Outcome measures in the lung

A. U. Wells¹, J. Behr² and R. Silver³

The indolent progression of lung disease in SSc has caused great difficulty in therapeutic studies as outcome measures need to be sensitive. Idiopathic pulmonary fibrosis (IPF) has been more widely studied and can usefully be extrapolated to SSc, but is more rapidly progressive. In IPF, forced vital capacity (FVC) trends are the most accurate serial surrogate for mortality. Serial gas transfer trends have a lower prognostic value in IPF and may be confounded by pulmonary vascular disease in SSc. Unresolved issues include the optimal time interval between pulmonary function tests and the mode of expression of change (percentage change from baseline vs absolute change). It has yet to be determined whether changes in pulmonary function variables are best analysed continuously or categorically (i.e. according to whether a threshold for ‘significant change’ has been reached). The 6-min walk distance has proved disappointing as an outcome variable due to major inter-test variability over the course of therapeutic studies, ascribable to extra-pulmonary factors. Serial CT is promising in principle but an optimal scoring system has proved elusive. Dyspnoea and quality of life scales provide useful ancillary information as to the likelihood that pulmonary function trends are clinically significant. For the time being, serial change in FVC appears to be the best primary end-point.

Key words: Scleroderma, Pulmonary fibrosis, Outcome, Serial pulmonary function, Serial exercise tests.

Introduction

The selection of outcome measures in lung disease in SSc in therapeutic trials poses considerable difficulties, in part because of the paucity of placebo-controlled data and the absence of a clear-cut treatment effect in other studies. Currently, two placebo-controlled studies of oral [1] and intravenous [2] cyclophosphamide in SSc lung disease are reported. The results of a placebo-controlled evaluation of bosentan are not yet available. In these studies, outcome measures were selected on the basis of clinical experience in monitoring diffuse lung disease and pulmonary hypertension. To date, the detailed consideration of outcome measures in systemic disease has not been mirrored in the respiratory system, although data have recently been presented orally on the validation of high-resolution CT of the lungs, the 6-min walk test and patient-reported outcomes [3].

At this early stage, the indolent nature of progression of interstitial lung disease in SSc in patients enrolled in recent studies appears to be the cardinal constraint. In the oral cyclophosphamide study of Tashkin et al. [1], the forced vital capacity (FVC) declined significantly (i.e. by ≥ 10% from baseline) over 1 yr in <15% of the subjects. The problem was confounded by the fact that the treatment effect in this trial amounted largely to the prevention of disease progression in patients with more extensive fibrotic abnormalities on CT, with little short-term reversibility. When therapeutic success is tantamount to stability, and little overall change is seen, the sensitivity of outcome measures to clinically important change is difficult to quantify. It is also increasingly clear that mortality is not a realistic end-point in therapeutic studies in SSc and that a surrogate for mortality is required.

Because of the paucity of data on outcome measures in the lung in SSc, recent longitudinal studies of idiopathic pulmonary fibrosis (IPF) provide an important alternative source of information and are emphasized in this review. The more rapid progression of IPF, compared with lung disease in SSc [4], needs to be kept in mind. However, it is logical to extrapolate in this way because IPF and non-specific interstitial pneumonia (NSIP), the expected histological pattern in SSc, share some clinical, pulmonary function and CT characteristics at presentation. Serial outcome measures can be divided broadly into pulmonary function tests (PFTs) at rest, exercise data, CT and patient-reported outcomes.

Serial PFTs

Historically, FVC and carbon monoxide diffusing capacity (DLco) have been used most often to quantify disease progression in clinical reports of lung disease in SSc, idiopathic NSIP and IPF. The most compelling observations endorsing serial PFT as a surrogate for mortality come from studies of mixed patient groups with IPF or idiopathic NSIP [5–7], in which serial decline in FVC or DLco at 6–12 months predicted increased mortality. In SSc, this question has been explored in only one under-powered study [8], which demonstrated a relationship between decline in DLco at 3 yrs and mortality. In IPF and idiopathic NSIP, FVC trends more consistently predict mortality [5–7], and further work in larger patient cohorts is required to determine whether this also holds true in SSc. Serial DLco is essentially unsatisfactory as an outcome measure in SSc as a decline does not discriminate between progression of interstitial disease and pulmonary vascular deterioration. Because serial resting PFTs are the only outcome measures shown to be linked to mortality in both SSc and idiopathic disease, they are increasingly preferred as primary endpoints, with FVC chosen in the US oral cyclophosphamide study [1] and in recent IPF studies. However, a number of issues remain to be clarified.

(i) The optimal time interval for repetition of PFT in treatment trials has not been established as, in the above studies of pulmonary function trends against mortality, no shorter time interval than 6 months was used to define pulmonary function trends. In most studies, including the oral cyclophosphamide study [1], PFTs were repeated every 3–4 months.

(ii) Pulmonary function trends are usually analysed as percentage changes from baseline values (rather than percentage changes of normal predicted values), although the IFIGENIA study of anti-oxidant therapy in IPF was an important exception, in that absolute changes of VC and DLco were used to define the primary end-points [9]. The strong linkage between mortality and PFT trends, expressed as percentage change from baseline, justifies the former approach.

(iii) Similarly, it is not clear whether PFT trends should be analysed as continuous data (with sub-group comparisons made using t-testing or non-parametric ranked analyses).

¹Interstitial Lung Disease Unit, Royal Brompton Hospital, London, UK.
²Department of Internal Medicine, Grosshadern Clinic, Ludwig Maximilians, University of Munich, Munich, Germany and ³Medical University of South Carolina, Charleston, SC, USA.

Submitted 1 May 2008; accepted 3 July 2008.

Correspondence to: A. U. Wells, Interstitial Lung Disease Unit, Royal Brompton Hospital, Emmanuel Kaye Building, Manresa Rd, Chelsea, London SW3 6LR, UK.
E-mail: A.Wells@rbht.nhs.uk
This approach is usual in treatment studies but an alternative approach is to examine the prevalence of 'significant' decline (categorical analysis). Continuous analysis presupposes that disease progression is likely to be broadly unimodal, whereas categorical analysis, which may be more suited to bimodal patterns of progression, lends itself to 'time to decline' analyses, such as progression-free survival [10].

(iv) 'Significant' change is defined from reproducibility data. For FVC and other lung volumes such as total lung capacity (TLC), a 10% improvement or decline from absolute values at baseline is needed to ensure that change can be confidently ascribed to alterations in disease severity, rather than to measurement variation. For DLCO estimation, which is less reproducible, a 15% change is required. Because lung disease in SSc is not rapidly progressive in most cases, evaluation of the prevalence of 'marginal' changes in PFTs (e.g. 5–10% changes in FVC, 10–15% changes in DLCO) might be fruitful. The inclusion of marginal trends in categorical analysis would increase the power of treatment studies, even if ancillary support for disease progression (such as symptomatic or serial CT change) was required.

(v) There is no current evidence that variations in the methodology of commonly performed PFTs (TLC measurement via body plethysmography vs gas dilution methods; single-breath DLCO vs re-breathing techniques; slow vital capacity as opposed to FVC) have a major influence on the accuracy or sensitivity of serial PFT. Plainly, all possible steps must be taken to reduce variability, including follow-up measurement using the same technique in the same laboratory and conformity with internationally accepted standards [e.g. American Thoracic Society/European Respiratory Society (ATS/ERS) standards for the measurement of FVC].

Serial exercise data

The role of serial maximal exercise tests, both in routine monitoring and in treatment trials, has been defined in neither SSc nor IPF, although included as a secondary outcome variable in a recent IPF trial [9]. It is now known that in IPF, maximal exercise data and especially the degree of oxygen desaturation are very poorly reproducible, as judged by major inter-test variation at an interval of 1 week [11], and this makes the definition of significant change highly problematic.

In contrast, the six minute walk distance (6MWD) was shown to be strikingly reproducible in IPF [11], comparing favourably with most routine PFTs in this regard, and this finding was subsequently confirmed at baseline in the BUILD 2 study of bosentan in lung disease in SSc [12]. Fuelled in part by the precedence of therapeutic studies in pulmonary arterial hypertension, the 6MWD was selected as the primary variable in studies of bosentan in SSc and IPF [10, 12] and in a trial of etanercept, which is yet to be reported. Based upon published data [10] and preliminary communication in the other two studies, it is now increasingly clear that the 6MWD is likely to be a highly unsatisfactory outcome variable. In IPF at least, it exhibits huge variability in the longer term with a S.D. of ~100 m between the initial and final tests and striking changes in both directions, despite relatively minor changes in other measures [10]. This endpoint may be even more problematic in SSc, a disease in which musculoskeletal, cardiac and pulmonary vascular complications may have variable but sometimes major effects on exercise capacity [13]. It is also likely that in the longer term, rehabilitation, deconditioning and adaptive ventilatory changes in stable disease to reduce the work of breathing confound the 6MWD, which can no longer be recommended as a primary outcome variable in future treatment studies in interstitial lung disease in SSc.

Serial CT data

In principle, the sensitivity of CT (compared with chest radiography) makes serial evaluation highly desirable in treatment trials. To date, no agreement has been reached on how disease should be scored on CT, with varying methods used in various therapeutic studies in SSc [1, 2] and IPF [14]. In particular, it is unclear what significance should be assigned to subtle evidence of disease progression on CT. Subjective visual assessment of image change, compared side by side, is usual clinical practice and allows a semi-quantitative estimation of the degree of change in lung regions [2]. Automated techniques, based upon measurement of changes in density, may eventually supplant visual evaluation but are currently wholly non-validated. Overall, serial change on CT is currently, at best, an exploratory secondary end-point, which might fruitfully be combined with marginal PFT changes in future validation work.

Patient-reported outcome variables

In the hierarchy of things that matter, improvements in symptoms are often valued more highly by licensing agencies than surrogates for mortality, however strong. Despite the fact that no single methodology has been validated in interstitial lung disease, inclusion of one or more dyspnoea scales for serial evaluation has been considered obligatory in treatment studies and these have included the Mahler transitional index in SSc [1] and the St George’s respiratory questionnaire in IPF [10]. The inclusion of a quality of life questionnaire that is less specific to pulmonary disease is also usual. Further work to establish the optimal means of quantifying changes in dyspnoea in interstitial lung disease is urgently required. For the moment, it can be argued that the evaluation of changes in dyspnoea and quality of life provide important ancillary support that small average changes in PFTs are more likely to be clinically significant and, importantly, that pulmonary benefits are not outweighed by systemic toxicity.

In conclusion, change in FVC appears to be the best single outcome measure in lung disease in SSc but lacks sensitivity in shorter term studies. Possible solutions include the evaluation of marginal FVC trends (perhaps validated but change on CT) and the development of more complex composite end-points, integrating serial functional trends, CT data and patient-reported outcome variables. However, these approaches are currently non-validated.

Rheumatology key messages

- Serial change in FVC, defined as percentage change from baseline, is the most widely accepted primary end-point for therapeutic studies in pulmonary fibrosis, based upon prognostic evaluation in IPF.
- Variables not suited for a primary role include a serial 6MWD (which exhibits unacceptable inter-test variability over time in pulmonary fibrosis) and serial CT evaluation (for which a scoring system has yet to be validated).
- The lack of sensitivity of current variables is a major constraint, justifying the exploration of composite end-points, possibly including marginal pulmonary function trends.

Acknowledgements

Supplement: This paper forms part of the supplement entitled ‘Update in systemic sclerosis’. This supplement was supported by an unrestricted grant from Encysive.

Disclosure statement: J.B. has received lecturing fees from Actelion, Altana, AstraZeneca, Bayer Schering Pharma, Boehringer, Encysive, GlaxoSmithKline, Pfizer, Merck, Nycomed, Schering, Zambon and ZLB Behring. J.B. is also involved in consultancies.
with Actelion, Bayer Schering Pharma, Pari and Zambon and has received industry grants from Actelion, Bayer, Mondobiotech and Mygene. R.S. and A.U.W. have both received research grant funding and honoraria as a member of a speakers’ bureau from Actelion.

References