Macklem PT. The use of helium-oxygen mixture during maximum expiratory flow to demonstrate obstruction in small airways in smokers. *J Clin Invest* 1975; **55**:1090-9.
- 5 Gelb AF, Molony PA, Klein E, Aronstam PS. Sensitivity of volume isoflow in the detection of mild airways obstruction. *Am Rev Respir Dis* 1975; **112**:401-5.
- 6 Cole TJ. Linear and proportional regression models in the prediction of ventilatory function. *J Roy Stat Soc (A)* 1975; **138**:297-338.
- 7 Zeck RT, Solliday NH, Cugell DW. Variability of the isoflow volume. *Am Rev Respir Dis* 1977; **115**(Suppl Abstr):392.
- 8 Brooks SM, Zipp T, Barber M, Carson A. Measurements of maximal expiratory flow rates in cigarette smokers and non-smokers using gases of high and low densities. *Am Rev Respir Dis* 1978; **118**:75-81.
- 9 Krzanowski WJ. The algebraic basis of classical multivariate methods. *Statistician* 1971; **20**:51-61.