Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow’s ear

Athol U Wells

In treatment trials in idiopathic pulmonary fibrosis (IPF), there is an unmet need for an accurate primary end point. A European-wide consensus exists that mortality is not a practicable primary end point for the demonstration of beneficial treatment effects in IPF. Serial, 6-min-walk test data are confounded by factors other than progression of interstitial lung disease. Alone among other candidate variables, trends in forced vital capacity (FVC) have consistently predicted mortality in IPF and can, thus, be viewed as the best marker of chronic disease progression. FVC trends are now the preferred primary end point in IPF treatment trials, although not a proven surrogate for mortality.

In pharmaceutical studies, FVC change is analysed as a continuous variable, or by designating thresholds for change as ‘significant’ and quantifying time to FVC decline. Analyses of continuous change are more sensitive. However, FVC change thresholds have important theoretical advantages. Progression in IPF occurs in a stepwise fashion in some patients, and is not necessarily captured by evaluation of FVC change as a continuous variable. The designation of ‘significant’ FVC decline allows patients with ‘treatment failure’ to exit a trial with no need to continue on a demonstrably ineffective blinded therapy. In addition, time to decline in FVC can be amalgamated with mortality in evaluations of ‘progression-free survival’.

The absence of consensus on the optimum FVC threshold for change in IPF prompted Richeldi et al to compare the prognostic value of candidate thresholds. The prevailing confusion on this question cannot be overstated. In 2000, significant FVC change in IPF was designated as a 10% change from baseline, but it was unclear whether this recommendation referred to ‘relative change’ (a 10% change from baseline values—eg, from 60% to 54% of predicted), or ‘absolute change’ (a 10% reduction in percentage predicted values—eg, from 60% to 50% of predicted). In some prognostic series, FVC decline was quantified as relative change, either as a continuous variable, or using a threshold of 10%. In other reports, an absolute change threshold of 10% was evaluated.

In the 2011 American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Asociación Latinoamericana de Tórax (ATS/ERS/JRS/ALAT) IPF guideline, it was concluded that a relative decline of 10% from absolute measured baseline values (eg, a reduction in FVC from 2.0 to 1.8 litres) constituted evidence of disease progression (in the absence of an alternative explanation such as lower respiratory infection). The disadvantage of absolute change thresholds is that, for example, a 10% absolute change is a relatively minor fall in mild disease, or when emphysema coexists with IPF (with preservation of spirometric values), but represents devastating progression in severe disease (eg, a fall in FVC from 40% to 30% and, thus, a 25% fall from baseline). A threshold for change which has different clinical implications in mild and severe IPF is fundamentally unsatisfactory. Relative change in FVC does not suffer from this problem, and more closely captures the original purpose of the designation of thresholds for change: to deal with the confounding effect of measurement variation, which is expressed as the SD of change from measured baseline values.

Richeldi and colleagues now provide the first comparative prognostic examination of absolute and relative FVC decline thresholds of both 10% and 5%. This long-overdue evaluation establishes that the two previously studied 10% thresholds have similar prognostic significance. However, relative change, as recommended by the ATS/ERS/JRS/ALAT expert group, provides a higher prevalence of decline signal, a vital consideration as an absolute FVC change of 10% has been an insensitive outcome measure. From the data of Richeldi et al, it can be concluded that current guideline recommendations with regard to change in FVC should be adopted in the evaluation of IPF therapies.

Richeldi et al also evaluated the prognostic significance of lesser changes in FVC, using 5% decline thresholds following reports that these thresholds have prognostic significance in IPF for both relative and absolute changes. The need for more sensitive measures of decline in IPF is unquestionable, but both thresholds have disadvantages. An absolute change threshold of 5% is confounded by the timing of patient birthdays (within or outside the interval between tests). The more sensitive relative change threshold of 5% from measured values deals with this problem. However, this threshold may be too low in multicentre treatment studies in which there is unavoidable variability in quality assurance in participating pulmonary function laboratories. The analyses of Richeldi et al establish that there is little difference in the marginal prognostic significance of these two approaches, with an absolute change threshold of 5% enjoying a slight advantage.

In considering the relative merits of the approaches explored by Richeldi et al, the true significance of decline thresholds should not be overlooked. It is often forgotten that a relative decline of 10% in FVC is not indicative of clinically significant change, but merely denotes true disease progression, as opposed to confounding by measurement variation. Reproducibility studies have established that the SD for FVC change due to measurement variation is less than 5% (with a figure of 5% appropriate for multicentre studies). The relative FVC threshold of 10% broadly corresponds to two SDs of change: in other words, a 10% decline due to measurement variation will be seen in only 2.5% of cases and is highly likely to denote true disease progression. However, as measurement variation results equally in the overstatement and the understatement of change, a measured decline of 10% represents, in reality, a true decline ranging from 1% to 19%.
equally likely to represent trivial decline and devastating disease progression. The recognition that a relative FVC decline of 10% merely establishes a high likelihood of true decline in an individual has important implications. The use of this threshold to evaluate the clinical significance of a cohort treatment effect is a logical non-sequitur. It was erroneously argued that the treatment benefit ascribable to antioxidant therapy in IPF (a mean FVC difference of approximately 8% of baseline values) was not clinically significant, as a 10% threshold was not achieved. The same mistaken view was widely espoused in the interpretation of FVC treatment benefits in subsequent trials of pirfenidone. However, measurement variation has no net effect on cohort change, as the overstatement and understatement of change in individuals occurs with equal frequency. The use of a measurement variation threshold to evaluate the clinical significance of a cohort treatment effect is wholly inappropriate. In reality, the minimal clinically important cohort difference in FVC lies somewhere between 5% and 6%. A nihilistic approach, in which cohort treatment effects of this amplitude are dismissed as ‘trivial’, is inexorably oversimplistic.

Equally important is the problem of misclassification of change. If a 10% decline in FVC, in the absence of an alternative cause, almost certainly represents true disease progression, a 5%–10% decline (whether relative or absolute) is moderately likely to represent true decline. This is best understood by considering the fact that a 5%–10% decline due to measurement variation can be expected in roughly 15% of patients, but will occur because of disease progression in well over half of IPF patients in the placebo arms of treatment trials of 1 year. It logically follows that the categorisation of a 5%–10% decline as ‘stable disease’ is more often than not a misclassification. False positivity and false negativity are equally important problems. There is a tendency in the design of IPF studies to ‘sanitise’ FVC thresholds in the belief that FVC decline must always be designated as definite—this is viewed approvingly as ‘a rigorous approach’. However, the consequence is that for every 10 patients with a decline of 5%–10%, perhaps two or three are correctly categorised as having stable disease, but in seven or eight cases, a false negative statement of decline is made.

This approach cannot be correct, nor is it acceptable to designate decline with a probability of only 70%–80% using 5% FVC thresholds, which had only marginal prognostic significance in the study of Richeldi et al. The use of a second outcome variable to adjudicate whether 5%–10% FVC reductions in individual patients represents true decline needs to be explored. The choice of a second variable should be validated by examining possible composite end points against subsequent mortality. Candidate second end points include independently measured pulmonary function indices (with carbon monoxide diffusing capacity the most intuitively attractive), dyspnoea scores and, perhaps, disease extent on CT. In this way, it can be hoped that the likelihood that a 5%–10% decline in FVC represents true decline will increase to an acceptable level of 95% when validated by change in a second variable, while false positive statements of FVC decline will be minimised. The value of improving the sensitivity of FVC cannot be questioned, provided this can be done without sacrificing accuracy. The current limitations of FVC as the least flawed of the flawed primary end points used in IPF are widely recognised. It is high time for them to be definitively addressed.

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCE