Thorax 1989;**44**:474–479

Repeatability of ventilatory function measurements in a population survey of 7 year old children

D P STRACHAN

From the Department of Community Medicine, University of Edinburgh

ABSTRACT The within subject variability of forced vital capacity (FVC), forced expiratory volumes in one second (FEV_1) and half a second ($FEV_{0.5}$), peak expiratory flow (PEF), and flow rates at 25–75%, 75–85%, 25%, 50%, and 75% of expired FVC were assessed among 7 year old children from the general population. Within occasion variability in 232 children was lowest for FVC (coefficient of variation (CV) 5%) and FEV_1 (CV 4%), and greatest for end expiratory flow rates. The precision of measurement for FEV_1 supports its use for bronchial provocation tests, particularly those using a graded challenge. In this context the value of PEF (CV 7%) and mid expiratory flow rates (CV 11%) is limited by their poorer repeatability. Between occasion variability was assessed in 171 children tested at an interval of one to four weeks. The difference between the variances between occasions and within occasions was attributed to biological variation; this accounted for a substantial component of the between occasion variance in all indices, particularly FEV_1 (73%) and PEF (66%). Together, within subject variability, sex, and height accounted for about half of the measured variance between subjects for all indices except FVC (68%). These results have implications for epidemiological studies.

Introduction

During an investigation of the relation between the home environment and respiratory disease,[1] measurements of ventilatory function and exercise induced bronchial lability were performed in a population sample of 7 year old children. No information could be found relating to the repeatability of ventilatory function measurements among untrained children of this age, which was particularly relevant to the definition of abnormal exercise induced bronchial lability.

This paper describes an investigation of within subject variability in baseline spirometric values among a subsample of the children participating in the main survey. An attempt was made to evaluate the magnitude of two sources of variability. The variation between measurements made on the same occasion was attributed to "measurement error," including unreliable performance by the subject. The difference between the within subject variances between occasions and on the same occasion was attributed to "biological variation" and assumed to be error free.

Address for reprint requests: Dr D P Strachan, Department of Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT.

Accepted 7 March 1989

Methods

SUBJECTS

Details of the random population sample selected for the main survey are given elsewhere.[1] The children were aged 6½–7½ years in September 1986. Parental consent was requested for clinical examination of the children at school, and ethical approval was obtained from the paediatric/reproductive medicine ethics of medical research subcommittee of the Lothian Health Board and from the research committee of the Department of Education, Lothian Regional Council.

MEASUREMENTS

An unheated Fleisch type pneumotachograph ("Compact," Vitalograph Ltd, Buckingham) with a paediatric mouthpiece adapter was used to record each spirogram. The manufacturer reports the accuracy of flow measurement as ±3% in the range 0–15 l/s, and tests of linearity across this range show absolute errors of +0·042 l/s at 1 l/s, +0·068 l/s at 2 l/s, and +0·076 l/s at 3 l/s, with negative errors in the range 12–15 l/s (Vitalograph Ltd, personal communication).

The same instrument was used throughout the survey, and all expiratory manoeuvres were supervised by the author. At the start of each two hour session the pneumotachograph was calibrated volumetrically using a 1 litre precision syringe. Three litres were delivered at fast and slow flow rates to check the linearity of the flow integration. The instrument was checked periodically during the session for calibration drift. Indoor temperature was recorded at the time of calibration with a digital thermohygrometer (Protimeter Diagnostic Mark III, Protimeter PLC, Marlow) and spirometric indices were corrected to body temperature automatically.

The following spirometric indices were calculated from each expiration: forced vital capacity (FVC), forced expiratory volume in one second (FEV_1), forced expiratory volume in the first half second ($FEV_{0.5}$), peak expiratory flow rate (PEF), mid expiratory flow rate from 25% to 75% of expired FVC (FEF_{25-75}), end expiratory flow rate from 75% to 85% of expired FVC (FEF_{75-85}), and instantaneous flow rates when 25%, 50%, and 75% of FVC had been expired (FEV_{25}, FEF_{50}, and FEF_{75}). Back extrapolation was used automatically in the calculation of zero time for timed forced expiratory volumes. The maximal values of FVC, FEV_1, and $FEV_{0.5}$ from the two best spirograms were recorded, but all flow rates except PEF were recorded from the best spirogram, according to the recommendations of the American Thoracic Society.[2] For greater comparability with surveys that have recorded only peak expiratory flow rate, the maximum achieved PEF was analysed (this was not always from the best spirogram).

DESIGN OF THE STUDY

Children were tested in pairs by the author, assisted by a research nurse. Standing height was measured with a vertical rule to the nearest centimetre below. After a period of instruction and two practice attempts, each child performed three forced expiratory manoeuvres, according to the methods recommended by the American Thoracic Society.[2] Tests were performed in the standing position and nose clips were not used. Two further expirations were performed if the two best results of the first set of three were not within 5% or 100 ml of each other according to the "best test" criteria of the American Thoracic Society (that is, the spirogram with the greatest sum of FEV_1 and FVC).[2]

Four of the larger schools were revisited to obtain duplicate measurements by spirometry on the same children on different occasions. On the second visit to these schools further measurements were taken after a five minute interval to assess within occasion variability of spirometric indices. The test sessions were separated by an interval of one to four weeks during the months of January to March 1987, and were not necessarily at the same time of day. Wheezy children were included, but those having inhaler treatment were tested at least six hours after their last dose. Children with upper respiratory symptoms at the time of either test were excluded from the analysis of between occasion variabiliy.

ANALYSIS

The within subject variability of each spirometric index was investigated by examining the distribution of differences between pairs of readings obtained from the same subject. The mean of this distribution describes the effect of the order of the measurements and the variance of the differences represents twice the within subject variance of a single reading. Within subject variances were estimated from pairs of readings on the same occasion, and from pairs of first readings on different occasions. The coefficient of variation (CV) was derived from the within subject standard deviation divided by the mean value for the corresponding index in the study sample. The within occasion and between occasion variability of FEV_1 was investigated for children within the lowest, middle, and highest tertiles of the distribution of measured FEV_1.

The true within subject, between occasion variance for each measurement was estimated as the difference between the measured within subject variance between occasions and the variance on the same occasion. Reliability coefficients[3] were calculated as one minus the ratio of measured within subject, between occasion variance to the variance of the distribution of first measurements among the children participating in this repeatability study.

For assessment of proportional, rather than absolute, change in airflow, the variability of the measurement is more conveniently expressed on a logarithmic scale. The within occasion variability was therefore also expressed as the standard deviation of the log (base 10) of a single reading. The coefficient of variation on an arithmetic scale is approximately equal to the antilogarithm of this value minus one.

Challenge tests generally require repeated measurements of the same spirometric index at intervals after a fixed challenge, or after increasing doses of a pharmacological agent. Random errors of measurement may give rise to a spurious "false positive" result, even if there has been no true change in ventilatory function. The probability that such a false positive result would occur by chance was calculated from estimates of within occasion variance (expressed on a logarithmic scale) for different criteria of "abnormality" and for one or more comparisons between prechallenge and postchallenge recordings.

Results

WITHIN OCCASION VARIABILITY

Duplicate readings at the same test session were available for 232 children. These included 27 (11·6%) who had a history of wheeze in the past year, of whom eight had wheezed in the past month and four regularly had inhalation treatment for asthma. There was no substantial order effect for any of the indices: the average differences between the first and second measurements were all less than 2% of the mean value for the corresponding index; for FEV_1 the average difference was 0·6%.

The within occasion variability of each index is expressed in table 1 on an arithmetic scale, with the coefficient of variation, and as the standard deviation of the logarithm of each reading. In terms of the coefficient of variation, FEV_1 was the least variable measurement (SD 60 ml, CV 4·3%), closely followed by FVC (SD 81 ml, CV 5·0%). Although flow rates during early expiration (PEF and FEF_{25}) were more variable in absolute terms, the coefficient of variation was greater for measurements from the terminal part of the spirogram (FEF_{75} and FEF_{75-85}).

The result of applying estimates of within occasion variability to the circumstances of challenge testing is shown in table 2. For both FEV_1 and PEF the chance probability of obtaining one or more abnormal results increases with the number of comparisons made between postchallenge measurements and the baseline reading, and decreases as the cut off defining abnormality becomes more extreme. The greater within occasion variability of PEF is reflected in higher false positive rates. Thus a decline of 20% in FEV_1 could be expected by chance in only 0·05% of tests based on a single before and after comparison, but a similar decline in PEF would occur in 2·3% of tests. If 10 postchallenge measurements were each compared with the baseline, then a decline of 20% in FEV_1 would occur by chance in 0·5% of subjects; whereas 21% would show a similar decline in PEF.

BETWEEN OCCASION VARIABILITY

Results of spirometry on the two occasions were analysed for 171 children. These included 20 (11·7%) with a history of wheeze in the past year, of whom six had wheezed in the past month and three regularly had inhaled treatment for asthma. The variability of each index between occasions is shown in table 3. The proportion of the between occasion variance attributable to measurement errors (as estimated by within occasion variances from table 1) is indicated in the right hand column. "Measurement errors" accounted for about half of the between occasion variance for most of the indices, but for PEF and FEV_1 the "biological" component was greater.

The mean and standard deviation of the first measurements on the 232 children included in the repeatability study and the coefficient of reliability for each index are shown in table 4. The most reliable measures were FVC and FEV_1, for which within subject variability accounted for less than one quarter of the observed variance between subjects. Indices of flow during early expiration (PEF and FEV_{25}) were more reliable than FEF_{75} and FEF_{75-85}.

Table 1 *Within subject variability of spirometric indices in 232 children, derived from duplicate measurements on the same occasion*

	Standard deviation of a single measurement			
	Arithmetic scale		*Log (base 10) scale*	
Spirometric index	*SD*	*CV (%)*	*SD*	*Antilog SD*
FVC (ml)	81·25	5·0	0·0207	1·049
FEV_1 (ml)	60·28	4·3	0·0209	1·049
$FEV_{0·5}$ (ml)	75·73	7·7	0·0437	1·106
$FEV_{0·5-1}$ (ml)	54·38	12·7	0·0489	1·119
PEF (l/min)	13·06	7·0	0·0344	1·082
FEF_{25-75} (ml/s)	181·99	10·5	0·0529	1·130
FEF_{75-85} (ml/s)	142·24	18·0	0·0868	1·221
FEF_{25} (ml/s)	237·94	8·9	0·0448	1·109
FEF_{50} (ml/s)	208·95	11·1	0·0590	1·145
FEF_{75} (ml/s)	130·43	13·5	0·0686	1·171

FVC—forced vital capacity; FEV_1, $FEV_{0·5}$—forced expiratory volume in one second and in half a second; PEF—peak expiratory flow; FEF_{25-75}—mid expiratory flow rate from 25% to 75% of FVC; FEF_{75-85}—expiratory flow rate from 75% to 85% of FVC; FEF_{25}, FEF_{50}, FEF_{75}—instantaneous flow rates when 25%, 50%, and 75% of FVC has been expired; CV—coefficient of variation.

Table 2 *Chance probability (%) of an "abnormal" result in comparisons of forced expiratory volume in one second (FEV_1) and peak flow (PEF) with a baseline reading at the same test session, by criterion of abnormality and number of comparisons*

		Number of comparisons with baseline reading					
Criterion of "abnormality"		*1*	*2*	*3*	*4*	*6*	*10*
Largest reduction in FEV_1	>10%	6·02	11·69	17·01	22·01	31·12	46·28
	>15%	0·83	1·66	2·47	3·28	4·88	8·01
	>20%	0·05	0·10	0·15	0·20	0·30	0·50
Largest reduction in PEF	>10%	17·33	31·65	43·49	53·28	68·07	85·08
	>15%	7·32	14·11	20·40	26·23	36·64	53·26
	>20%	2·31	4·56	6·77	8·92	13·08	20·83

Table 3 *Within subject variability of spirometric indices in 171 children, derived from duplicate measurements on different occasions*

Spirometric index	*SD of single measurement**		*% between occasion variance due to "measurement error"*
	SD	*CV (%)*	
FVC (ml)	121·67	7·5	45
FEV_1 (ml)	117·08	8·3	27
$FEV_{0·5}$ (ml)	109·27	11·2	48
$FEV_{0·5-1}$ (ml)	61·53	14·3	78
PEF (l/min)	22·56	12·1	34
FEF_{25-75} (ml/s)	257·90	14·8	50
FEF_{75-85} (ml/s)	191·01	24·1	56
FEF_{25} (ml/s)	378·02	14·2	40
FEF_{50} (ml/s)	273·60	14·5	58
FEF_{75} (ml/s)	197·22	20·4	44

*Arithmetic scale.
Abbreviations as in table 1.

VARIABILITY AT DIFFERENT LEVELS OF FEV_1

In the lowest tertile (FEV_1 less than 1313 ml) the standard deviation of a single measurement on a given occasion was 64·5 ml and the standard deviation between occasions was 154·1 ml. The corresponding figures for the middle tertile were 61·0 ml and 90·2 ml and for the highest tertile (FEV_1 greater than 1412 ml) 54·7 ml and 92·9 ml. There was a weak but highly significant negative correlation between the mean and the standard deviation of each pair of readings ($r = -0·17$ for 232 pairs on the same occasion, $p = 0·01$; $r = -0·21$ for 171 pairs on different occasions, $p = 0·005$).

Discussion

Few publications present data relating to within subject variability of ventilatory function in children. In their extensive review Polgar and Promadhat[4] quote only one early study, which used reverse plethysmography.[5] Within occasion variability of vital capacity and FEV_1/VC ratio were assessed in 55 children aged 6–14 years, 38 of whom had asthma. The within subject standard deviation of VC (78 ml) was similar to that obtained for FVC in the present study (81 ml) (table 1). In more recent publications the emphasis has been on between occasion variability in older children. Leeder *et al*[6] measured FVC, FEV_1 $FEV_{0·5}$, PEF, and flow rates at 50% and 75% of forced expired vital capacity by pneumotachograph weekly for six weeks in 19 girls of mean age 15·8 years. They quote standard deviations for readings in the same subject on different occasions of 166 ml for FVC and 155 ml for FEV_1, and comment that the ratio of within subject variation to between subject variation was greater for flow rates than for lung volumes. Hutchison *et al*[7] performed repeated lung function tests on 20 healthy children (11 male) aged 10–16 years, using spirometry to determine lung volumes and body plethysmography to determine flow rates. They could not detect any significant effect of time of day or of the retest interval up to two months. Pooled within subject standard deviations can be derived from their data: 90 ml for FVC and 112 ml for FEV_1. Again, variability was significantly greater for flow rates than for lung volumes. In contrast, Cotes *et al*,[8] in a study of 13 twins aged 8–16 years, found a lower day to day coefficient of variation for PEF as measured by a Wright meter (2%) than for $FEV_{0·75}$, FEV_1, or FVC assessed by dry spirometry (4%).

In view of the much younger children studied here, it is surprising how close the within subject variability of FEV_1 and FVC, both within occasions and between occasions, is to previously published figures. Strictly, the conclusions relate to the method used, and this may be particularly relevant to measurements of PEF, which differ consistently between the Wright peak flow meter and pneumotachograph recordings.[4] Nevertheless, when considered in absolute terms, it appears that the between occasion standard deviation for FEV_1 and FVC may be substantially independent of age, and of

Table 4 *Distribution and reliability of spirometric indices among 232 children and the proportion of measured variance attributable to various factors*

Spirometric index	*Mean*	*SD*	*Reliability coefficient*	*Percentage of measured variance explained by*			
				within subject variability	*height*	*sex*	*other*
FVC (ml)	1632	262	0·78	22	37	9	32
FEV_1 (ml)	1407	237	0·76	24	31	4	41
$FEV_{0·5}$ (ml)	978	204	0·71	29	21	2	48
$FEV_{0·5-1}$ (ml)	430	106	0·66	34	13	5	48
PEF (l/min)	186	41	0·70	30	23	1	46
FEF_{25-75} (ml/s)	1739	448	0·67	33	7	0	60
FEF_{75-85} (ml/s)	792	281	0·51	49	2	0	49
FEF_{25} (ml/s)	2669	635	0·65	35	15	0	50
FEF_{50} (ml/s)	1887	480	0·68	32	8	1	59
FEF_{75} (ml/s)	965	317	0·61	39	4	0	57

Abbreviations as in table 1.

the order of 100–150 ml for each index. This would be consistent with the observations that in older children the standard deviation of FEV_1 was independent of the actual volume expired in one second,[7] and implies a smaller coefficient of variation with increasing lung volume. Among the 7 year olds in the present study there was a weak inverse relation between within subject, between occasion standard deviation and mean level of FEV_1. This may reflect the fact that the smaller, younger children were close to the age threshold at which reproducible spirometry is feasible. Alternatively, children with reactive airways may have had spontaneous bronchospasm at one attendance, which both lowered their mean achieved FEV_1 and increased the variation between the two recordings.

For epidemiological studies that concentrate on baseline spirometry between occasion variability is of greatest relevance. Within occasion variability was used as an indicator of "measurement error," though calibration error is a potential source of between occasion variability, which may have persisted despite attempts to monitor calibration throughout the study. The automatic correction to BTPS took into account changes in ambient temperature, but not changes in barometric pressure. Nevertheless, for most indices, and particularly for FEV_1, true between occasion variation appears to be of considerable importance. This presumably reflects the degree of biological variability in airflow among children in this age group.

Statements about the proportion of variance of spirometric indices that is explained or unexplained by putative causes may be misleading if they do not take both measurement error and biological variation into account. Table 4 emphasises that within subject (between occasion) variability, height, and sex account for about half of the between subject variation in most spirometric indices at 7 years of age. The corollary of this is that the amount of "unexplained" variation is substantially smaller than estimated from typical epidemiological data, but the proportionate contribution of factors that do explain some of the true between subject variation is correspondingly greater.

The estimates of within occasion variability were of particular interest in the main study, where one of the principal outcomes was a short term change in ventilatory function after exercise.[1] In the evaluation of any diagnostic or screening test high repeatability is a necessary but not sufficient condition for high validity. Forced vital capacity may be one of the more repeatable lung function indices, but it lacks validity as a measure of airway calibre. Both PEF and FEV_1 have been widely used to measure bronchoconstriction in physiological and pharmacological challenge tests. It has been suggested that in children given a graded histamine challenge the proportionate changes in each index are about equal, and that PEF may be used interchangeably with FEV_1.[9] Table 2 shows that, for any given criterion of abnormality, a test based on change in FEV_1 will be considerably more specific than the equivalent test based on PEF. The calculations in table 2 assume no true change in ventilatory function. In practice, the specificity of tests for bronchial reactivity will also depend on the effect of the challenge on the chosen index of function in normal airways.

The choice of a 20% reduction in FEV_1 as the conventional criterion of abnormality in bronchial challenge procedures[10] appears to be justified, even in this young age group. On purely statistical grounds, there must be considerable reservations about the predictive value of lesser degrees of bronchoconstriction, particularly when these are based on PEF.[11 12] Although mid expiratory flow rates may be more sensitive indicators of exercise induced bronchoconstriction,[13] their diagnostic value is limited by their greater within subject variability. The greater precision of measurement for FEV_1 may be essential if ventilatory function is to be measured repeatedly, after graded doses of a pharmacological challenge or to assess individual response to bronchodilator treatment.

Field work was supported by the Asthma Research Council and was carried out while I held a Wellcome research training fellowship in clinical epidemiology.

References

1 Strachan DP. Damp housing and childhood asthma; validation of symptom reporting. *Br Med J* 1988; **297**:1223–6.
2 American Thoracic Society Statement. Snowbird workshop on the standardization of spirometry. *Am Rev Respir Dis* 1979;**119**:831–8.
3 Cochran WG. Errors of measurement in statistics. *Technometrics* 1968;**10**:637–66.
4 Polgar G, Promadhat V. *Pulmonary function testing in children: techniques and standards*. Philadelphia: Saunders, 1971.
5 Engstrom I, Escardo FE, Karlberg P, Kraepelien S. Respiratory studies in children. VI: Timed vital capacity in healthy children and in symptom-free asthmatic children. *Acta Paediatr Scand* 1959; **48**:114–20.
6 Leeder SR, Swan AV, Peat JK, Woolcock AJ, Blackburn CRB. Maximum expiratory flow-volume curves in children: changes with growth and individual variability. *Bull Eur Physiopathol Respir* 1977;**13**: 249–60.
7 Hutchison AA, Erben A, McLennan LA, Landau LI, Phelan PD. Intrasubject variability of pulmonary function testing in healthy children. *Thorax* 1981; **36**:370–7.

8 Cotes JE, Dabbs JM, Hall AM, Axford AT, Laurence KM. Lung volumes, ventilatory capacity and transfer factor in healthy British boy and girl twins. *Thorax* 1973;**28**:709–15.

9 Henry RL, Mellis CM, South RT, Simpson SJ. Comparison of peak expiratory flow rate and forced expiratory volume in one second in histamine challenge studies in children. *Br J Dis Chest* 1982;**76**:167–70.

10 Chai H, Farr FS, Froelich LA, *et al.* Standardisation of bronchial challenge procedures. *J Allergy Clin Immunol* 1975;**56**:327.

11 Silverman M, Anderson SD. Standardisation of exercise tests in asthmatic children. *Arch Dis Child* 1972; **47**:882–9.

12 Toop L. Active approach to identifying asthma in general practice. *Br Med J* 1985;**290**:1629–31.

13 Bierman CW, Pierson WE. Exercise and asthma: summary. *Pediatrics* 1975;**56**(suppl):950–2.