New concepts for expressing forced expiratory volume in 1 s: conclusions need nuances

To the Editors:

We read with interest the recent publication by MILLER and PEDERSEN [1] on determining the best way of expressing lung function impairment for prediction of mortality. The authors ranked the methods of expressing forced expiratory volume in 1 s (FEV₁). It was concluded that the FEV₁ quotient (Q) was the best predictor. We believe that this outcome is in need for some nuances. This ranking is, in our view, only suitable when the parameters under investigation are all valid and when based on proper testing. We feel that this is not the case. We will illustrate this *via* the FEV₁ standardised residuals (SR) and focus on the receiver operating curve (ROC) comparison chosen for the ranking.

The authors point at the phenomenon that the number of residual standard deviation (RSD) values an elderly subject can decline is less than that of a younger subject. Based on the American Thoracic Society (ATS)/European Respiratory Society (ERS) approach, that observation is correct, but at the same time it is odd. The controversy is caused by the fact that current (ERS) prediction equations for FEV₁ deliver an age-independent RSD value, while the predicted value slowly declines with age.

The absolute minimal observable FEV₁ value equals zero and is based on the notion that the maximum number of RSDs a subject can deviate from the predicted value is easily calculated, but also limited. The equation (observed - predicted)/RSD collapses into predicted/RSD and so, with age, the maximum number of RSDs declines, starting at the youngest ages. This goes against statistical reasoning!

This odd behaviour points at problems with the statistical modelling that was used to draft the reference equations. To draw valid conclusions from multiple regression analysis performed on a sample for a general population, several assumptions must be met. First, data should range between $-\infty$ and $+\infty$. Secondly, homoscedasticity, meaning that the variance is the same for all values of the predictor values, should be present. An FEV₁ being obligatory to >0 , combined with decreasing values with age must lead to a conclusion that the FEV₁ distribution can be truncated and skewed, and that the variance is age dependent. When this is ignored in the statistical modelling the FEV₁SR approach will be flawed. As a direct consequence, any lower limit of normal defined as $-1.645 \times \text{RSD}$ is also flawed. The following example demonstrates this: the lower limit of normal of the maximal expiratory flow at 25% of forced vital capacity for a 70-yr-old male, 1.70 m in height is $-0.003 \text{ L}\cdot\text{s}^{-1}$ (ERS equation). Lower limits of normal tend to be too low for older subjects.

The lack of modelling age effects on variance and skewed/truncated distributions render the current ATS/ERS approach to convert deviations from predicted into z-scores less optimal. Currently, statistical models are available where reference

equations can be generated while the underlying (non-normal) distribution is modelled, as well as the variance, skewness and kurtosis. Thus, the current statistical models enable the assumptions necessary for a valid interpretation of multiple regression analysis to be met. Age effects on these parameters can be incorporated and valid age-dependent percentiles are possible [2, 3].

As mentioned previously, using a less optimal reference has consequences for the ranking reported by the authors. Based on the notion that not all parameters examined are correct, their ranking might change when these are devoid of the inborn errors. The conclusions drawn must, therefore, be nuanced.

A second issue is the way the ranking was obtained. The authors used all-cause mortality ROC areas to rank the various FEV₁ expressions. However, they did not test whether the differences in these correlated areas significantly differed. A ROC area can be interpreted as the probability to obtain a correct diagnosis, in this case dead or alive, and we observe that the area for FEV₁Q is marginally larger than, for example, the raw FEV₁. It differs by a mere 0.025 or the probability to obtain a correct diagnosis differs by only $0.025 \times 100\% = 2.5\%$. In our opinion this is not clinically relevant and the two FEV₁ expressions will not behave very differently in this respect. Set aside the discussion whether this all-cause mortality is suitable since the ROC areas are close to 0.5, indicating barely usable predictors.

In conclusion we feel that that using a less optimal reference has consequences for the ranking reported by the authors. The conclusions drawn must, therefore, be nuanced.

F.A.A. Mohamed Hoessein and P. Zanen

*Division of Heart and Lungs, Dept of Respiratory Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.

Correspondence: P. Zanen, Division of Heart and Lungs, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands. E-mail: pzanen@umcutrecht.nl

Statement of Interest: None declared.

REFERENCES

- 1 Miller MR, Pedersen OF. New concepts for expressing forced expiratory volume in 1 s arising from survival analysis. *Eur Respir J* 2010; 35: 873–882.
- 2 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *Appl Statist* 2005; 54: 507–554.
- 3 Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* 2004; 23: 3053–3076.

DOI: 10.1183/09031936.00064710