Review: Clinical Trial Methodology

CLINICAL INVESTIGATION

New approaches to the design of clinical trials in idiopathic pulmonary fibrosis

Clin. Invest. (2013) 3(6), 531–544

Idiopathic pulmonary fibrosis (IPF) is one of the major challenges for respiratory medicine, since prognosis is particularly poor and few therapeutic options are available – in fact, at present the only approved drug is pirfenidone. Overcoming this challenge will be dependent on successful design and completion of randomized clinical trials. The last decade witnessed an unprecedented increase in quality and quantity of trials in IPF; however, most have been negative and potential obstacles have emerged. In particular, the choice of the best end point; that is, one that is both clinically meaningful and at the same time feasible, as well as the management of missing data still represent issues are not fully resolved. The increasingly competitive environment and the heterogeneity in approach by different regulatory agencies also need to be considered. During the next few years, more and more trials will be designed and completed in the hope of developing quicker and safer treatments to IPF patients.

Keywords: end point • fibrosis • lung function • missing data • mortality • randomized • statistical analysis • trials

Stefania Cerri[1], Cinzia Del Giovane[2], Sara Balduzzi[2], Francesco Soncini[1], Antonia Sdanganelli[1] & Luca Richeldi*[1]

[1]Center for Rare Lung Disease, Department of Medical & Surgical Sciences, University of Modena & Reggio Emilia, Modena, Italy
[2]Section of Statistics, Department of Diagnostic, Clinical Medicine & Public Health, University of Modena & Reggio Emilia, Modena, Italy
*Author for correspondence:
Tel.: +39 059 422 5181
Fax: +39 059 422 4231
E-mail: luca.richeldi@unimore.it

Idiopathic pulmonary fibrosis (IPF) is a specific form of interstitial lung disease of unknown etiology limited to the lungs and characterized by progressive accumulation of extracellular matrix and fibroblasts within the pulmonary interstitium. This ultimately results in complete distortion of the lung architecture leading to respiratory failure and death [1]. It has been known since its first description that the disease is progressive and irreversible and its reported median survival time from the diagnosis is approximately 2–5 years. However, very little is known about the actual natural history of the disease and no validated measures are currently available to predict which course the disease will take at the time of diagnosis. In the last decade many large, randomized, placebo-controlled clinical trials have been completed; follow-up data from the patients enrolled in the placebo groups of these trials have provided precious insights into the natural history of the disease. A lesson learnt from the experience in clinical trials is that the clinical course of the disease may vary considerably even among homogenously selected groups of patients (identified by specific predefined inclusion criteria). While disease progression seems to be the rule for the majority of cases, it is becoming well acknowledged that patients with preserved pulmonary function over time and prolonged survival can be encountered [2,3]. On the other end of the 'clinical spectrum', some patients experience rapid progression and much shorter survival time than averagely observed [4]. Whether these different clinical courses identify different disease phenotypes is still largely unknown and certainly there are no '*a priori*' measurements that allow assigning a possible clinical course to an individual patient at the time of diagnosis. In addition, at any time acute exacerbations of the disease represent unpredictable events that

FUTURE SCIENCE part of fsg

suddenly deteriorate clinical conditions and can be fatal in the majority of cases [5,6].

The last decade in IPF research has seen an increased interest in the design and accomplishment of large, well-conducted, international, multicenter, randomized, placebo-controlled clinical trials exploring the efficacy of different drugs. The input in performing well-designed clinical trials in IPF pairs with the new interest of pharmaceutical companies for the development of more effective drugs for this disease, targeting specific pathways, and with the progresses in understanding the pathogenesis mechanisms. As a result, the number of randomized clinical trials (RCTs) and the number of patients enrolled in each trial has increased and rigorous criteria for patient selection and outcome measures have been adopted.

The aim of this paper is to deliver suggestions for the design of future clinical trials in IPF, focusing on study design and methodology including selected outcomes, sample size, inclusion criteria and patient selection, statistical analysis and management of missing data. To this end, we performed a systematic literature review of published RCTs in IPF by extending the previously used search strategy 2012 up to December [7]; we then assessed the issues concerning design and methodology, to provide our opinions for the design of future trials.

Study design

Single center versus multicenter RCTs

Double-blinded, placebo-controlled RCTs represent the best study design to assess the efficacy of new drugs. In a disease with low prevalence, the collaborative efforts in the context of international multicenter studies are mandatory to ensure the enrolment of relatively large numbers of patients who have been homogeneously selected. Among the 20 RCTs included in our review (Table 1), three were single center [8–10] and 17 were multicenter [2,3,11–25]. As has happened in previous trials, it is highly likely that future clinical trials will continue to target patients with mild-to-moderate disease, which inevitably limits eligibility and requires having multiple centers involved to be able to enrich the population enrolled. On the other hand, while relying on an international global network may have its advantages (in terms of numbers of patient and speed of enrolment), some limitations of such an approach might be taken into account. For example, as therapies targeting specific pathogenic pathways are being developed, it should be considered that different genetic backgrounds might have an impact on the effects of the drug itself. Furthermore, it should be highlighted that different standards of care across countries might affect relevant outcomes such as hospitalizations or acute exacerbations. However, despite possible limitations, it is likely that the involvement of multiple centers across the world will continue to be the preferred choice for all Phase II and III RCTs in the future.

The importance of including a true placebo arm

From our review, four RCTs did not include a placebo arm [8,10,14,17], two compared active intervention versus placebo on top of standard therapy [9,13], while 14 were true placebo-controlled studies [2,3,11,12,15,16,18–25]. The importance of having a true placebo arm in IPF clinical trials has been recently highlighted by the results of the PANTHER trial conducted by the IPF-NET in the USA [21]. In 2005, Demedts *et al.* published the results of the IFIGENIA trial in which the triple therapy with prednisone, azathioprine and *N*-acetylcysteine was compared with the so-called standard treatment of prednisone and azathioprine and placebo [13]. This study showed that the triple therapy resulted in a statistically significant reduction in lung-function decline over 12 months [13]. The results of this study were largely criticized because of the lack of a true placebo arm and because of the statistical methods (in particular the least square last observation carried forward [LOCF]) used to impute the large number of missing measurements due to the high dropout rate; it was also questionable whether *N*-acetylcysteine alone had any effect in IPF. Therefore, the IPF-NET designed a three-arm RCT assessing the efficacy of the triple therapy as compared with *N*-acetylcysteine alone or placebo. The results of a predefined interim analysis, showing a significant increase in all-cause mortality and hospitalization among patients receiving the triple therapy as compared with the placebo, led to the early termination of the triple therapy treatment, while the other two arms are continuing [21]. The lesson learnt from this trial raise the question of whether a true placebo arm should always be included in RCTs in IPF, which will allow potential harmful effects from any given treatment to be assessed.

The use of a true placebo arm might become an issue in the near future, as new drugs are being approved by regulatory agencies and becoming available on the market. Pirfenidone is the first drug being approved in Europe for the treatment of mild-to-moderate IPF. The drug is already marketed in Japan and India and was also recently approved in Canada. Based on a meta-analysis of the results from the three RCTs, pirfenidone was shown to confer a 30% benefit in terms of progression-free survival (PFS), compared with the placebo [7]. In the USA, the US FDA denied approval of the drug and an additional RCT is on going. Additional drugs with positive results in Phase II trials are currently being investigated in Phase III trials, with results expected within a few years. The approval of new drugs for the treatment of IPF will raise the question of whether no

Table 1. Patient selection and inclusion criteria used in randomized, controlled trials in idiopathic pulmonary fibrosis.

Study (year)	Diagnosis of IPF	Additional inclusion criteria	Patients (n)	Duration	Sample size selection	Ref.
Noth *et al.* (2012)	Modified ATS/ERS/JRS/ALAT criteria (2000)	Progressive IPF defined as: history of worsening of dyspnea; or absolute decline of FVC ≥ 10%; or absolute decline of DLCO ≥ 15%; or reduction in arterial oxygen saturation ≥ 5%; or progression of radiographic findings	145	48 weeks	At least 95 adjudicated primary end points to achieve a 90% power to detect a difference in 48-week event-free rates of 70% for the warfarin group versus 50% for the placebo group (two-sided type I error rate of 0.05)	[20]
Raghu *et al.* (2012)	Modified ATS/ERS/JRS/ALAT criteria (2000)	FVC ≥ 50%; DLCO ≥ 30%; diagnosed within 48 months or less before enrolment	155	60 weeks	Approximately 93% power to detect a significant difference at the two-sided level of 0.05 for the hypothesized difference of 0.15 l in FVC between study groups at 60 weeks	[21]
Malouf *et al.* (2011)	ATS/ERS consensus statement (2000)	Diagnosis confirmed by SLB	89	36 months	80% power to detect a 50% difference in time disease progression between arms	[18]
Richeldi *et al.* (2011)	ATS/ERS consensus statement (2000)	FVC ≥ 50% predicted; 30% < DLCO < 79% predicted; P_aO_2 ≥ 55 mmHg (when breathing ambient air) at altitudes of up to 1500 m or P_aO_2 ≥ 50 mmHg at altitudes above 1500 m	428	12 months	80% power to detect a difference of 0.1 l in the annual decrease in FVC between groups	[23]
King *et al.* (2011)	ATS/ERS consensus statement (2000)	Diagnosis within the last 3 years, confirmed by SLB; honeycombing on HRCT in <5% of parenchyma in ≥ three out of six specified zones	616	1 year	202 events needed to detect a 35% relative risk reduction in the primary end point, overall 90% power (two-sided type I error rate of 0.05)	[16]
Noble *et al.* (2011)	ATS/ERS consensus statement (2000)	FVC ≥ 50% DLCO ≥ 35% 6MWD ≥ 150 m	779	72 weeks	Based on the placebo experience in the Raghu (2004) study and the magnitude of pirfenidone treatment effect on VC in the Azuma (2005) study. Sample size was increased during the study	[19]
Zisman *et al.* (2010)	ATS/ERS consensus statement (2000)	DLCO < 35% predicted	180	12 weeks	90% power to detect an improvement of ≥ 20% on the 6MWD at 12 weeks, based on 30% expected response rate with sildenafil and 10% assumed response rate with placebo (overall type I error of 0.05)	[25]

6MET: 6-min steady-state exercise test; 6MWD: 6-min walking distance; ALAT: Latin American Thoracic Association; ATS: American Thoracic Society; CXR: Chest X-ray; DLCO: Carbon monoxide diffusing capacity; ERS: European Respiratory Society; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; JRS: Japanese Respiratory Society; $P_{(A-a)}O_2$: Alveolar–arterial gradient of oxygen; P_aO_2: Arterial partial pressure of oxygen; PFS: Progression-free survival; SLB: Surgical lung biopsy; SpO_2: Peripheral oxygen saturation; TBB: Transbronchial biopsy; TLC: Total lung capacity; UIP: Usual interstitial pneumonia; VC: Vital capacity.

Table 1. Patient selection and inclusion criteria used in randomized, controlled trials in idiopathic pulmonary fibrosis (cont.).

Study (year)	Diagnosis of IPF	Additional inclusion criteria	Patients (n)	Duration	Sample size selection	Ref.
Taniguchi *et al.* (2010)	ATS/ERS consensus statement (2000)	≥5% difference between resting SpO_2 and the lowest SpO_2 during 6MET Lowest SpO_2 ≥ 85% while breathing air during the 6MET	267	52 weeks	80% power to detect assumed differences of the mean changes in the lowest SpO_2 from baseline to week 52 between groups, at a significance level of 0.1 (two-sided). Same statistical power calculated based on change in VC did not alter the sample size	[24]
Daniels *et al.* (2010)	ATS/ERS consensus statement (2000)	Diagnosis between 3 and 36 months FVC ≥ 55% predicted DLCO ≥ 35% predicted Clinical worsening (>10% decrease in % predicted FVC, worsening CXR or worsening dyspnea)	121	96 weeks	Based on the primary end point (assuming no dropout). 80% power (two-tailed, $\alpha = 0.05$), assuming 2-year follow up with a placebo event rate of 51% and treatment event rate of 24% To account for study drop-out, sample size increased by 20%	[12]
King *et al.* (2009)	ATS/ERS consensus statement (2000)	Diagnosis within the last 48 months, with clinical symptoms for at least 3 months Disease progression in the past 12 months 55% < FVC < 90% predicted 35% < DLCO < 90% predicted 6MWD ≥ 150 m	826	64 weeks	90% power to detect a treatment effect equivalent to a 50% reduction in 3-year mortality with a log-rank test ($\alpha = 0.025$, one-sided)	[3]
Raghu *et al.* (2008)	ATS/ERS consensus statement (2000)	Diagnosis within 2 years Measurable progression of dyspnea and progressive fibrosis on CXR; or no improvement in lung function (≤10% improvement in FVC % predicted, ≤15% improvement in DLCO); or 10 mmHg worsening in $P_{(A-a)}O_2$ with or without exertion	87	48 weeks	Based on a preliminary report 80% power to detect a difference of 17% between groups with a two-sided test at the 0.05 significance level, assuming a SD of 25%	[22]
King *et al.* (2008)	ATS/ERS consensus statement (2000)	Diagnosis within the last 3 years 150 m ≤ 6MWD ≤ 499 m	154	12 months	Based on the primary end point; Wilcoxon rank sum test to test the null hypothesis of no difference between groups at 12 months 90% power to detect a mean difference of 45 m, assuming normal distribution and a common SD of 75 m. Type I error (two-sided) = 0.05; type II error = 0.10	[15]
Demedts *et al.* (2005)	ATS/ERS consensus statement (2000)	NA	155	12 months	80% power ($\alpha = 0.05$ by two-sided test) to detect a treatment difference of 15% for VC and 20% for DLCO at 1 year (expected withdrawal rate of 25% including deaths)	[13]

6MET: 6-min steady-state exercise test; 6MWD: 6-min walking distance; ALAT: Latin American Thoracic Association; ATS: American Thoracic Society; CXR: Chest X-ray; DLCO: Carbon monoxide diffusing capacity; ERS: European Respiratory Society; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; JRS: Japanese Respiratory Society; $P_{(A-a)}O_2$: Alveolar–arterial gradient of oxygen; P_aO_2: Arterial partial pressure of oxygen; PFS: Progression-free survival; SLB: Surgical lung biopsy; SpO_2: Peripheral oxygen saturation; TBB: Transbronchial biopsy; TLC: Total lung capacity; UIP: Usual interstitial pneumonia; VC: Vital capacity.

Table 1. Patient selection and inclusion criteria used in randomized, controlled trials in idiopathic pulmonary fibrosis (cont.).

Study (year)	Diagnosis of IPF	Additional inclusion criteria	Patients (n)	Duration	Sample size selection	Ref.
Kubo *et al.* (2005)	Based on histology (SLB or TBB), or HRCT evaluation or both	Patients admitted to the hospital Demonstration of progressive deterioration of IPF despite conventional therapy without prednisolone	56	3 years	NA	[17]
Azuma *et al.* (2005)	ATS/ERS consensus statement (2000)	HRCT evidence of definite or probable UIP pattern	109	9 months	Based on a simulation study with the use of lowest SpO_2 achieved during a 6MET after 1 year. Statistical power greater than 80% to detect assumed efficacy at the significance level of 0.025	[11]
Raghu *et al.* (2004)	ATS/ERS consensus statement (2000)	Clinical symptoms for at least 3 months 50% < FVC < 90% predicted DLCO ≥ 25% predicted P_aO_2 > 55 mmHg while breathing ambient air at rest Definite or probable HRCT pattern	330	12 months	94% power to detect a 50% reduction in the primary end point (PFS at 1 year)	[2]
Ziesche *et al.* (1999)	Based on histology	Decrease in lung function ≥10% during the 12 months, despite continuous or repeated treatment with glucocorticoids, immunosuppressive agents or both for at least 6 of the 12 months	18	12 months	NA	[10]
Douglas *et al.* (1998)	Clinical diagnosis of IPF and either HRCT or histopathology criteria	NA	26	12 months	NA	[14]
Raghu *et al.* (1991)	Supported by lung biopsy	Previously untreated patients with progressive disease (i.e., progressive dyspnea ± progressive radiological parenchymal abnormality ± ≥10% decrease in FVC or TLC or ≥20% reduction in DLCO	27	12 months	NA	[9]
Johnson *et al.* (1989)	Untreated fibrosing alveolitis	Diagnosis made on clinical grounds and confirmed whenever feasible by SLB	43	36 months	NA	[8]

6MET: 6-min steady-state exercise test; 6MWD: 6-min walking distance; ALAT: Latin American Thoracic Association; ATS: American Thoracic Society; CXR: Chest X-ray; DLCO: Carbon monoxide diffusing capacity; ERS: European Respiratory Society; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; JRS: Japanese Respiratory Society; $P_{(A-a)}O_2$: Alveolar–arterial gradient of oxygen; P_aO_2: Arterial partial pressure of oxygen; PFS: Progression-free survival; SLB: Surgical lung biopsy; SpO_2: Peripheral oxygen saturation; TBB: Transbronchial biopsy; TLC: Total lung capacity; UIP: Usual interstitial pneumonia; VC: Vital capacity.

treatment will still be a valuable option in IPF. Until now, none of the available treatments (such as corticosteroids and immunosuppressive agents) was proven to be efficacious in IPF and, therefore, there were no strong recommendations for the use of any drug. As such, among many experts, no treatment was considered to be less harmful than any other therapeutic option. If different regulatory agencies apply different standards and approaches to the approval of new drugs, we could create a paradoxical situation in which placebo-controlled RCTs are feasible in some countries and not in others. It appears reasonable to think that, until new drugs are strongly recommended for the treatment of IPF based on robust evidence, it will continue to be ethically acceptable to design studies with a placebo as a comparator. A valuable alternative option would be to allow the use of approved treatments during the trial in patients reaching predefined criteria for disease progression.

Outcomes in IPF

Primary & secondary end points

Until now, despite quite uniform criteria for patient selection among different clinical trials, the choice of primary and secondary outcomes in published RCTs has been largely heterogeneous. Table 2 summarizes primary and secondary outcomes in all RCTs in IPF. Taken collectively, lung function changes, survival or disease progression measures, and patient-reported outcomes (such as quality of life or symptom scores) are almost invariably taken into account in all RCTs; however, definition of each individual outcome and its priority can vary considerably between trials. Overall survival has actually been considered a secondary outcome in the majority of RCTs already published, given the intrinsic impracticalities linked to this end point, although mortality studies in IPF are theoretically feasible [26].

Lung function decline, as assessed mainly by repeated measurements of forced vital capacity (FVC) over time, is one of the major outcome measures used both in clinical trials and in routine clinical practice to assess disease status and progression. The definition of change in FVC that has been used to assess drug efficacy in clinical trials is not univocal as well as the statistical analysis applied to measure the effect. Such heterogeneity in outcome measurements produce difficulties in pooling results from two or more studies in a systematic review. Some studies have chosen to look at FVC as a continuous variable and express the outcome in terms of absolute change; other studies have preferred to look at categorical changes of percent values above a predefined threshold. Categorical changes in FVC% predicted ≥10% over 6–12 months consistently appear to be strongly correlated with poor outcomes [27]. Additionally, a recent retrospective analysis on two large cohorts of patients with IPF showed that a relative decline ≥10% in FVC % predicted (e.g., from 70 to 63% predicted) at 12 months, while occurring almost twice as often as the absolute change, is equally able to predict 2-year transplant-free survival, as compared with an absolute decline ≥10% (e.g., from 70 to 60% predicted) [28]. This means that by using a 10% decline in FVC as a time-to-event end point, it could be possible to substantially decrease the number of patients needed to show an effect of a treatment when using a relative change instead of an absolute one. Only one trial chose to assess the efficacy of the study drug on lung function by looking at the annual rate of FVC decline [23]. This type of measurement allows modeling of a linear decrease of lung function for every patient, based on all FVC measurements available and does not imply imputation of missing data, given the *a priori* assumption of the progressive nature of the disease.

The pros and cons of different primary outcomes in IPF have already been discussed above and elsewhere in the literature [26,29]. While nobody could argue that mortality is a clinically meaningful outcome, several obstacles appear to limit the feasibility of RCTs that would like to use mortality as a primary outcome. On the other hand, if regulatory agencies will specifically require that any new drug should demonstrate a survival benefit, this might raise important questions in the design of future trials. Furthermore, for trials already designed with a different primary outcome (such as decline in FVC) and currently ongoing and recruiting patients, a decision of pretending to provide a mortality signal would possibly imply significant changes in the study design, particularly concerning the sample size, the duration of follow up and planned statistical analyses.

Clinically meaningful outcomes

A major area of criticism in IPF research concerns which outcome measure should be regarded as clinically meaningful and, should therefore become the primary outcome in all future clinical trials on this disease. While the choice of the primary outcomes matter, in the design of RCTs it should always be kept in mind that the adoption of a scientifically rigorous and patient-centered methodology is highly likely to produce an output that will be translated in a significant improvement of the quality of healthcare for all stakeholders, such as patients, physicians and policy makers [30].

A further element of complexity is related to whether the term clinically meaningful can have the same weight from a patient's perspectives as compared with physicians' or even regulatory agencies' perspectives. One group of authors has recently proposed that all-cause

Table 2. Reported outcomes in randomized, controlled trials assessing idiopathic pulmonary fibrosis treatments.

Study (year)	Primary outcomes	Secondary outcomes	Imputation of missing data	Ref.
Noth *et al.* (2012)	Composite end point (based on time to all-cause mortality; nonelective, non-bleeding hospitalization; decrease in absolute FVC ≥ 10%)	All-cause mortality All-cause hospitalization Respiratory-related hospitalization AE-IPF Bleeding Cardiovascular morbidity/mortality Lung function Change in 6MWD D-dimer levels QoL scores	None	[20]
Raghu *et al.* (2012)	Lung function (FVC)	Mortality Time to death Time to disease progression AE-IPF	NA	[21]
Malouf *et al.* (2011)	Time to disease progression (defined as time to the second of any two of a 10% change in FVC or TLC; 15% change in DLCO; a 4% change in resting room air SaO_2)	Transplant-free survival QoL scores 6MWD	Missing data before withdrawal or death were handled by imputation	[18]
Richeldi *et al.* (2011)	Lung function (FVC; annual rate of decline)	Lung function (%pred.) Change in SpO_2 Change in 6MWD QoL and symptom scores AE-IPF Overall survival Death from respiratory cause	LOCF (for secondary end points)	[23]
King *et al.* (2011)	Time to IPF worsening or all-cause mortality	QoL and symptoms scores Lung function (absolute values) Time to IPF worsening Time to death	None	[16]

%pred.: % predicted; 6MET: 6-min steady-state exercise test; 6MWD/T: 6-min walking distance/test; A–a: Alveolar–arterial; ABG: Arterial blood gases; AE-IPF: Acute exacerbation of idiopathic pulmonary fibrosis; ANCOVA: Analysis of covariance; CRP: Clinical, radiological and physiological score; DLCO: Diffusing capacity of the lung for carbon monoxide; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; LOCF: Last observation carried forward; QoL: Quality of Life; SaO_2: Arterial oxygen saturation; SpO_2: Peripheral oxygen saturation; TLC: Total lung capacity; VC: Vital capacity.

Table 2. Reported outcomes in randomized, controlled trials assessing idiopathic pulmonary fibrosis treatments (cont.).

Study (year)	Primary outcomes	Secondary outcomes	Imputation of missing data	Ref.
Noble *et al.* (2011)	Lung function (FVC)	Lung function (categorical change >10% FVC) Progression-free survival Time to IPF worsening Change in 6MWD and lowest SpO_2 during 6MWT Symptom scores Categorical change in HRCT Survival	Missing values as a result of death were assigned to the worst rank in ANCOVA analyses, and worst possible outcome in mean change analyses and categorical analyses. Other missing data were imputed with the average value from three patients with the smallest sum of squared differences at each visit with data that were not missing	[19]
Zisman *et al.* (2010)	20% improvement in 6MWD	Change in 6MWD QoL and dyspnea scores Lung function (%pred.) Change in ABG Change in A–a gradient Change in SpO_2 Overall survival	NA	[25]
Taniguchi *et al.* (2010)	Lung function (VC)	Progression-free survival Change in SpO_2 during 6MET	LOCF	[24]
Daniels (2010)	Progression-free survival (disease progression; i.e., >10% FVC decline, or death)	Lung-function (%pred. DLCO) Change in resting A–a gradient Change in 6MWD Questionnaire scores Change in modified CRP score Overall survival	LOCF	[12]
King *et al.* (2009)	Overall survival	Progression-free survival Lung function Change in 6MWD Dyspnea score Acute respiratory decompensation AE-IPF	Average value of the three patients with the smallest sum of squared differences up to the missed visit Missing values due to death were replaced with the worst possible value for each variable	[3]
Raghu *et al.* (2008)	Lung function (FVC, DLCO; %pred. change) Change in ABG	Lung function (TLC) Change in SaO_2 Change in 6MWD QoL and symptom scores Radiographic progression Progression-free survival	LOCF	[22]

%pred.: % predicted; 6MET: 6-min steady-state exercise test; 6MWD/T: 6-min walking distance/test; A–a: Alveolar–arterial; ABG: Arterial blood gases; AE-IPF: Acute exacerbation of idiopathic pulmonary fibrosis; ANCOVA: Analysis of covariance; CRP: Clinical, radiological and physiological score; DLCO: Diffusing capacity of the lung for carbon monoxide; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; LOCF: Last observation carried forward; QoL: Quality of Life; SaO_2: Arterial oxygen saturation; SpO_2: Peripheral oxygen saturation; TLC: Total lung capacity; VC: Vital capacity.

Table 2. Reported outcomes in randomized, controlled trials assessing idiopathic pulmonary fibrosis treatments (cont.).

Study (year)	Primary outcomes	Secondary outcomes	Imputation of missing data	Ref.
King *et al.* (2008)	Change in 6MWD	Progression-free survival Lung function QoL and symptom scores	LOCF	[15]
Demedts *et al.* (2005)	Lung function (VC, DLCO; absolute change)	Lung function (%pred. change) Change in CRP score QoL and symptom scores Change in maximum exercise index Change in HRCT score Number of adverse effects and withdrawals Mortality	LOCF	[13]
Kubo *et al.* (2005)	Overall survival Hospital-free period	Death	NA	[17]
Azuma *et al.* (2005)	Change in SpO_2 during 6MET	Lung function Change in ABG AE-IPF Progression by HRCT QoL	LOCF	[11]
Raghu *et al.* (2004)	Progression-free survival	Survival Lung function Change in ABG Questionnaire scores Extent of fibrosis	NA	[2]
Ziesche *et al.* (1999)	Lung function (FVC, TLC) Change in ABG	NA	NA	[10]
Douglas *et al.* (1998)	Progression-free survival Lung function (FVC ± DLCO; categorical change)	NA	NA	[14]
Raghu *et al.* (1991)	Lung function	Overall survival	NA	[9]
Johnson *et al.* (1989)	Survival Progression-free survival	NA	NA	[8]

%pred.: % predicted; 6MET: 6-min steady-state exercise test; 6MWD/T: 6-min walking distance/test; A–a: Alveolar–arterial; ABG: Arterial blood gases; AE-IPF: Acute exacerbation of idiopathic pulmonary fibrosis; ANCOVA: Analysis of covariance; CRP: Clinical, radiological and physiological score; DLCO: Diffusing capacity of the lung for carbon monoxide; FVC: Forced vital capacity; HRCT: High-resolution computed tomography; IPF: Idiopathic pulmonary fibrosis; LOCF: Last observation carried forward; QoL: Quality of Life; SaO_2: Arterial oxygen saturation; SpO_2: Peripheral oxygen saturation; TLC: Total lung capacity; VC: Vital capacity.

mortality and all-cause nonelective hospitalization should be regarded as the desirable primary outcome in all clinical trials in IPF research, arguing that they are the only outcomes that meet the definition of clinically meaningfulness [29].

The definition of clinically meaningful outcome implies that the selected end point should directly provide a measure on how a patient feels, functions or survives [31]. Based on this definition, nobody could argue that mortality is in fact a clinically meaningful outcome and that FVC can only be considered a surrogate end point: this does not automatically mean that FVC is not clinically relevant. Whether mortality should be considered the only valuable end point in IPF is a matter of debate.

Another distinction that can be made in IPF studies is among all-cause mortality and disease-specific mortality. If we think of the disease itself, which affects adults with a median age of 65 years – who might suffer from relevant comorbidities – it is not unlikely that patients might die of other causes rather than IPF. From the experience gained in large RCTs, all-cause mortality rate among patients randomized in the placebo arm can be variable (ranging from 6.2 to 17%), but is generally low (11% on average). It is possible that patient selection criteria, which specifically targets mild-to-moderate disease, can explain the low mortality rates observed in clinical trials. The recently published results from the PANTHER trial, showed a very low mortality rate in the placebo arm (1%) at 32 weeks of follow up [21]. Although the reasons for this observation remain unclear, because the trial was prematurely stopped at approximately 50% of the predefined follow-up time, it would have been reasonable to expect that deaths would have been concentrated in the second part of the study and, therefore, that the estimate of mortality as based on this prematurely discontinued trial could be artificially low. However, based on this data, any trial that would use mortality as a primary outcome, will have to either increase the sample size, or ensure a longer follow up, or both, with an increased risk of capturing more deaths from other causes rather than IPF. Moreover, a longer follow-up time might turn out to be impracticable because it will pose the ethical issue of keeping those patients whose disease is progressing in a trial and for whom alternative treatments might become available, given the increasing number of drugs that are currently entering in Phase II and III trials. Additionally, in any RCTs, it should also be taken into account that the treatment under investigation might actually cause unexpected mortality for other causes, which may also become an argument against the use of mortality in clinical trials. In such a situation, mortality will turn out to be a harm signal rather than an efficacy signal and, therefore, it will be more suitable as a secondary outcome.

Disease-specific mortality might be a more suitable outcome, but still it will require larger numbers of patients to be enrolled and longer follow up.

Despite the undoubted clinical meaningfulness of mortality as a primary outcome, its intrinsic limits have raised the question of whether other outcome measures should be preferred instead. Among others, change in FVC appears to have the characteristics to be used as a primary outcome in RCTs in IPF. It is a matter of debate whether FVC could be considered a clinically meaningful end point *per se* or a surrogate (although not fully validated) end point for mortality. Both views have pros and cons, which have already been discussed in two recently published papers [26,29]. There is consistent data that have shown that FVC can be considered the most clinically reliable measure of disease progression. The measurement itself is highly reliable, as demonstrated by the high correlation (r = 0.93) between two consecutive FVC measurements taken less than a month apart (at screening visit and at randomization) in patients enrolled in a large RCT [27]. From the data set of the same clinical trial, which was powered on mortality and did not show any statistically significant difference between the placebo and the treatment arms [3], it has also been demonstrated that absolute changes in percentage predicted FVC at 24 weeks were the stronger predictors of mortality in the subsequent year, with a risk of death approximately fivefold higher in patients with a FVC decline greater than 10%, as compared to patients with more stable lung function. An absolute decline in percentage predicted FVC between 5 and 10% also conferred an increase of more than twofold in the risk of death after 1 year [27].

The use of a composite index such as PFS, has recently been proposed as a valid outcome for IPF RCTs [32]. As discussed by Vancheri and du Bois, many similarities exist between IPF and cancer, not only at the pathobiological level, but also in terms of the magnitude of the effect that can be realistically expected as a response to treatment on disease progression [32]. Therefore, as PFS is largely being used in oncology RCTs, it should not be unrealistic to consider using PFS as a primary outcome in future trials in IPF.

Generally, more attention should be given to the selection of appropriate outcomes when RCTs have to be designed to compare the effects of different interventions, as the risk of outcome reporting bias can be minimized. Outcome reporting bias is recognized to be an important problem in RCTs [33] and, consequently, it has effects on systematic reviews [34].

In order to make a good choice about the main outcomes that should be included in all trials on IPF, we

suggest following the principles of the Core Outcome Measures in Effectiveness Trials Initiative [101]. The Core Outcome Measures in Effectiveness Trials Initiative aims to promote and support the identification of a minimum core outcome set (COS), which should be evaluated in all trials for a specific clinical area. This initiative is based on a rigorous methodology used to develop standardized sets of outcomes. Following this methodology, to better identify the COS in each specific clinical area, investigators need firstly to define the scope (e.g., different outcomes could be important for different type of patients or treatments included), then to involve different stakeholders (e.g., researchers, healthcare practitioners, industry representatives, patients) and then to collect information from all of them using a consensus method in order to achieve, at the end of the process, an agreement on the COS. Therefore, if this methodology is implemented, it is expected to find one of the selected outcomes within the COS as the primary outcome to be used in most trials. It is important to mention here that the use of such an approach does not limit the possibility to include other outcomes in a particular study [35].

Sample size

As shown in Table 1, which summarizes inclusion criteria, sample size and duration of all RCTs on IPF, the number of patients randomized in each trial has increased over time. The smallest trial was conducted in 1999 where only 18 patients were enrolled [10], whilst the largest involved 826 patients with IPF and was published in 2009 [3]. Generally, after 2008, all trials enrolled more than 100 patients on average. With regard to the methodologies used for sample size calculation, they have been different in different trials and changed based on the year of publication. In more recent trials, attention has been taken on reporting the criteria used for the calculation of the sample size; such rigor in fact was not adopted in previous trials where this information was not available. In general, the estimation of the sample sizes has been based on the primary end point and a power between 80 and 90% has been chosen. A criterion that could be taken into account for the calculation of the sample size of a future study is to base the estimation on the evaluation of the potential impact of this new study on the results of a meta-analysis in which it would be included. Recent researches have identified the statistical methods to be applied for this purpose [36,37]. The rationale for this suggestion is that meta-analysis of RCTs provides high evidence compared with a single trial, but sometimes it turns out to be inconclusive, as the confidence interval of the pooled effect size includes the situation of no effect of a treatment. Therefore, it should be more reasonable to plan the sample size estimation from results obtained in a meta-analysis than from previous single studies, taking into account the information from pre-existing evidence. Such criteria also allow a proper allocation of funds and of patients, especially considering the high medical need related to IPF.

Inclusion criteria & patient selection

As shown in Table 1, patient selection appears to be more uniform in clinical trials published in the last decade as opposed to studies performed in the late 1980s and 1990s. Starting from the study published by Raghu in 2004 [2], the internationally accepted criteria for diagnosing IPF described in the first American Thoracic Society (ATS)/European Respiratory Society (ERS) Consensus Statement [38] have been adopted in selecting eligible patients. The ATS/ERS Consensus Statement published in 2000 represented the reference document for the diagnostic criteria of IPF up to March 2011 when the new ATS/ERS/Japanese Respiratory Society/Latin American Thoracic Association Guidelines were published [1]. The adoption of international standards for the diagnosis of IPF has ensured the inclusion in clinical trials of patients with a more precise definition of the disease than what happened in the past.

With regard to disease severity, the vast majority of large, randomized, multicenter clinical trials that have been published in the last decade, with few exceptions, have specifically targeted patients with mild-to-moderate disease, as defined by lung function criteria (essentially FVC, percentage predicted, ≥50 or 55% and carbon monoxide diffusing capacity, percentage predicted, ≥30 or 35%). History of disease progression or clinical deterioration is no longer a mandatory inclusion criteria in more recent clinical trials, with the exception of the Zisman [25] and Noth [20] studies, which specifically enrolled more severe patients with progressive disease in 2010 and 2012, respectively.

Therefore, in general, the inclusion criteria appear to reflect the expected magnitude of the effect on the selected primary and secondary outcomes, as well as the natural history of the disease itself. The exclusion of severe patients from trials might, therefore, impact the labeling of approved drugs, limiting their use to the mild-to-moderate group and thus reducing the therapeutic chances for the group of severe patients (who are also those more in need of an effective treatment). The lack of evidence on the efficacy of any given drug among severe patients will possibly result in the exclusion of such patients from the recommendations for treatment with new drugs. On one hand, one could argue whether it would be ethical to expose severe patients to the risks of participating in RCTs; on the other, it is difficult to imagine that the response to

an effective drug would be worse in more advanced patients. The respiratory disease field is not new to such issues: if we think of a more common disease such a chronic obstructive pulmonary disease, many of the drugs currently in use have been separately tested in patients with different levels of disease severity.

Another reason that would justify the choice of selecting patients with mild-to-moderate disease for the enrolment in RCTs might be related to the need to retain patients able to provide valid FVC measurements during the whole study duration. In fact, we know that the disease is progressive and irreversible, and treatments currently available or under investigation will not be able to reverse the process, but only to slow down the disease progression at best. Therefore, if any given treatment shows a positive effect, it is thought to be more likely that such effect should be seen and be clinically relevant in early disease rather than in a more advanced one. In addition, given that the majority of patients with IPF tend to have a progressive clinical course, all RCTs are designed to expect significant results in primary and secondary outcomes within a standard time frame (e.g., 1 year). In fact, as reported in Table 1, the study durations in recent trials range between 48 and 72 weeks, with the exception of the study published by Zisman in 2010, which was specifically looking at short-term outcomes in severe IPF patients [25]. This reflects the fact that patients with a more severe disease (e.g., FVC, percentage predicted, <50%), have a significantly shorter survival [39] and higher risk of mortality at 1 year [40], as compared with patients with mild-to-moderate impairment of lung function. Therefore, for the purposes of the majority of clinical trials currently ongoing or under design, severe IPF patients are consistently excluded since they are more likely to die prior to the completion of a clinical trial. Whether the exclusion of severe patients would limit the population of patients in which a drug can be used, remains to be seen.

In the future, it is possible that knowledge accumulated on genomics, biomarkers and a better understanding of disease pathogenesis and natural history, will allow for the improvement of patient selection in RCTs, particularly for those studies that will evaluate the effect of molecules targeting specific biochemical pathways. A similar transition has happened in the lung cancer field: in recent years, the appraisal of the occurrence of particular genetic mutations among otherwise similar histopathologic diseases have allowed researchers to identify subsets of patients who can benefit from targeted treatments rather than conventional therapies.

Statistical analysis & management of missing data

The use of different methods of statistical analysis among different trials to measure the effect of any given treatment on a similar outcome may represent an issue when data are combined in meta-analyses. One demonstration of such a limit is shown in the meta-analysis of the clinical trials on pirfenidone. In fact, in a recently published Cochrane review, the results of the four RCTs on pirfenidone have been combined in a meta-analysis [7]. However, although change in FVC was considered as either a primary or a secondary outcome in all four studies, and met statistical significance in three of them [11,19,24], the use of different methods for statistical analysis did not allow the results on lung function of all four trials to be combined in a single meta-analysis.

Another important issue in RCTs is related to the method used to adjust for missing data. Any imputation of missing data virtually introduces a statistical bias; depending on the natural history of the disease and on the effect of treatment, the management of missing data might have a significant impact on the measured effect on prespecified outcomes. As shown in Table 2, many trials in IPF have used the LOCF method for the imputation of missing data, usually this method should never be considered appropriate. In particular, when analyzing FVC change over time, the LOCF method implies to assume that the last measurement available is the one used to calculate the effect at the end of trial. In a progressive disease such as IPF, the closer the measurement is taken to the end of the study, the higher the likelihood that the change from baseline reflects the actual outcome. However, if a study suffers from high dropout rates, particularly when this happens in the treatment arm, the use of the LOCF method might result in an over-estimation of the effect of the study drug. Therefore, the method used for imputation of missing data should always be carefully chosen and predefined.

Future perspective

IPF is a challenging disease in terms of designing a clinical trial to assess the effect of any intervention. This is mainly a consequence of the low prevalence of the disease and of many existing gaps in knowledge, first of all the lack of causative agent(s). However, it is reasonable to expect in the future that the wide availability of chest high-resolution computed tomography, coupled with precisely defined and possibly simplified diagnostic algorithms as identified in guideline documents, will increase the number of patients identified and potentially participant in clinical trials. In addition, the participation of large countries, such as China and India, will expand the target population, thus increasing the feasibility of trials. This will also allow the assessment of the effect of different genetic backgrounds.

Hopefully, more effective drugs will be identified and registered for use: one molecule has already reached this stage in Europe. This may become a potential obstacle to the enrolment of patients in placebo-controlled trials,

although there will probably be room for well-designed trials with 'rescue' options and studies designed with multiple arms. The identification of subgroups of 'responders' will become more and more of a priority, with the final aim of increasing the clinical benefit to patients, while reducing costs and time required for completion of trials.

Clinically relevant primary end points will be identified and more precisely measured, although it is possible that trials based on overall mortality will be designed and completed, nonetheless, such studies will require enormous investments and efforts. For this reason, clinical research will help with the identification of meaningful and feasible end points, which are more likely be based on lung function (IPF being a disease limited to the lungs) and quality of life. Innovative ways of measuring these end points, such as daily home measurement of lung function or handheld-based recording of quality of life components, will increase both accuracy and feasibility.

Statistical analyses will need to be homogenized and improved, although the only effective way of avoiding biases would be avoiding any missing value; however, in the real world, dropouts will always happen in an IPF trial. Better strategies to address key factors in study design, such as duration, prespecified statistical analyses and imputation of missing data, will be identified.

All of the above components will contribute to increasing feasibility and 'clinical meaningfulness' of clinical trials in IPF, thus leading to the final aim of delivering the most effective and safest treatments to patients in a reasonable timeframe.

Financial & competing interests disclosure

L Richeldi has served as a paid consultant for Boehringer Ingelheim, Intermune, Celgene, Sanofi-Aventis and Gilead for design and conduct of clinical trials in IPF. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Study design
- Collaborative efforts in the context of international multicenter studies are mandatory to ensure the enrolment of large numbers of patients with idiopathic pulmonary fibrosis (IPF).
- The use of a true placebo arm might become an issue in the near future, as new drugs are being approved by regulatory agencies and becoming available in the market.

Outcomes in IPF
- When defining a 'clinically meaningful' outcome, the different perspectives of patients, physicians and regulatory agencies should be considered and possibly integrated.
- A minimum core outcome set should be identified following a proper methodology.

Sample size
- The number of patients randomized in each trial has increased over time.
- The calculation of sample size for a future trial should take into account the effect of this new trial on the results of already existing meta-analyses.

Inclusion criteria & patients' selection
- The exclusion of severe patients from trials is likely to impact the labeling of approved drugs, thus reducing the therapeutic chances for this group of patients.
- The adoption of international standards for the diagnosis of IPF helped with the identification of IPF populations similar across studies, while, mostly, only patients with mild-to-moderate disease have been enrolled so far in clinical trials.

Statistical analysis & management of missing data
- Different statistical analyses have been used to reduce the potential bias related to missing data from dropout studies.
- Many trials in IPF have used the last observation carried forward method for the imputation of missing data: usually this method should never be considered appropriate.

References

1 Raghu G, Collard HR, Egan JJ *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am. J. Respir. Crit. Care. Med.* 183(6), 788–824 (2011).

2 Raghu G, Brown KK, Bradford WZ *et al.* A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 350(2), 125–133 (2004).

3 King TE Jr, Albera C, Bradford WZ *et al.* Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicenter, randomised, placebo-controlled trial. *Lancet* 374(9685), 222–228 (2009).

4 Selman M, Carrillo G, Estrada A *et al.* Accelerated variant of idiopathic pulmonary fibrosis: clinical behavior and gene expression pattern. *PLoS One* 2(5), e482 (2007).

5 Kim DS, Park JH, Park BK, Lee JS, Nicholson AG, Colby T. Acute exacerbation of idiopathic pulmonary fibrosis: frequency and clinical features. *Eur. Respir. J.* 27(1), 143–150 (2006).

6 Song JW, Hong SB, Lim CM, Koh Y, Kim DS. Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *Eur. Respir. J.* 37(2), 356–363 (2011).

7 Spagnolo P, Del Giovane C, Luppi F *et al.* Non-steroid agents for idiopathic pulmonary fibrosis. *Cochrane Database Syst. Rev.* 2010(9), CD003134 (2010).

8 Johnson MA, Kwan S, Snell NJ, Nunn AJ, Darbyshire JH, Turner-Warwick M. Randomised controlled trial comparing prednisolone alone with cyclophosphamide and low dose prednisolone in combination in cryptogenic fibrosing alveolitis. *Thorax* 44(4), 280–288 (1989).

9 Raghu G, Depaso WJ, Cain K *et al.* Azathioprine combined with prednisone in the treatment of idiopathic pulmonary fibrosis: a prospective double-blind, randomized, placebo-controlled clinical trial. *Am. Rev. Respir. Dis.* 144(2), 291–296 (1991).

10 Ziesche R, Hofbauer E, Wittmann K, Petkov V, Block LH. A preliminary study of long-term treatment with interferon gamma-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 341(17), 1264–1269 (1999).

11 Azuma A, Nukiwa T, Tsuboi E *et al.* Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care. Med.* 171(9), 1040–1047 (2005).

12 Daniels CE, Lasky JA, Limper AH, Mieras K, Gabor E, Schroeder DR. Imatinib treatment for idiopathic pulmonary fibrosis: randomized placebo-controlled trial results. *Am. J. Respir. Crit. Care. Med.* 181(6), 604–610 (2010).

13 Demedts M, Behr J, Buhl R *et al.* High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 353(21), 2229–2242 (2005).

14 Douglas WW, Ryu JH, Swensen SJ *et al.* Colchicine versus prednisone in the treatment of idiopathic pulmonary fibrosis. A randomized prospective study. Members of the Lung Study Group. *Am. J. Respir. Crit. Care. Med.* 158(1), 220–225 (1998).

15 King TE Jr, Behr J, Brown KK *et al.* BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care. Med.* 177(1), 75–81 (2008).

16 King TE Jr, Brown KK, Raghu G *et al.* BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care. Med.* 184(1), 92–99 (2011).

17 Kubo H, Nakayama K, Yanai M *et al.* Anticoagulant therapy for idiopathic pulmonary fibrosis. *Chest* 128(3), 1475–1482 (2005).

18 Malouf MA, Hopkins P, Snell G, Glanville AR. An investigator-driven study of everolimus in surgical lung biopsy confirmed idiopathic pulmonary fibrosis. *Respirology* 16(5), 776–783 (2011).

19 Noble PW, Albera C, Bradford WZ *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 377(9779), 1760–1769 (2011).

20 Noth I, Anstrom KJ, Calvert SB *et al.* A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care. Med.* 186(1), 88–95 (2012).

21 Raghu G, Anstrom KJ, King TE Jr, Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N. Engl. J. Med.* 366(21), 1968–1977 (2012).

22 Raghu G, Brown KK, Costabel U *et al.* Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. *Am. J. Respir. Crit. Care. Med.* 178(9), 948–955 (2008).

23 Richeldi L, Costabel U, Selman M *et al.* Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 365(12), 1079–1087 (2011).

24 Taniguchi H, Ebina M, Kondoh Y *et al.* Pirfenidone in idiopathic pulmonary fibrosis. *Eur. Respir. J.* 35(4), 821–829 (2010).

25 Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N. Engl. J. Med.* 363(7), 620–628 (2010).

26 du Bois RM, Nathan SD, Richeldi L, Schwarz MI, Noble PW. Idiopathic pulmonary fibrosis: lung function is a clinically meaningful end point for Phase III trials. *Am. J. Respir. Crit. Care. Med.* 186(8), 712–715 (2012).

27 du Bois RM, Weycker D, Albera C *et al.* Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am. J. Respir. Crit. Care. Med.* 184(12), 1382–1389 (2011).

28 Richeldi L, Ryerson CJ, Lee JS *et al.* Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 67(5), 407–411 (2012).

29 Raghu G, Collard HR, Anstrom KJ *et al.* Idiopathic pulmonary fibrosis: clinically meaningful primary end points in Phase 3 clinical trials. *Am. J. Respir. Crit. Care. Med.* 185(10), 1044–1048 (2012).

30 Gabriel SE, Normand SL. Getting the methods right--the foundation of patient-centered outcomes research. *N. Engl. J. Med.* 367(9), 787–790 (2012).

31 Temple R. A regulatory authority's opinion about surrogate end points. In: *Clinical Measurement In Drug Evaluation*. Nimmo W, Tucker G (Eds). Wiley, New York, NY, USA (1995).

32 Vancheri C, du Bois RM. Progression-free end point for idiopathic pulmonary fibrosis trials: lessons from cancer. *Eur. Respir. J.* 41(2), 262–269 (2012).

33 Dwan K, Altman DG, Arnaiz JA *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 3(8), e3081 (2008).

34 Kirkham JJ, Dwan KM, Altman DG *et al.* The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 340, c365 (2010).

35 Williamson PR, Altman DG, Blazeby JM *et al.* Developing core outcome sets for clinical trials: issues to consider. *Trials* 13, 132 (2012).

36 Roloff V, Higgins JP, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. *Stat. Med.* 32(1), 11–24 (2013).

37 Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. *Stat. Med.* 26(12), 2479–2500 (2007).

38 American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am. J. Respir. Crit. Care. Med.* 161(2 Pt 1), 646–664 (2000).

39 Nathan SD, Shlobin OA, Weir N *et al.* Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* 140(1), 221–229 (2011).

40 du Bois RM, Weycker D, Albera C *et al.* Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care. Med.* 184(4), 459–466 (2011).

Website

101 Core Outcome Measures in Effectiveness Trials. www.comet-initiative.org