



New concepts for expressing forced expiratory volume in 1 s arising from survival analysis

M.R. Miller* and O.F. Pedersen#

ABSTRACT: Spirometric lung function is partly determined by sex, age and height (Ht). Commonly, lung function is expressed as a percentage of the predicted value (PP) in order to account for these effects.

Since the PP method retains sex, age and Ht bias, forced expiratory volume in 1 s (FEV₁) standardised by powers of Ht and by a new sex-specific lower limit (FEV₁ quotient (FEV₁Q)) were investigated to determine which method best predicted all-cause mortality in >26,967 patients and normal subjects.

On multivariate analysis, FEV₁Q was the best predictor, with a hazard ratio for the worst decile of 6.9 compared to 4.1 for FEV₁PP. On univariate analysis, the hazard ratios were 18.8 compared to 6.1, respectively; FEV₁·Ht⁻³ was the next-best predictor of survival. Median survival was calculated for simple cut-off values of FEV₁Q and FEV₁·Ht⁻³. These survival curves were accurately fitted ($r^2=1.0$) by both FEV₁Q and FEV₁·Ht⁻³ values expressed polynomially, and so an individual's test result could be used to estimate survival (with SD for median survival of 0.22 and 0.61 yrs, respectively).

It is concluded that lung function impairment should be expressed in a new way, here termed the FEV₁Q, or, alternatively, as FEV₁·Ht⁻³, since these indices best relate spirometric lung function to all-cause mortality and survival.

KEYWORDS: Cox's proportional hazards, respiratory function tests, spirometry

From the first scientific recording of lung function data [1, 2], it was appreciated that the values obtained were dependent upon the subject's sex, age and height (Ht). This led to the practice of trying to take these influences into account by using prediction equations and then relating the subject's result to the expected value (percentage of the predicted value (PP)). In the second edition of *Respiratory Function in Disease* by BATES *et al.* [3], it was suggested that, if a lung function index value were <80% pred, then it was likely to be abnormal. This method was widely embraced [4] and has endured, but there are several reasons why this is not a helpful rule of thumb [5, 6]. The European Respiratory Society (ERS) was the first to recommend the use of standardised residuals (SRs), which are in essence a z-score, for determining whether or not an index is outside the normal range [7]. The lower limit of normal (LLN) is -1.645 SRs, which is an estimate of the lower 5th percentile, and this method has been recommended by the most recent American Thoracic Society (ATS) and ERS statement [8] for determining whether or not a

result is abnormal, with PP suggested as the method for expressing severity.

Whenever lung function is related to a predicted value, it requires accurate prediction equations against which to compare the subject's data. The equation must be obtained from a relevant population of subjects, using comparable equipment and with rigorous technical standards applied. Exactly how normal these subjects are can be hard to define, and a population that is too pure may be unrepresentative. Even the best prediction equation has quite wide 95% confidence limits for its predicted value, and so this, in itself, is an inexact science. An alternative approach has been to standardise spirometric lung function using a power relationship with Ht that helps to account for some size and sex difference. This method was shown, in the Framingham study [9], to be helpful in relating lung function data to subsequent survival, and was found by FLETCHER *et al.* [10] to be the best method for evaluating longitudinal decline in function in chronic obstructive pulmonary disease (COPD).

AFFILIATIONS

*Dept of Medicine, University Hospitals Birmingham National Health Service Trust, Birmingham, UK.

#Institute of Public Health, Dept of Environmental and Occupational Medicine, University of Aarhus, Aarhus, Denmark.

CORRESPONDENCE

M.R. Miller
Dept of Medicine
University of Birmingham
Selly Oak Hospital
Birmingham
B29 6JD
UK
E-mail: martin.miller@uhb.nhs.uk

Received:

Feb 14 2009

Accepted after revision:

Aug 27 2009

First published online:

Sept 09 2009

The purpose of the present article is to explore the limitations and advantages of these various methods, and to explore other and perhaps better methodologies so that these can be tested by other researchers in order to help determine the best way forward. The relationship between lung function and all-cause mortality was used to explore this since this was information that was readily available. It has previously been shown that forced expiratory volume in 1 s (FEV₁) is a predictor of all-cause mortality in the general population [9, 11–13], and that this relationship is even stronger for mortality caused by respiratory diseases related to airflow obstruction [11].

DATA AND METHODS

Data sets

In order to help explore various methods for using spirometric lung function data, three sets of data were used. One was obtained from routine lung function tests performed at the University Hospitals Birmingham National Health Service Trust (Birmingham, UK). These data comprise the results of the most recent attendance tests obtained from 11,972 patients (53% male) referred for whatever reason for lung function tests, and all had their survival registered up to October 2008 in UK National Health Service data records. Tests were performed on equipment validated to conform to ATS/ERS specifications [14], and the data were obtained following these test criteria. The second set of data were from the Copenhagen City Heart Study (CCHS) [15], kindly released to us by P. Lange (Hvidovre Hospital, Hvidovre, Denmark) in order to facilitate exploration of a novel approach to using lung function data [11]. The lung function

data from subjects entered into the CCHS during the period 1976–1978 and their survival up to December 2002 were released for analysis. The methods and background to this large project have been described previously [12, 15]. FEV₁ and forced vital capacity (FVC) were recorded without prior bronchodilatation whilst sitting using a Vitalograph bellows spirometer (Maids Moreton, UK). Only subjects with at least two measurements within 5% of each other were included, and the highest value obtained was recorded. Any subject whose recorded FEV₁ was <0.3 L (n=4) or whose FEV₁ exceeded FVC (n=19) were excluded, leaving data from 13,900 subjects (46% male) for analysis. The third set of data comprised 1,095 patients (41% male) with COPD who had had their post-bronchodilator FEV₁ recorded and were then followed for 15 yrs in order to explore predictors of survival [16]. When all three data sets were combined and only those aged ≥20 yrs retained, this left 26,967 subjects for analysis with regard to FEV₁ and survival prediction.

Methods for expressing FEV₁ impairment

FEV₁ has been expressed in a number of ways, as PP (FEV₁PP), using European Coal and Steel Community (ECSC) reference equations, and as FEV₁ divided by Ht squared (FEV₁·Ht⁻²) [13, 17] and Ht cubed (FEV₁·Ht⁻³) [10]. FEV₁ has also been presented as a SR (FEV₁SR), which is derived from:

$$\text{FEV}_{1\text{SR}} = (\text{observed FEV}_{1\text{SR}} - \text{predicted FEV}_{1\text{SR}}) / \text{RSD}$$

where RSD is the residual standard deviation of the prediction equation used [7]. Although the ECSC reference equations are

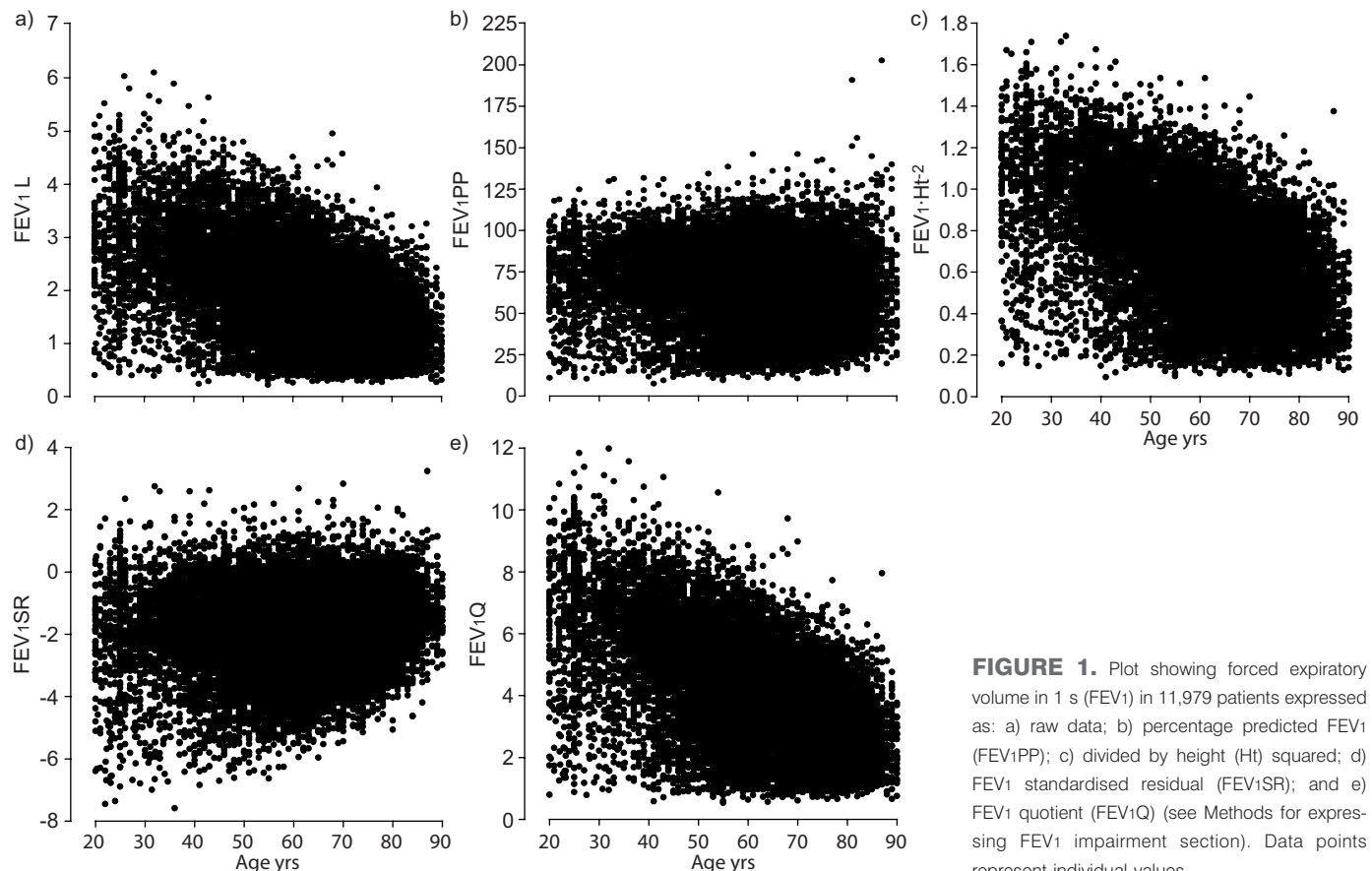


FIGURE 1. Plot showing forced expiratory volume in 1 s (FEV₁) in 11,979 patients expressed as: a) raw data; b) percentage predicted FEV₁ (FEV₁PP); c) divided by height (Ht) squared; d) FEV₁ standardised residual (FEV₁SR); and e) FEV₁ quotient (FEV₁Q) (see Methods for expressing FEV₁ impairment section). Data points represent individual values.

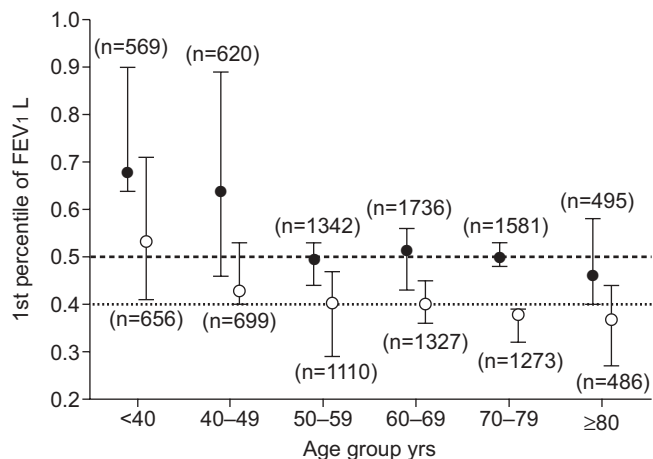


FIGURE 2. First percentiles of forced expiratory volume in 1 s (FEV₁) in male (●) and female patients (○) by age group. -----: proposed generic male first percentile;: proposed generic female first percentile. Vertical bars represent 95% confidence limits and numbers of subjects are shown.

only relevant for people aged ≤ 70 yrs, they are frequently used beyond this age, and we have found the equations to be as good at prediction as other more age-specific equations up to an age of 95 yrs [18]. Plotting the patient data showed that, regardless of age, there was a flat lower limit to FEV₁, as shown in figure 1. It was found that the lower 1st percentile of FEV₁ in the patient group of nearly 12,000 subjects differed between the sexes (0.5 L for males and 0.4 L for females), but did not vary significantly with age at ages >50 yrs, where more reliable estimates of the 1st percentile were possible, as shown in figure 2. It was then decided to standardise FEV₁ using these sex-specific lowest 1st percentiles, and this index was termed the FEV₁ quotient (FEV₁Q). It is an index of the number of turnovers of a nominal lower limit of lung function remaining, and takes into account some sex and size differences in lung function. When using data from a single sex, FEV₁Q has no advantage over raw FEV₁.

Statistics

All analysis was undertaken using Stata/SE version 9.1 (StataCorp, College Station, TX, USA). Cox's regression models for predicting survival from FEV₁ were derived together with age and sex as predictors, and then without these predictors, since the object was to determine the best method for using lung function data in a clinical setting in which other factors would not be explicitly accounted for in decision making. All models were confirmed to abide by the assumptions implicit in proportional hazard analysis.

RESULTS

Each method of expressing lung function impairment is taken in turn, with results provided to support or reject its use in this context. Finally, results on survival in the present large data set are explored.

Percentage of the predicted value

PP methodology has sustained itself over the years, but it has no statistical basis and can be misleading when comparing different lung function indices. Figure 3 shows idealised data for males and females indicating that the true LLN is at different PPs for different ages, and that this differs between the sexes. It also varies with Ht. Thus, if the procedure of relating to a predicted value is an attempt to account for age, Ht and sex, this method indeed conceals influences for each of these three domains that potentially corrupt the result. The problems with PP get worse if it is desired to compare results from different indices because the PP that relates to the estimated 5th percentile is very different according to the index considered. Table 1 shows the LLN (estimated 5th percentile) expressed as PP for both sexes for a variety of indices using the ECSC equations [7]. The values for LLN range from 58% pred for residual volume to 87% pred for FEV₁ as a percentage of FVC. If the PP 80% rule [3] were used, it might be falsely assumed that the result for residual volume was extremely low but that the FEV₁ as a percentage of FVC was acceptable, whereas they are indeed both equivalent and at the LLN. This table indicates how difficult and potentially misleading it is to

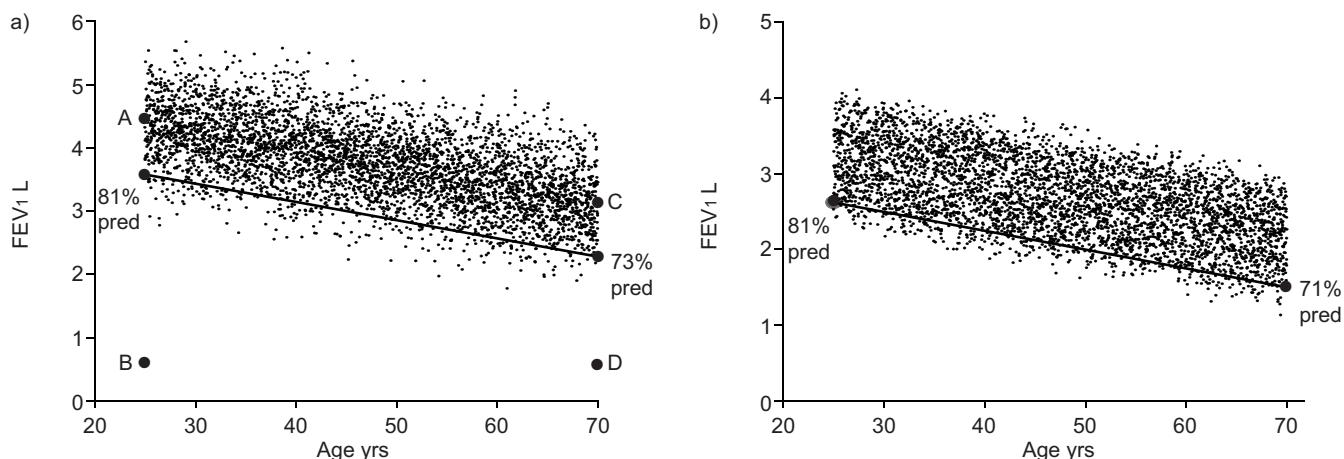


FIGURE 3. Hypothetical populations of: a) 1,000 males; and b) 1,000 females with normally distributed forced expiratory volume in 1 s (FEV₁) data that fit the European Coal and Steel Community prediction equations with the estimated 5th percentile shown (sloping line; values shown are for ages of 25 and 70 yrs). Points B and D represent males of height 1.77 m and aged 25 and 70 yrs, respectively, each with an FEV₁ of 0.6 L, and points A and C are the predicted values of B and D respectively. For further explanation, see Standardised residuals section.

TABLE 1 Lower limit of normal of various lung function indices expressed as a percentage of the predicted value at two extremes of age[#]

	Males		Females	
	25 yrs	70 yrs	25 yrs	70 yrs
FEV ₁	81	73	81	71
FVC	81	75	81	72
FEV ₁ (% FVC)	86	84	87	86
PEF	80	75	79	74
FRC	70	73	69	70
RV	59	74	58	72
TLC	84	84	80	80
RV (% TLC)	62	78	65	78
T _{L,CO}	81	74	80	73

FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; PEF: peak expiratory flow; FRC: functional residual capacity; RV: residual volume; TLC: total lung capacity; T_{L,CO}: transfer factor of the lung for carbon monoxide. [#]: for males and females of average height (1.74 and 1.66 m, respectively).

use PP to look for patterns of abnormality amongst lung function indices.

If PP were a valid method of expressing severity of impairment, then it might be expected that the lowest FEV₁PP seen in patients would be roughly the same irrespective of sex, age and Ht. Looking at this another way, if FEV₁PP were a valid measure of severity, then young patients with cystic fibrosis would die with a larger raw FEV₁ than older people because their predicted value is larger than that found in older subjects. Figure 1 shows the FEV₁ data from all of the present patients plotted against age expressed as raw FEV₁, FEV₁PP, FEV₁·Ht⁻², FEV₁SR and FEV₁Q. For raw FEV₁, it is striking that the lower boundary was roughly the same irrespective of age. For FEV₁PP, the lower limit in the young subjects was lower than that seen in older subjects, *i.e.* young subjects can survive with an FEV₁ that is a much lower PP than can older subjects. For FEV₁·Ht⁻² and FEV₁Q, the lower boundary is flat, much as for raw FEV₁. Figure 1 confirms what is known from clinical practice, *i.e.* that young cystic fibrosis patients can survive with an absolute FEV₁ just as low as can 70 yr olds, and so can survive with a much lower PP [19]. This suggests that FEV₁PP is also not the best method for estimating severity.

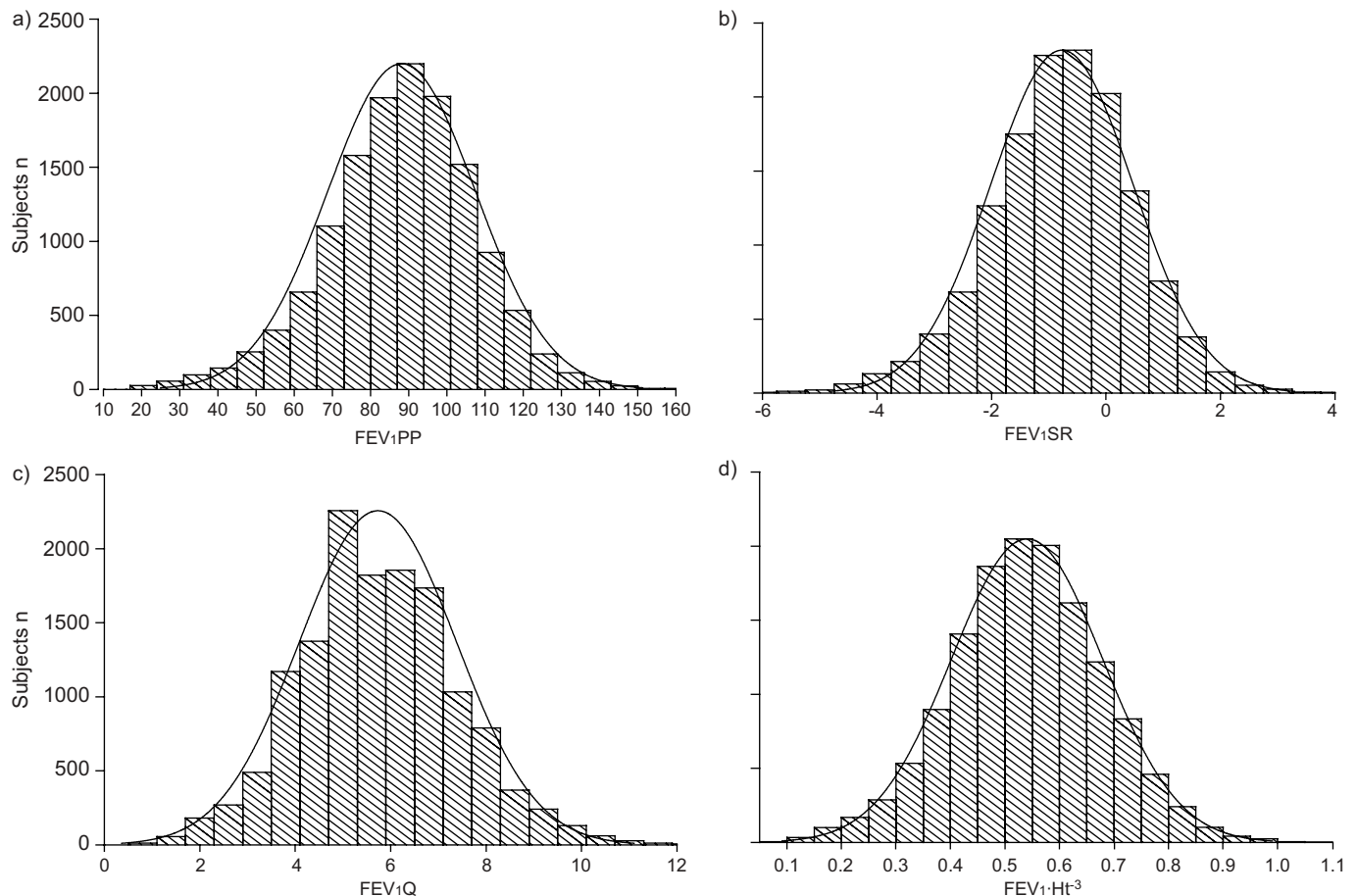


FIGURE 4. Histograms, with a normal distribution curve superimposed, showing Copenhagen City Heart Study forced expiratory volume in 1 s (FEV₁) data expressed as: a) percentage predicted FEV₁ (FEV₁PP); b) FEV₁ standardised residual (FEV₁SR); c) FEV₁ quotient (FEV₁Q); and d) divided by height (Ht) cubed.

TABLE 2 Age, survival and forced expiratory volume in 1 s expressed as a standardised residual (FEV₁SR) in each of the three components of the combined data set

	Subjects	Males	Age yrs	Survival yrs	FEV ₁ SR
CCHS	13900	45.9	53.0±12.0	19.7±7.6 (26.8)	-0.79±1.25
COPD	1095	41.2	61.8±9.9	7.9±3.7 (14.6)	-3.21±1.02
Patients	11972	53.3	60.7±15.1	3.1±2.3 (20.9)	-1.60±1.48
Total	26967	49.0	56.8±13.9	11.8±10.0 (26.8)	-1.25±1.46

Data are presented as n, % or mean±sd, with maximum in parentheses for survival. CCHS: Copenhagen City Heart Study; COPD: chronic obstructive pulmonary disease.

Standardised residuals

Use of SRs is the method endorsed by the ATS and ERS in their recommendations for determining whether or not an individual's lung function is outside the normal range [7, 8]. The SR is commonly used in statistical analysis, with the term being synonymous with a z-score, and was first used in the context of lung function data in a study looking at patterns of abnormality in smokers [20]. The advantage of this technique is that the units are the same for all types of index, and the SR indicates where a subject's result lies with regard to the Gaussian distribution of the normal population. Since 1.645 SRs below the predicted value is an estimate of the lower 5th percentile (1.96 SRs below estimates the 2.5th percentile), a level of deviation from predicted where clinical interest is to be directed can be decided upon. The ATS and ERS [8] have suggested that 1.645 SRs below the predicted value is the level to use in patients to define the LLN. Implicit in this is that 5% of people who have been judged to be completely normal would now be considered as abnormal (*i.e.* they are false positives). For patients or symptomatic subjects, this may be acceptable, but, if an asymptomatic population of nondiseased subjects were being tested, the estimated 2.5th percentile might be chosen instead in order to minimise the number of false positive results.

Using SRs to express the degree of abnormality below the LLN is more problematic since the predicted values for younger

TABLE 3 Forced expiratory volume in 1 s standardised residual (FEV₁SR) and survival in each of seven age ranges

Age range yrs	Subjects n	Dead	Males	FEV ₁ SR	Survival yrs
20 –<30	1126	10.0	45.8	-1.3±1.8	15.1±11.0
30 –<40	2308	13.7	48.1	-0.9±1.5	17.3±10.5
40 –<50	4496	27.2	45.9	-1.0±1.4	16.8±10.1
50 –<60	7459	44.6	46.0	-1.2±1.5	14.0±9.8
60 –<70	6639	56.8	52.0	-1.4±1.5	9.3±8.3
70 –<80	3762	51.0	53.9	-1.5±1.3	4.6±4.9
80 – 97	1177	59.6	51.6	-1.2±1.2	3.0±3.1

Data are presented as % or mean±sd, unless otherwise indicated.

subjects are higher and thus, in terms of the number of RSDs available to fall, the younger are able to go lower. This can be seen in figure 3a for males, where point A is the predicted value for a male aged 25 yrs of Ht 1.77 m and point B is for a male of the same age and Ht with an FEV₁ of 0.6 L. Points C and D are the equivalent for a male aged 70 yrs. The baseline of zero FEV₁ is 8.6 and 6.1 SRs below the predicted values for these males aged 25 and 70 yrs, respectively. The FEV₁ at B represents 7.44 SRs below the predicted value and point D represents an FEV₁ of 4.9 SRs below the predicted value. It is not possible for the male aged 70 yrs to have an FEV₁ of 7.44 SRs below the predicted value since this would require a negative FEV₁, which is nonsensical. If the two subjects at B and D were indeed equivalently disabled, showed equivalent symptoms and had similar survival projections, then the SR method would not appear to reflect properly the degree of impairment. This is borne out in figure 1, where the SRs go much lower in the younger subjects than in the older subjects.

Standardising by powers of height

The Framingham study [9] showed that FEV₁ divided by Ht gave a reasonable prediction of long-term survival. FLETCHER *et al.* [10] showed that FEV₁·Ht⁻³ as a means of standardising FEV₁ was the best method for evaluating lung function decline. This form of standardisation by Ht takes some sex and size differences into account, and it is these differences that make use of raw FEV₁ problematic, especially when considering data from both sexes together.

Regression of log FEV₁ against log Ht in the CCHS data gives a slope of 3.7, but only 0.33 of the variance in ln FEV₁ was explained by ln Ht. The fit was not very good and this slope for both sexes suggested Ht to the power of three or four might provide the best fit. Figure 4 shows histograms of CCHS FEV₁ data expressed as FEV₁PP, FEV₁SR, FEV₁Q and FEV₁·Ht⁻³. Since these data were randomly acquired from a normal

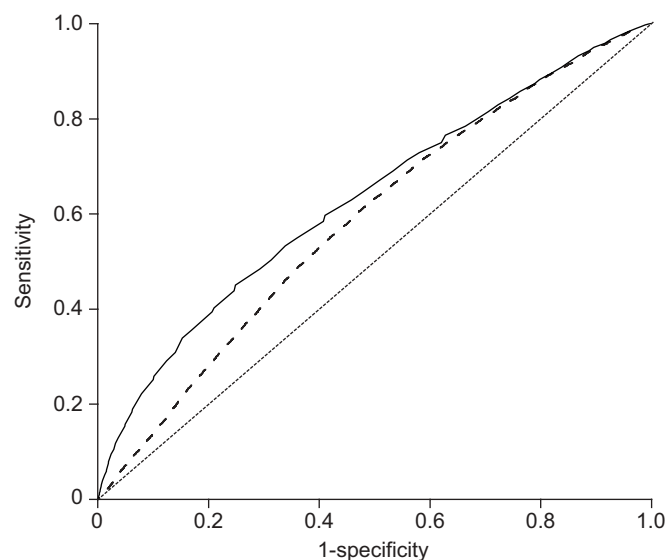


FIGURE 5. Receiver operating characteristic curves for forced expiratory volume in 1 s (FEV₁) expressed as FEV₁ quotient (FEV₁Q; —) and percentage predicted FEV₁ (FEV₁PP; - - -) for predicting survival in 26,967 subjects. ·····: reference line of no effect.

TABLE 4 Hazard ratios for prediction of all-cause mortality using Cox's regression analysis[#]

	FEV ₁ Q	FEV ₁ ·Ht ⁻²	FEV ₁ ·Ht ⁻³	FEV ₁ PP
Chi-squared	10408	10376	10320	10257
Sex				
Females	1	1	1	1
Males	1.8	2.0	1.8	1.5
Age decile				
1 (youngest)	1	1	1	1
2	1.6	1.7	1.7	1.8
3	2.2	2.3	2.3	2.7
4	2.9	3.1	3.1	3.9
5	3.5	3.8	3.9	5.0
6	4.2	4.5	4.7	6.3
7	5.3	5.8	6.1	8.6
8	6.6	7.3	7.7	11.5
9	8.5	9.4	10.0	15.8
10 (oldest)	13.0	14.6	15.6	27.3
Lung function decile				
1 (best)	1	1	1	1
2	1.4	1.4	1.3	1.1
3	1.6	1.6	1.6	1.2
4	2.0	1.7	1.6	1.2
5	2.0	2.0	1.9	1.4
6	2.6	2.3	2.1	1.6
7	3.0	2.7	2.5	1.8
8	3.7	3.4	3.2	2.4
9	4.8	4.4	4.2	2.9
10 (worst)	6.9	6.5	5.9	4.1

The youngest, best lung function and female sex groups were used as the reference groups. Chi-squared for each model are shown as a measure of fit, and, although the differences in Chi-squared appear small, the models in the leftmost columns are significantly better than those in the rightmost columns ($p < 0.05$). FEV₁Q: forced expiratory volume in 1 s (FEV₁) quotient; FEV₁·Ht⁻²: FEV₁ divided by height squared; FEV₁·Ht⁻³: FEV₁ divided by height cubed; FEV₁PP: percentage predicted FEV₁. #: in 26,967 subjects, using sex, age and lung function indices as predictors, with the latter two split into deciles.

population with respect to their lung function, the distribution for a satisfactory method of expressing lung function for both sexes together should be normal. For raw data, the histogram would indeed show two separate distributions for males and females, with their known size differences (skewness of 0.57). When expressed as PP the distribution was negatively skewed (-0.27), and the same was true for SRs (-0.26). FEV₁Q had a skewness of 0.23, but FEV₁·Ht⁻² gave a better fit for a normal distribution, with skewness of 0.15, and the fit was best for FEV₁·Ht⁻³, with skewness of 0.00.

Testing lung function impairment and survival

The hardest end-point for lung function impairment to be tested against is survival, and this is a clearly defined end-point. Table 2 shows the mean ages and mean survival for each component of this large data set, and table 3 shows the number of subjects, split into 10-yr age bands, with their mean \pm SD FEV₁SR and survival, and the percentage of subjects who had

died. Receiver operating characteristic (ROC) curves were calculated to investigate which method of expressing FEV₁ was best, on its own, at predicting survival, and the area under the curve was best for FEV₁Q (0.631 (95% confidence limit 0.624–0.637)), with FEV₁·Ht⁻³ almost the same (0.626 (0.619–0.633)); next best was FEV₁·Ht⁻² (0.621 (0.614–0.628)), followed by raw FEV₁ (0.606 (0.599–0.612)) and FEV₁PP (0.586 (0.579–0.592)), with FEV₁SR being worst at 0.571. Figure 5 shows the ROC curves for FEV₁Q and FEV₁PP, with FEV₁Q being more specific and no less sensitive than FEV₁PP.

The best FEV₁ predictor of survival on multivariate analysis was determined from Cox's regression models, which were derived using each index, sex and age as predictors, with lung function in deciles. The best model for predicting survival was with FEV₁Q, followed by FEV₁·Ht⁻², FEV₁·Ht⁻³ and then FEV₁PP, with each model significantly better than the next ($p < 0.05$ (likelihood-ratio test)), with the hazard ratios for the results shown in table 4. The FEV₁Q column in table 4 shows that the hazard ratios for older age groups were smaller for FEV₁Q than for FEV₁PP, but that the opposite was true for the hazard ratios associated with worsening lung function. This indicates that age *per se* plays a smaller part in prediction of survival for the model using FEV₁Q than for the model with FEV₁PP. The data were then split into data set A, comprising 12,181 subjects with an FEV₁SR ranging from 0.0 to -1.645 (mean \pm SD -0.82 \pm 0.46), and data set B with 9,630 subjects with an FEV₁SR of $<$ -1.645 (mean \pm SD -2.81 \pm 0.95), *i.e.* all were below the LLN. Cox's models were generated for each of these two data sets using sex and quintiles of both function and age to determine whether or not the ability of the various FEV₁ indices to predict survival was different in those with better (data set A) or worse (data set B) lung function. The best model for predicting survival in both data sets was with FEV₁Q, followed, in order, by FEV₁·Ht⁻³, FEV₁·Ht⁻² and then FEV₁PP. For each data set, the FEV₁Q model was significantly better than the other three models, and the FEV₁PP model was significantly worse than the others ($p < 0.001$ (likelihood-ratio test)). Models with FEV₁SR were very much worse with little utility. Thus the superiority of FEV₁Q in predicting survival was not affected by the range of lung function being considered.

Each method of expressing lung function was then split into the top quartile, as the reference group for normal survival, and then the remaining values for the index were divided into a further nine bins of subjects using cut-off levels derived as follows. If X_q were the value defining the upper quartile, the other bins were defined at X_q , $X_q \times 9/10$, $X_q \times 8/10$. . . $X_q \times 1/10$. The worst two groups were combined since the number of subjects was $<$ 10 in the lowest groups. Regression models using only these bins as predictors were derived without sex or age as predictors, with the results shown in table 5. Again, FEV₁Q was the best predictor, with FEV₁·Ht⁻³ being the next best. Lastly, Cox's regression models were derived for simple numerical cut-off levels of FEV₁Q and FEV₁·Ht⁻³ that might easily be applied in lung function laboratories ($<$ 1.0, 1.0–1.9, 2.0–2.9 . . . 6.0–6.9 and \geq 7.0 for FEV₁Q, with the cut-off levels for FEV₁·Ht⁻³ being numerically a tenth of these). Median survival was calculated for each group, with the results shown in figure 6. The survival curves in figure 6 predicted from FEV₁Q and FEV₁·Ht⁻³ could each be accurately fitted by polynomial functions, as presented in table 6.

TABLE 5 Hazard ratios for prediction of all-cause mortality using Cox's regression analysis[#]

	FEV1Q	FEV1·Ht ⁻³	FEV1·Ht ⁻²	FEV1PP
Chi-squared	6405	5603	5343	3490
Lung function				
1 (best)	1.00	1.00	1.00	1.00
2	1.67 (1.49–1.87)	1.50 (1.34–1.67)	1.57 (1.41–1.74)	0.97 (0.88–1.06)
3	2.27 (2.05–2.51)	2.20 (1.98–2.44)	2.00 (1.81–2.22)	1.11 (1.01–1.21)
4	3.07 (2.77–3.41)	2.40 (2.17–2.66)	2.33 (2.11–2.58)	1.16 (1.06–1.27)
5	3.66 (3.31–4.05)	2.98 (2.70–3.30)	2.80 (2.53–3.09)	1.33 (1.22–1.46)
6	4.80 (4.34–5.32)	3.74 (3.38–4.13)	3.51 (3.18–3.87)	1.61 (1.48–1.76)
7	6.31 (5.70–6.97)	4.68 (4.24–5.17)	4.46 (4.04–4.91)	1.91 (1.76–2.08)
8	7.99 (7.23–8.83)	6.57 (5.95–7.26)	6.01 (5.45–6.62)	2.76 (2.54–3.00)
9	12.29 (11.10–13.60)	9.67 (8.75–10.69)	9.02 (8.17–9.95)	3.92 (3.60–4.27)
10 (worst)	18.76 (16.96–20.76)	14.87 (13.46–16.42)	14.02 (12.72–15.47)	6.11 (5.62–6.65)

Data are presented with the 95% confidence limit in parentheses. The upper quartile of best lung function was used as the reference group. Chi-squared for each model are shown as measure of fit. FEV1Q: forced expiratory volume in 1 s (FEV1) quotient; FEV1·Ht⁻²: FEV1 divided by height squared; FEV1·Ht⁻³: FEV1 divided by height cubed; FEV1PP: percentage predicted FEV1. [#]: in 26,967 subjects, using 10 groupings of lung function index (see Testing lung function impairment and survival section) alone as predictor.

DISCUSSION

The present study has shown that the currently widely used PP method is significantly inferior to other methods for expressing FEV1 when considering the relation between lung function and subsequent survival. In this context, it has been shown that standardisation using a power of Ht is much better than using FEV1PP or FEV1SR. Overall FEV1·Ht⁻³ has a slight edge as the best power of Ht for removing sex and size bias, and, in a random normal population, this measure was normally distributed. However, the best way of expressing impairment was with use of FEV1Q, a novel method we propose for expressing lung function data. Like FEV1PP and FEV1SR,

however, it depends on sex, since the denominator is sex-dependent. Choosing the best method in a given circumstance is dependent upon the aspect of clinical care or management that is relevant. We chose the method that was best for predicting all-cause mortality as this information was readily available to us, and it is a well-defined end-point that is ultimately the most important outcome in any medical condition. The results presented here may not be correct for an alternative end-point, such as symptoms like breathlessness, and this aspect needs testing in other appropriate data sets. Focussing on respiratory mortality alone might further enhance the prediction; it has previously been shown, in a

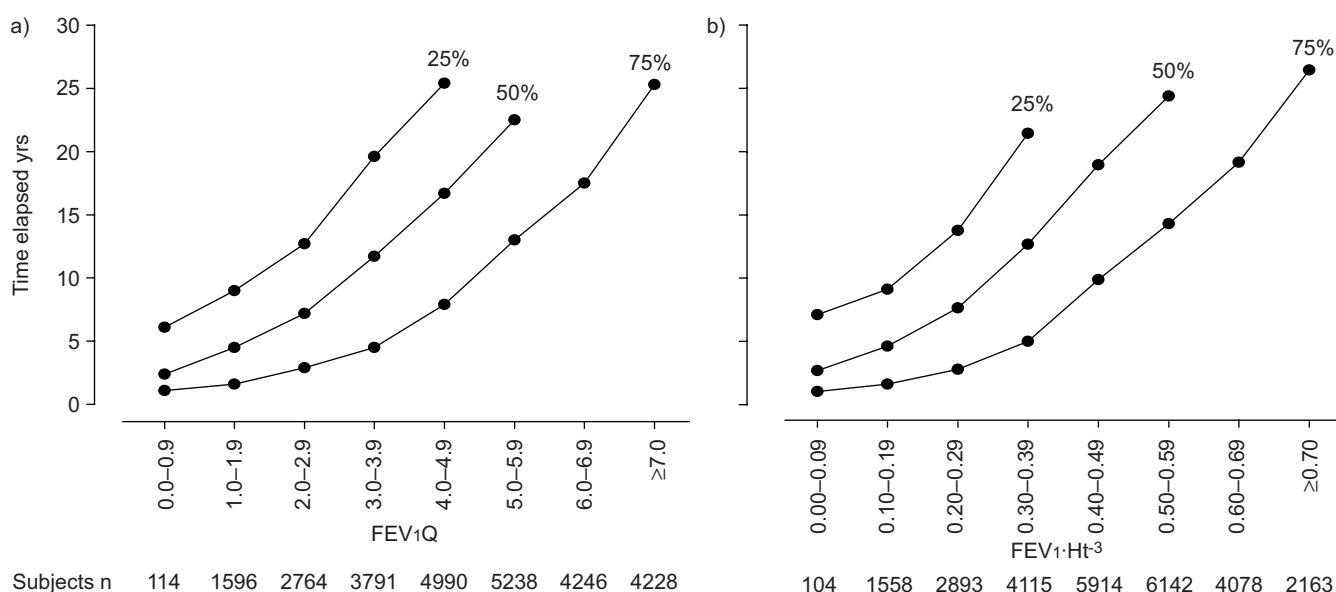


FIGURE 6. Isopleths of the time elapsed with 75% (lower quartile survival), 50% (median) and 25% (upper quartile) of the population still surviving for each of the bins of cut-off levels for: a) forced expiratory volume in 1 s (FEV1) quotient; and b) FEV1 divided by height (Ht) cubed. Numbers of subjects in each group are shown.

TABLE 6 Constants and coefficients for polynomial prediction of the expected lower quartile, median and upper quartile for survival[#] using either forced expiratory volume in 1 s (FEV1) quotient (Q) or FEV1·height (Ht)⁻³

Predictor	Constant	Index	Index ²	Index ³	RSD	r ²	Limits [†]
FEV1Q							
Lower quartile	1.18	-0.24	0.29	0.02	0.47	0.9984	0.5–7.5
Median	1.71	1.08	0.49		0.22	0.9995	0.5–6.0
Upper quartile	4.99	1.71	0.64		0.71	0.9960	0.5–5.0
FEV1·Ht⁻³							
Lower quartile	1.12	-5.38	52.08		0.50	0.9979	0.05–0.75
Median	1.74	11.92	54.63		0.61	0.9968	0.05–0.60
Upper quartile	7.24	-9.07	141.75		0.07	1.0000	0.05–0.40

RSD: residual standard deviation. [#]: in years; [†]: minima and maxima for the index for which the prediction is valid.

general population sample, that the hazard ratio of cut-off levels of FEV1·Ht⁻² for predicting death caused by respiratory disease were 10 times higher than for all-cause mortality in the more severely affected subjects [11].

FEV1 has been found by many authors to relate to survival in the general population [9, 11–13], but the exact reason for this is not clear. It is possible that this link occurs because genes that are associated with worse lung function are in some way co-located with genes that determine susceptibility to common diseases, such as cancer or cardiovascular disease. In support of this, the Framingham study found that FEV1 did not relate to survival in the elderly [9]. However, it has recently been shown that lung function still predicts survival in a cohort of 95 yr olds [18], when a putative link between lung function and other disease risks would have been mainly spent. This suggests that having lower lung function may mean that other diseases are more likely to be fatal, for example by predisposing to pneumonia following a stroke, but a firm causal link has not yet been proven.

The problems with the PP method relate to the proportional assumption implicit in this expression and the numerical aspects that lead to PP retaining unwanted age, sex and Ht bias. A further issue is that all assessments of impairment are currently based on looking at how far a subject has fallen from an estimated predicted point that is deemed appropriate for the subject's sex, age and Ht. This predicted value includes a lot of uncertainty, and so the resulting index includes this uncertainty, and perhaps contributes to why the PP method is not the best index. We chose to turn the issue on its head and concentrate on looking at how far a subject is above the bottom line. Using a zero value as the bottom line is no good since, with data for both sexes, size differences obscure the signal. Using Ht standardisation is effective in this respect and it seems, from the combined data here, that standardisation by Ht cubed is better than using lower powers of Ht.

We here propose a new concept for expressing spirometric data, which was suggested to us from the observation in figure 1 that there is an absolute lower limit of FEV1 seen in

TABLE 7 Example forced expiratory volume in 1 s (FEV1) results in four different subjects

	1	2	3	4
Age yrs	55	65	80	65
Height m	1.56	1.76	1.50	1.76
Sex	F	M	F	M
FEV1 L	3.00	0.67	0.67	1.20
FEV1SR	2.14	-4.95	-1.73	-3.91
FEV1Q	7.50	1.34	1.68	2.40
MS estimated from FEV1Q (X) yrs	33.8	3.0	3.7	5.6
Pred FEV1 L	2.19	3.19	1.33	3.19
Pred FEV1 as FEV1Q	5.47	6.39	3.31	6.39
MS estimated from Pred FEV1Q (Y) yrs	19.4	25.4	8.7	25.4
Survival difference (X-Y) yrs	14.3	-22.4	-5.0	-19.8
Survival ratio (X/Y)	1.74	0.12	0.42	0.22

FEV1SR: FEV1 standardised residual; FEV1Q: FEV1 quotient; MS: median survival; Pred: predicted; F: female; M: male.

laboratory testing. It was then found that this limit is slightly lower in females and, if a subject's FEV1 is standardised by the relevant sex-specific lower limit, this gives the number of turnovers of FEV1 left for the subject. This index, FEV1Q, is the best overall predictor for use with regard to predicting survival. This is true with multivariate (table 4), as well as univariate, analysis (table 5). This represents a change in thinking to concentrate on what function it is known a subject has left to survive on, rather than on what it is thought that they might have lost. In a clinical setting, there is always awareness of the age and sex of the patient and appreciation that survival is related to both of these attributes. However, an index may be of greater utility to a clinician if it does not require any additional manipulation in order to take these into account. With FEV1Q and FEV1·Ht⁻³, this is possible, but, if FEV1PP were to be used, then this index must be manipulated with a complex function in order to take into account the age and sex of the subject before it can accurately be used to predict survival potential. Since an age effect is retained in FEV1Q and FEV1·Ht⁻³, it is possible that these indices might not be so suitable if it were necessary to focus solely on the extent of lung function abnormality *per se* and totally avoid any age effect, or, alternatively, if there were a research need to tease out the exact effect of age, as distinct from lung function, on an aspect of medical interest. Although FEV1PP attempts to account for age effects on lung function by use of prediction equations, this method introduces other age, sex and Ht biases from the equation used, and the assumption of proportionality and these effects introduce noise in the signal and reduce the overall ability of a researcher to predict mortality. In analysis of the effect of time-related exposures on lung function, such as in occupational medicine, there may be an advantage in using FEV1Q or FEV1·Ht⁻³ as it does not hide such biases. It has been shown here that FEV1Q and FEV1·Ht⁻³ are the best indices for investigating all-cause mortality, but it is another question that remains to be tested as to how well they relate to symptoms and other markers of lung disease.

The results in figure 6, based on simple cut-off points of the lung function index alone, can be used clinically to judge the severity of a situation with regard to survival by using the appropriate polynomial prediction from table 6. This represents a potential benefit to patients for initiating treatment strategies and for disease severity stratification in future research into outcomes of lung diseases. Thus, for example, a subject's FEV₁ would first be tested to see whether or not it was outside the expected range (e.g. below the lower 90% confidence limit, i.e. an FEV₁SR of <-1.645), and then the severity of any abnormality could be judged using the estimated median survival for the FEV₁Q. This survival could, if needed, be related to the estimated survival for the predicted FEV₁Q. In relating any other measurements that might arise in research to severity, using the FEV₁Q itself would suffice. Table 7 shows some examples of how lung function results can be expressed using the present results.

The method for standardising FEV₁ using the lowest sex-specific FEV₁ (FEV₁Q) is a way of avoiding the difficulties of using raw FEV₁ because of sex and size differences between individuals. This method expresses an individual's FEV₁ as the number of turnovers of the bottom line level of lung function that remain. It has been found that decline in FEV₁ in normal subjects is greater in males than females, and is defined in terms of absolute loss of volume [21]. For never-smokers aged >50 yrs the annual FEV₁ decline has been estimated to be 28 mL in males and 22 mL in females [22], giving a female to male ratio of 0.78, which is approximately the same as the ratio of a female to male 1st percentile FEV₁ of 0.8. In patients with asthma, FEV₁ decline has been found to be ~50 mL·yr⁻¹ in intrinsic asthmatics and ~23 mL·yr⁻¹ in extrinsic asthmatics [23], and ~32 mL·yr⁻¹ in asthmatic nonsmokers aged 40–60 yrs and ~26 mL·yr⁻¹ in those aged >60 yrs [21]. None of these studies suggested that this loss was a proportional effect. These longitudinal findings can be easily applied to FEV₁Q data in that, for nonsmokers aged 40–60 yrs, a decrease in FEV₁Q of 1.0 would take ~18 yrs [22]. In smokers aged >60 yrs, their accelerated loss would equate to a decrease in FEV₁Q of 1.0 every 10 yrs [21]. These estimates of FEV₁Q decline are independent of sex because the proportional difference in lung function decline between males and females appears to be an approximate fit for the observed sex difference in 1st percentile FEV₁.

As FEV₁ declines, the likelihood of terminal hypoventilatory failure increases, and the fact that females appear to have a smaller absolute lower limit of FEV₁ can be explained in two ways. Ventilatory dead space [24, 25] and airway volume [26], which are closely correlated to anatomical dead space, have been shown to be smaller in females and are related to their overall smaller stature. Secondly, females also exhibit a lower basal metabolic rate than males [27] and so have a lower basal ventilatory demand. Thus females may be able to survive to a lower absolute FEV₁ than males.

It is concluded that the FEV₁PP method is not ideal for expressing lung function impairment and should be dropped in favour of a new method of expressing FEV₁ impairment called FEV₁Q, with FEV₁·Ht⁻³ being the next best alternative. Future work should determine how these expressions of spirometric lung function impairment relate to symptoms and whether other lung function indices can be managed in a similar way.

STATEMENT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

We thank P. Lange (Hvidovre Hospital, Hvidovre, Denmark) and the research group of the Copenhagen City Heart Study (Epidemiological Research Unit, Bispebjerg University Hospital, Copenhagen, Denmark) for releasing the spirometric data for this analysis. We thank A. Dirksen (Gentofte University Hospital, Hellerup, Denmark) for the inclusion of the 1,095 chronic obstructive pulmonary disease patients followed for 15 yrs.

REFERENCES

- Hutchinson J. On the capacity of the lungs, and on the respiratory function, with a view to establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Trans* 1846; 29: 137–252.
- Thackrah CT. The Effect of Arts, Trades and Professions and of Civic States and Habits of Living on Health and Longevity. London, Longman, 1831.
- Bates DV, Macklem PT, Christie RV. Respiratory Function in Disease. 2nd Edn. Philadelphia, PA, W.B. Saunders, 1971.
- Sobol BJ, Sobol PG. Per cent of predicted as the limit of normality in pulmonary function testing: a statistically valid approach. *Thorax* 1979; 34: 1–3.
- Oldham P. Per cent of predicted as the limit of normality in pulmonary function testing: a statistically valid approach. *Thorax* 1979; 34: 569–570.
- Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265–267.
- Quanjer PH, Tammeling GJ, Cotes JE, et al. Lung volumes and forced ventilatory flows. Report Working Party. Standardization of Lung Function Tests. European Community for Steel and Coal. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.
- Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- Sorlie PD, Kannel WB, O'Connor G. Mortality associated with respiratory function and symptoms in advanced age. The Framingham Study. *Am Rev Respir Dis* 1989; 140: 379–384.
- Fletcher CM, Peto R, Tinker CM, et al. The Natural History of Chronic Bronchitis and Emphysema. Oxford, Oxford University Press, 1976; pp. 183–186.
- Miller MR, Pedersen OF, Lange P, et al. Improved survival prediction from lung function data in a large population sample. *Respir Med* 2009; 103: 442–448.
- Lange P, Nyboe J, Appleyard M, et al. Spirometric findings and mortality in never-smokers. *J Clin Epidemiol* 1990; 43: 867–873.
- Chinn S, Gislason C, Aspelund T, et al. Optimum expression of adult lung function based on all-cause mortality: results from the Reykjavik study. *Respir Med* 2007; 101: 601–609.
- Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
- Schnohr P, Jensen G, Lange P, et al. The Copenhagen City Heart Study. *Eur Heart J* 2001; 3: Suppl. H, H1–H83.
- Hansen EF, Phanareth K, Laursen LC, et al. Reversible and irreversible airflow obstruction as predictor of overall mortality in asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999; 159: 1267–1271.
- Miller MR, Pedersen OF, Dirksen A. Improved staging of chronic obstructive lung disease. *Int J COPD* 2007; 2: 657–663.
- Jacobsen PK, Sigsgaard T, Pedersen OF, et al. Lung function as a predictor of survival in the very elderly: the Danish 1905 cohort study. *J Am Geriatr Soc* 2008; 56: 2150–2152.
- Sharples L, Hathaway T, Dennis C, et al. Prognosis of patients with cystic fibrosis awaiting heart and lung transplantation. *J Heart Lung Transplant* 1993; 12: 669–674.

- 20** Miller MR, Pincock AC, Grove DM. Patterns of spirogram abnormality in individual smokers. *Am Rev Respir Dis* 1985; 132: 1034–1040.
- 21** Lange P, Parner J, Vestbo J, *et al.* A 15-year follow-up study of ventilatory function in adults with asthma. *New Engl J Med* 1998; 339: 1194–1200.
- 22** Xu X, Weiss ST, Dockery DW, *et al.* Comparing FEV₁ in adults in two community-based surveys. *Chest* 1995; 108: 656–662.
- 23** Ulrik CS, Backer V, Dirksen A. A 10 year follow up of 180 adults with bronchial asthma: factors important for the decline in lung function. *Thorax* 1992; 47: 14–18.
- 24** Bush ML, Asplund PT, Miles KA, *et al.* Longitudinal distribution of O₃ absorption in the lung: gender differences and intersubject variability. *J Appl Physiol* 1996; 81: 1651–1657.
- 25** Fowler WS. Lung function studies. II. The respiratory dead space. *Am J Physiol* 1948; 154: 405–416.
- 26** Zeman KL, Scheuch G, Sommerer K, *et al.* *In vivo* characterization of the transitional bronchioles by aerosol-derived airway morphology. *J Appl Physiol* 1999; 87: 920–927.
- 27** DuBois EF. Basal Metabolism in Health and Disease. Philadelphia, PA, Lea & Febiger, 1936.