The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation

Emma Loveman, Vicky R Copley, Jill Colquitt, David A Scott, Andy Clegg, Jeremy Jones, Katherine MA O'Reilly, Sally Singh, Claudia Bausewein and Athol Wells

NHS

National Institute for Health Research

The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation

Emma Loveman,[1]* Vicky R Copley,[1] Jill Colquitt,[1]
David A Scott,[2] Andy Clegg,[1] Jeremy Jones,[1]
Katherine MA O'Reilly,[3] Sally Singh,[4]
Claudia Bausewein[5] and Athol Wells[6]

[1]Southampton Health Technology Assessments Centre, University of Southampton, Southampton, UK
[2]Oxford Outcomes, Oxford, UK
[3]Department of Respiratory Medicine, Mater Misericordiae University Hospital, Dublin, Ireland
[4]Cardiac and Pulmonary Rehabilitation, University Hospitals of Leicester NHS Trust, Leicester, UK
[5]Department of Palliative Medicine, University Hospital of Munich, Munich, Germany
[6]Interstitial Lung Disease Unit, Royal Brompton and Harefield NHS Trust, London, UK

*Corresponding author

This report should be referenced as follows:

Loveman E, Copley VR, Colquitt J, Scott DA, Clegg A, Jones J, *et al*. The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation. *Health Technol Assess* 2015;**19**(20).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents*®/ Clinical Medicine.

Health Technology Assessment

> ## Criteria for inclusion in the *Health Technology Assessment* journal
> Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.
>
> Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme
The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

Abstract

The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation

Emma Loveman,[1]* Vicky R Copley,[1] Jill Colquitt,[1] David A Scott,[2] Andy Clegg,[1] Jeremy Jones,[1] Katherine MA O'Reilly,[3] Sally Singh,[4] Claudia Bausewein[5] and Athol Wells[6]

[1]Southampton Health Technology Assessments Centre, University of Southampton, Southampton, UK
[2]Oxford Outcomes, Oxford, UK
[3]Department of Respiratory Medicine, Mater Misericordiae University Hospital, Dublin, Ireland
[4]Cardiac and Pulmonary Rehabilitation, University Hospitals of Leicester NHS Trust, Leicester, UK
[5]Department of Palliative Medicine, University Hospital of Munich, Munich, Germany
[6]Interstitial Lung Disease Unit, Royal Brompton and Harefield NHS Trust, London, UK

*Corresponding author emma.loveman@soton.ac.uk

Background: Idiopathic pulmonary fibrosis (IPF) is a life-limiting lung disease that generally affects people over 60 years old. The main symptoms are shortness of breath and cough, and as the disease progresses there is a considerable impact on day-to-day life. Few treatments are currently available.

Objectives: To conduct a systematic review of clinical effectiveness and an analysis of cost-effectiveness of treatments for IPF based on an economic model informed by systematic reviews of cost-effectiveness and quality of life.

Data sources: Eleven electronic bibliographic databases, including MEDLINE, EMBASE, Web of Science, and The Cochrane Library and the Centre for Reviews and Dissemination databases, were searched from database inception to July 2013. Reference lists of relevant publications were also checked and experts consulted.

Methods: Two reviewers independently screened references for the systematic reviews, extracted and checked data from the included studies and appraised their risk of bias. An advisory group was consulted about the choice of interventions until consensus was reached about eligibility. A narrative review with meta-analysis was undertaken, and a network meta-analysis (NMA) was performed. A decision-analytic Markov model was developed to estimate cost-effectiveness of pharmacological treatments for IPF. Parameter values were obtained from NMA and systematic reviews. Univariate and probabilistic sensitivity analyses were undertaken. The model perspective is NHS and Personal Social Services, and discount rate is 3.5% for costs and health benefits.

Results: Fourteen studies were included in the review of clinical effectiveness, of which one evaluated azathioprine, three *N*-acetylcysteine (NAC) (alone or in combination), four pirfenidone, one BIBF 1120, one sildenafil, one thalidomide, two pulmonary rehabilitation, and one a disease management programme. Study quality was generally good, with a low risk of bias. The current evidence suggests that some treatments appear to be clinically effective. The model base-case results show increased survival for five pharmacological treatments, compared with best supportive care, at increased cost. General recommendations cannot be made of their cost-effectiveness owing to limitations in the evidence base.

Limitations: Few direct comparisons of treatments were identified. An indirect comparison through a NMA was performed; however, caution is recommended in the interpretation of these results. In relation to the economic model, there is an assumption that pharmacological treatments have a constant effect on the relative rate of per cent predicted forced vital capacity decline.

Conclusions: Few interventions have any statistically significant effect on IPF and a lack of studies on palliative care approaches was identified. Research is required into the effects of symptom control interventions, in particular pulmonary rehabilitation and thalidomide. Other research priorities include a well-conducted randomised controlled trial on inhaled NAC therapy and an updated evidence synthesis once the results of ongoing studies are reported.

Study registration: This study is registered as PROSPERO CRD42012002116.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

List of abbreviations

6MWD	6-minute walk distance	HR	hazard ratio
6MWT	6-minute walk test	HRCT	high-resolution computed tomography
AIC	Akaike information criterion	HRQoL	health-related quality of life
ATAQ-IPF	A Tool to Assess QOL in IPF	ICD	*International Classification of Diseases*
ATS	American Thoracic Society	ICER	incremental cost-effectiveness ratio
BAI	Beck Anxiety Inventory	IIP	idiopathic interstitial pneumonia
BDI	Baseline Dyspnoea Index	ILD	interstitial lung disease
BMI	body mass index	IPF	idiopathic pulmonary fibrosis
BNF	*British National Formulary*	IPFCRN	Idiopathic Pulmonary Fibrosis Clinical Research Network
BSC	best supportive care		
CCT	controlled clinical trial	ISHLT	International Society for Heart and Lung Transplantation
CEAC	cost-effectiveness acceptability curve	ITT	intention to treat
CFA	cryptogenic fibrosing alveolitis	K-BILD	King's Brief Interstitial Lung Disease
CI	confidence interval	LAS	lung allocation score
COPD	chronic obstructive pulmonary disease	LOCF	last observation carried forward
		LS	least squares
CQLQ	Cough Quality of Life Questionnaire	MCID	minimal clinically important difference
CRD	Centre for Reviews and Dissemination	MRC	Medical Research Council
CrI	credible interval	NAC	*N*-acetylcysteine
CRP	clinical, radiological, physiological	NICE	National Institute for Health and Care Excellence
CTC	Common Terminology Criteria		
DIC	deviance information criterion	NMA	network meta-analysis
DL_{CO}	diffusing capacity of the lung for carbon monoxide	NMB	net monetary benefit
		NSIP	non-specific interstitial pneumonia
EQ-5D	European Quality of Life-5 Dimensions	ONS	Office for National Statistics
		PaO_2	partial pressure of oxygen in arterial blood
ERS	European Respiratory Society		
EVPI	expected value of perfect information	PCT	primary care trust
		PFS	progression-free survival
FEV_1	forced expiratory volume in 1 second	PSA	probabilistic sensitivity analysis
FVC	forced vital capacity		
HIV	human immunodeficiency virus		

PSS	Perceived Stress Scale		TLC	total lung capacity
QALY	quality-adjusted life-year		TPMT	thiopurine *S*-methyltransferase
QoL	quality of life		TTO	time trade-off
RCT	randomised controlled trial		TTP	time to progression
RR	risk ratio		UCSDSBQ	University of California, San Diego, Shortness-of-Breath Questionnaire
SD	standard deviation			
SF-36	Short Form questionnaire-36 items		UIP	usual interstitial pneumonia
SGRQ	St George's Hospital Respiratory Questionnaire		UNOS	United Network for Organ Sharing
			URTI	upper respiratory tract infection
SGRQ-I	St George's Hospital Respiratory Questionnaire IPF		VAS	visual analogue scale
			VC	vital capacity
SHTAC	Southampton Health Assessements Centre		VOI	value of information
			WTP	willingness to pay
SpO$_2$	peripheral oxygen saturation			
THIN	The Health Improvement Network			

Plain English summary

Idiopathic pulmonary fibrosis (IPF) is a serious lung condition of unknown cause that scars and stiffens the lung tissue; it generally affects people over 60 years old. The main symptoms are shortness of breath and a cough, and as the disease progresses there is a considerable impact on day-to-day life. Few treatments are available. We evaluated the benefits and harms of available treatments by considering the most up-to-date, high-quality evidence, using a systematic approach. Searches for evidence from 11 databases were made and evidence was reviewed by two authors, using predefined criteria, to consider its relevance. All included studies were assessed for their quality, and data from each study were extracted into a standardised template. A narrative review and statistical methods to combine study data were applied. A statistical model was developed to compare the cost-effectiveness of pharmacological interventions for the treatment of IPF. Standard methods were used.

Fourteen studies were included in the review of clinical effectiveness. These evaluated six drugs and three non-pharmacological interventions. Results were mixed. There are few treatments which have a significant benefit. Harms from these treatments were not significant in most cases. Treatments are unlikely to be considered cost-effective. There were few studies on interventions in symptom management and palliative care. Further research is required in a number of areas including the effects of symptom control interventions and an inhaled therapy. A number of ongoing studies are yet to report and these may provide further evidence for the best approach to take for treating IPF.

Scientific summary

Background

Idiopathic pulmonary fibrosis (IPF) is a serious lung disease, the exact cause of which is not known. It generally affects people over 60 years of age and the main symptoms are shortness of breath and a cough, which can have a considerable impact on day-to-day life. IPF was once thought to progress at a steady, predictable rate, but it is now known that this is often not the case. Many people with IPF deteriorate rapidly, while others have periods of relative stability. In general, people with IPF survive for between 2 and 5 years. Evidence shows that the number of people with IPF is increasing, although the reasons for this are unclear. IPF is a difficult condition to manage, particularly in the later stages. Few treatments are available for IPF and none offers a cure. Treatments aim to reduce symptoms and improve survival. The type of treatment offered can vary and with a number of new treatments emerging it is timely to establish which are effective and provide the best value for money to the NHS.

Objectives

To evaluate the clinical effectiveness and cost-effectiveness of the different treatment strategies used within the NHS for IPF through systematic reviews of the evidence for clinical effectiveness, cost-effectiveness and quality of life (QoL), and economic modelling relevant to the UK setting to estimate the cost-effectiveness of the different treatments.

Methods

Search strategies were developed and applied to 11 electronic bibliographic databases (including The Cochrane Library, MEDLINE and EMBASE) from database inception to July 2013. Bibliographies of retrieved papers were screened and experts contacted to identify any additional published and unpublished references.

Titles and abstracts (where available) were screened for potential eligibility by two reviewers independently using inclusion criteria that were defined a priori. Screening of the full text of retrieved papers was performed by one reviewer and checked by a second. For the systematic review of clinical effectiveness, studies were eligible for inclusion if the participants had a confirmed diagnosis of IPF and the interventions under study were currently used to manage symptoms or modify IPF. Randomised controlled trials (RCTs) and controlled clinical trials were eligible for inclusion. Data extraction and assessment of methodological quality were undertaken by one reviewer and checked by a second. Differences in opinion were resolved through discussion at each stage or consultation with a third reviewer if necessary. Data were synthesised through a narrative review with tabulation of the results of included studies. Where appropriate, the studies were combined in a meta-analysis and heterogeneity was assessed. A network meta-analysis (NMA) focusing on pharmacological treatments for IPF and assessing forced vital capacity (FVC) end points was undertaken on 10 studies. The FVC end point was measured on two continuous scales and the NMA used the standardised mean difference approach.

Systematic literature searches were undertaken to identify full economic evaluations of interventions to manage IPF, and to assess the health-related quality of life (HRQoL) of people with IPF. Studies reporting HRQoL in people with IPF were eligible for inclusion if they used either generic preference-based measures or the St George's Hospital Respiratory Questionnaire, a disease-specific instrument used in IPF.

Data were extracted in a standardised form by a health economist and checked by a systematic reviewer, with any differences resolved through discussion.

A cost–utility decision-analytic model was developed to compare the cost-effectiveness of pharmacological interventions for the treatment of IPF. The model incorporates three survival curves, which are used to inform the probabilities of transition from three health states: unprogressed IPF, progressed IPF and lung transplant. Treatment effects are obtained from NMA. Utility values are applied to the health states to estimate total quality-adjusted life-years (QALYs). Costs are included for treatments, treatment monitoring, acute exacerbations, lung transplant and adverse events. The outcome of the economic evaluation is reported as cost per QALY gained.

Results

Eight hundred and fourteen references were identified by searches for clinical effectiveness. Fourteen studies were included, of which one evaluated azathioprine, three *N*-acetylcysteine (NAC) (alone or in combination), four pirfenidone, one BIBF 1120 (nintedanib), one sildenafil, one thalidomide, two pulmonary rehabilitation, and one a disease management programme. Study quality was generally good with a low risk of bias; however, where there were areas of greater risk of bias, these have been highlighted.

In patients with mild to moderate IPF, 10 studies evaluating five pharmacological interventions (azathioprine, BIBF 1120, NAC, pirfenidone and thalidomide) were included. In a small RCT, treatment with azathioprine and prednisolone led to an improvement in survival compared with placebo and prednisolone when an age-adjusted analysis was used. There was no effect on lung function. This trial had an unclear risk of bias and it is possible that the trial included participants who would have been diagnosed with non-specific interstitial pneumonia, which may in part explain the treatment effect. Follow-up was 12 months. BIBF 1120 300 mg/day was more favourable than placebo on some measures of lung function, rates of acute exacerbations and the number of deaths; however, the primary outcome of annual rate of decline in FVC was not statistically significantly different between groups in this 54-month study. Treatment with NAC was evaluated in three studies: in combination with azathioprine and prednisolone in two and as a single agent in an inhaled format in one. Follow-up was approximately 12 months in these studies. Study results were mixed, with no benefit from triple therapy on FVC compared with placebo in one study; however, there was a benefit on vital capacity when compared with double therapy in another study. Inhaled single-therapy NAC did not have a statistically significant effect compared with a control. Secondary outcomes were reported, similarly with mixed results across the three studies. The two studies with triple-therapy interventions had a low risk of bias; however, the study using nebulised NAC had an unclear risk of bias. Pirfenidone was studied in four RCTs, and meta-analysis of FVC shows that pirfenidone appears to demonstrate an effect when compared with placebo treatment. However, caution is required in interpreting these data as the outcomes pooled were different, and as a consequence a standardised mean difference analysis was undertaken; in addition, the timing of assessment of these outcomes varied (from 48 weeks to 72 weeks). Results for secondary outcomes were generally seen to be less favourable to pirfenidone. In a small crossover study, thalidomide appeared to improve cough, cough-related quality life and respiratory-related QoL, compared with treatment with placebo.

One study assessed sildenafil for those with moderate to severe IPF; the participants in this study also had evidence of pulmonary hypertension. Results on the primary outcome, a 20% improvement on the 6-minute walk test, were not statistically significant between the sildenafil and placebo groups. Results for secondary outcomes were mixed, with some favourable to sildenafil and others favouring placebo. This study followed participants for 12 weeks.

Adverse events from the pharmacological interventions were generally mild to moderate and were reasonably well balanced between the treatments and the placebo arms across the studies, with the

exception of thalidomide. Severe adverse events appeared to be more common in one study in those treated with triple therapy.

Three studies evaluated non-pharmacological treatments for populations with IPF. Two compared pulmonary rehabilitation with a control; the other compared a disease-management approach with a control. Results are uncertain with regard to pulmonary rehabilitation as differences favouring the intervention were seen for some outcomes but not others. The included studies had an uncertain risk of bias and outcomes were assessed at 10 and 12 weeks in the two trials, respectively, immediately after the cessation of the intervention. The third study reported limited evidence on the effects of a disease-management programme in IPF: there were no statistically significant differences in dyspnoea, and QoL results were mixed. This study has an uncertain risk of bias and follow-up was at 6 weeks, immediately after the programme had completed.

The fixed-effects NMA found only BIBF 1120 and pirfenidone to have a statistically significant improvement in FVC over placebo. A head-to-head comparison of BIBF 1120 versus pirfenidone showed a trend favouring BIBF, but this was not statistically significant. Caution is required in the interpretation of the results of the NMA.

One full economic evaluation of treatment for patients with IPF was identified. This examined the benefits of a testing strategy prior to treatment with NAC triple therapy but did not examine the cost-effectiveness of IPF treatment. The systematic review of QoL studies included 23 studies; results varied, but generally appeared to show that IPF has an adverse effect on HRQoL compared with population norms, and that HRQoL is likely be diminish as IPF becomes more severe.

The model base-case results show increased survival for five of the treatments compared with best supportive care (BSC), at increased cost. Only one treatment, inhaled NAC, is cost-effective at a willingness-to-pay (WTP) threshold of £30,000, but its treatment effect does not achieve statistical significance in either the single primary study or the NMA. The treatment effect of inhaled NAC compared with BSC is associated with an expected value of partial perfect information of £15.8M at a WTP threshold of £20,000.

Discussion and limitations

This evidence synthesis reports the clinical effectiveness and cost-effectiveness of a range of interventions which are currently used, or proposed to be used, to manage IPF in the UK. No previous systematic reviews have included all potentially relevant treatments for IPF, and there have been only limited economic evaluations in this area. The results of this report complement recent national guidance in the UK. The current evidence suggests that there are few treatments that have any effect on surrogate outcomes which can be linked through evidence to patient-related outcomes such as mortality. There is a scarcity of studies on interventions in symptom management and palliative care in IPF.

This evidence synthesis has been undertaken following the principles for conducting systematic reviews and economic evaluations. Limitations to this evidence synthesis include there being few direct comparisons of treatments identified. An indirect comparison through a NMA was performed; however, caution is recommended in the interpretation of these results. In relation to the economic model, there is an assumption that pharmacological treatments have a constant effect on the relative rate of FVC percentage decline.

Research is required into the effects of symptom control interventions, in particular pulmonary rehabilitation and thalidomide. Other research priorities include a well-conducted RCT on inhaled NAC therapy and an updated evidence synthesis once the results of ongoing studies are reported.

Conclusions

This evidence synthesis has identified limited evidence of the effectiveness of a number of available treatments for IPF. Pirfenidone and BIBF 1120 appear to be clinically effective; however, general recommendations cannot be made in terms of their cost-effectiveness owing to limitations in the evidence base. Further research is required in a number of areas as outlined above.

Study registration

This study is registered as PROSPERO CRD42012002116.

Funding

Chapter 1 Introduction

Background

Description of underlying health problem

Idiopathic pulmonary fibrosis (IPF) is a debilitating respiratory condition for which there is no cure. It is characterised by diffuse scarring (fibrosis) and mild inflammation of the lung tissue, leading to a gradual worsening of lung capacity. IPF is classed as an idiopathic interstitial pneumonia (IIP), which is a group of interstitial lung diseases (ILDs) also known as diffuse parenchymal lung disease.[1] IPF is the most common type of IIP, accounting for over 50% of this category of lung disease.[2]

Initially believed to develop as a result of a chronic inflammatory process, the mechanism that results in IPF is now more widely thought to be due to fibrotic processes involving the epithelial alveolar cells.[3] The disease is thought to arise as a result of recurrent injury to epithelial alveolar cells. Many different cells types have been implicated in the development of IPF, including fibroblasts, myofibroblasts, alveolar macrophages and endothelial cells. IPF is a disease characterised by aberrant wound healing in which excessive (and perhaps abnormal) extracellular matrix is deposited in the lung, thereby distorting the architecture and disrupting function. This lung injury and scarring eventually leads to a decline in lung function, which culminates in respiratory failure and death.[4] Shortness of breath on exercise and a chronic dry cough are the prominent symptoms.[1]

The natural history of IPF is not fully understood. It is a progressive chronic condition and was once thought to progress at a steady, predictable rate. However, this is often not the case, with many people deteriorating rapidly and others having periods of relative stability in their condition.[2,5] In some individuals, unexpected deterioration can occur with a sudden worsening of symptoms and resultant hypoxaemia (decreased partial pressure of oxygen in blood).[2,5] These episodes are usually without clinically apparent infection or other identifiable cause. Known as 'acute exacerbations', these are thought to occur in about 10–15% of cases and are often fatal episodes.[6]

Distinguishing IPF from other IIPs can be difficult as presentation can be similar. International consensus statements published in 2000[7] and 2002[8] provided guidelines for the definition of IPF following the identification of a new subgroup of ILD, non-specific interstitial pneumonia (NSIP), which had a substantially better survival.[1] Prior to the identification of this group, some 20–35% of people diagnosed as IPF would have had NSIP,[8] although in older populations the relative likelihood of NSIP may be lower. These guidelines also recognised that the term cryptogenic fibrosing alveolitis (CFA) was synonymous with IPF.[8] Prior to this, in the UK, the term CFA corresponded to a characteristic clinical presentation seen in IPF but common also to other IIPs.[9] In 2011, a joint statement from the American Thoracic Society (ATS), the European Respiratory Society (ERS), the Japanese Respiratory Society and the Latin-American Thoracic Society (hereafter referred to as the ATS/ERS 2011 guideline for ease of reference) provided updated guidance for the diagnosis of IPF (see *Diagnosis*).[10] No changes to the definition of IPF were made in the 2011 guideline (see below for discussion of diagnostic criteria).

Idiopathic pulmonary fibrosis is known to affect males more than females and in particular affects those in middle age. The disease is uncommon in people < 50 years of age,[2,11] and there is a peak prevalence in the eighth decade. Factors associated with the condition, for which there is no known cause, include cigarette smoking, environmental exposure, and possibly infective agents such as influenza, Epstein–Barr virus and hepatitis C.[7] Older age, male gender and smoking are all associated with shorter survival times (see *Prognosis and progression*).[12]

Epidemiology

The epidemiology of IPF is uncertain. Most estimates in the literature are based on populations aged > 55 years and the number of incident cases of IPF appears to be increasing, although the reasons for this are unclear.[11,13] One particular difficulty with estimating the descriptive epidemiology of IPF is due to the use of different case definitions for IPF. Many studies may not use the currently used definition of IPF (based on ATS/ERS 2011 guidelines) and, as such, estimates may include people with other ILDs.

Large-scale population-based assessments of the epidemiology of IPF are limited; however, four published studies have been identified which provide estimates of the incidence and/or prevalence of IPF. Two of these are based on populations in the UK and two on populations in the USA.

Incidence of idiopathic pulmonary fibrosis

In the UK, two studies have been published using data from The Health Improvement Network (THIN), which is a longitudinal primary care database recorded by general practitioners as part of routine clinical primary care. In the first of these studies, Gribbin and colleagues[11] used data from 1991 to November 2004 from 255 general practices, which the authors state represented approximately 25% of general practices that were using the particular primary care software at that time. The study identified 920 people over the age of 40 years who had received their first diagnosis of IPF during the period of study. The study reports that participants with a clinical diagnosis of IPF were included but no definition was provided. The mean age at presentation was 71 years and 62% were male. As seen in *Table 1*, the overall crude incidence rate was 4.6 [95% confidence interval (CI) 4.3 to 4.9] per 100,000 person-years. *Table 1* also shows that the incidence rates were generally higher in men than in women, and incidence also increased with increasing age.

In the most recent study using data from THIN's database, Navaratnam and colleagues[13] estimated incidence rates of IPF from all information available in the database up until July 2009 and including 446 practices. The definition of a diagnosis of IPF was not provided and although authors note that the diagnosis in the data sets has been validated, they also note that it is possible that some cases of other fibrotic lung diseases may be included. Crude annual incidence rates were calculated stratified by sex and age in 5-year age bands over the age of 55 years (see *Table 1*). After the exclusion of individuals with a variety of comorbidities, the study identified 2074 incident cases during the period of study. This equated to a crude overall incidence rate of 7.44 per 100,000 person-years (95% CI 7.12 to 7.77). As in the Gribbin study,[11] the authors found that the majority of incident cases were in men (63%) and that incidence also generally increased with age.

Fernández Pérez and colleagues[14] undertook a population-based historical cohort study in Olmsted County, MN, USA, between 1997 and 2005. The study used the ATS/ERS 2002[8] consensus statement for the case definition of IPF, with incidence rates calculated as of the date the patient met the criteria. Two sets of criteria were used: a narrow criterion based on evidence of usual interstitial pneumonia (UIP) on surgical lung biopsy or a definite pattern on high-resolution computed tomography (HRCT), and a broad criterion based on surgical lung biopsy or a definite or possible pattern on HRCT. The age- and sex-adjusted incidence rate of IPF among residents aged ≥ 50 years using the narrow IPF criterion was 8.8 per 100,000 person-years (95% CI 5.28 to 12.38 per 100,000 person-years). Using the broader criterion, the rates were estimated to be 17.43 cases per 100,000 person-years (95% CI 12.42 to 22.44 cases per 100,000 person-years). Incidence rates were directly adjusted for age, or age and sex, using the population structure of white persons in the USA in the year 2000. The incidence rates were seen to be higher in men than in women and also generally higher in the older age categories.

In a registry study using data from a large US health-care claims database (approximately 3 million people in 20 states), Raghu and colleagues[15] also estimated incidence of IPF using a narrow and broad base criteria. The broad criteria included all people aged ≥ 18 years, who were eligible for health benefits, and had had at least one medical appointment with an *International Classification of Diseases* (ICD) code for IPF and no medical appointments with an ICD code for any other ILD. To meet the narrow case definition, the person was required to satisfy the broad criteria and have had at least one surgical lung biopsy.

TABLE 1 Estimated incidence rates of IPF from four population-based registry studies

UK-based	
Gribbin et al. 2006[11]	
Population-based cohort study	
Population: from 255 general practices 1991 until November 2004	
Incidence of IPF (95% CI) per 100,000 person-years	
Crude rate	4.6 (4.3 to 4.9)
Men	5.69 (5.24 to 6.18)
Women	3.44 (3.10 to 3.82)
Age (years)	
< 55	0.54 (0.43 to 0.67)
55–64.9	7.30 (6.27 to 8.50)
65–74.9	17.06 (15.20 to 19.14)
75–84.9	25.37 (22.67 to 28.40)
≥ 85	22.37 (18.04 to 27.74)
Navaratnam et al. 2011[13]	
Population-based cohort study	
Population: from 446 general practices until July 2009	
Incidence of IPF (95% CI) per 100,000 person-years	
Crude rate	7.44 (7.12 to 7.77)
Men	9.46 (8.96 to 9.98)
Women	5.46 (5.07 to 5.86)
Age (years)	
≤ 54	0.86 (0.75 to 1.00)
55–59	10.48 (9.06 to 12.13)
60–64	20.76 (18.34 to 23.50)
65–69	36.45 (32.99 to 40.27)
70–74	47.57 (43.26 to 52.32)
75–79	47.38 (42.76 to 52.49)
80–84	60.05 (52.47 to 68.73)
> 85	34.82 (27.55 to 44.01)

USA-based		
Fernández Pérez et al. 2010[14]	**Broad definition**[a]	**Narrow definition**[a]
Population-based cohort study		
Population: 128,000 residents; 596 people initially thought to have IPF were screened from Olmsted County, MN, USA, between 1997 and 2005		

TABLE 1 Estimated incidence rates of IPF from four population-based registry studies (*continued*)

USA-based		
Fernández Pérez *et al.* 2010[14]	**Broad definition[a]**	**Narrow definition[a]**
Incidence of IPF (95% CI) per 100,000 person-years		
Age- and sex-adjusted, aged ≥ 50 years	17.43 (12.42 to 22.44)	8.8 (5.28 to 12.38)
Men, age-adjusted	24.02 (14.84 to 33.20)	13.38 (6.51 to 20.24)
Women, age-adjusted	13.43 (7.50 to 19.37)	6.08 (2.08 to 10.08)
Age, men (years)		
50–59	1.64	1.64
60–69	21.39	10.69
70–79	42.88	21.44
≥ 80	66.02	41.26
Age, women (years)		
50–59	1.55	1.55
60–69	12.44	4.98
70–79	36.79	16.72
> 80	11.78	3.93
Raghu *et al.* 2006[15]	**Broad definition[b]**	**Narrow definition[b]**
Population-based cohort study		
Population: from US health plan covering 20 states and approximately 3 million people, between 1996 and 2000		
Incidence of IPF per 100,000		
Age- and sex-adjusted	16.3	6.8
Age, men (years)		
18–34	2.8	0.9
35–44	1.1	0.0
45–54	11.4	6.2
55–64	35.1	12.2
65–74	49.1	21.3
≥ 75	97.6	38.5
Age, women (years)		
18–34	0.0	0.0
35–44	5.4	4.5
45–54	10.9	5.4
55–64	22.6	9.9
65–74	36.0	16.6
≥ 75	62.2	19.5

a Broad: usual interstitial pneumonia (UIP) on surgical lung biopsy or a definite or possible pattern on high-resolution computed tomography; narrow: UIP on surgical lung biopsy or a definite pattern on high-resolution computed tomography.
b Broad: aged ≥ 18 years, eligible for health benefits, ≥ 1 medical appointment with an *International Classification of Diseases* (ICD) code for IPF and none with an ICD code for any other ILD; narrow: all of the broad criteria and ≥ one surgical lung biopsy.

Data analysed were between 1996 and 2000 inclusive. Incidence rates were estimated by combining age- and sex-specific rates from the study with population weights from the US census. In this study, overall incidence rates for those diagnosed in 2000 were estimated to be 16.3 per 100,000 persons using the broad case definition, and 6.8 per 100,000 persons using the narrow case definition. The study demonstrated increasing incidence by age (see *Table 1*). Similar to other studies, incidence rates were also generally higher in men than in women.

Overall, the estimates suggest an incidence rate in the region of approximately 4.6 to 8.8 per 100,000, although this depends on the definitions of IPF used.[11,13–15]

In the two UK-based studies, analyses confirmed that the incidence of IPF had increased over the two time periods studied. In the Gribbin and colleagues[11] study, after adjusting for age and sex, the annual increase in incidence of IPF was estimated to be 11% (rate ratio 1.11, 95% CI 1.09 to 1.13; $p < 0.0001$). In the Navartnam and colleagues[13] study, after adjustment for age, sex and health authority, the estimated annual increase in incidence was 5% (rate ratio 1.05, 95% CI 1.03 to 1.06).

Prevalence

The two studies undertaken in the USA also reported estimated prevalence rates. Age- and sex-adjusted prevalence rates among people aged ≥ 50 years in the study undertaken in Olmsted county[14] was 27.9 (95% CI 10.4 to 45.4) cases per 100,000 people using the narrow criteria for IPF. For the broad criteria, the estimated prevalence was 63 (95% CI 36.4 to 89.6) cases per 100,000. In the Raghu study,[15] using data from a health-care claims database, the prevalence of IPF was estimated to be 42.7 per 100,000 persons using the broad definition, and 14.0 per 100,00 persons using the narrow definition. The prevalence was also seen to increase with age and was generally higher in men than in women in this study. Overall estimates suggest prevalence of between 14 and 30 per 100,000 persons when using a narrow definition for IPF.[14,15]

Mortality

Mortality rates were presented from the two UK studies using data from THIN's primary care database.[11,13] In the Navaratnam and colleagues[13] study, data were also analysed from routine mortality data from the Office for National Statistics (ONS) derived from death certificates in England and Wales between 1968 and 2008 and applied to the 2008 population. In the database cohort, the crude mortality rate was 228.8 per 1000 person-years (95% CI 193.8 to 216.7 per 1000 person-years). From the routine mortality data, the overall mortality rate standardised to the 2008 UK population over this period of time was 2.54 per 100,000 person-years (95% CI 2.52 to 2.56 per 100,000 person-years). In the Gribbin study,[11] the crude mortality rate for people with IPF in THIN's database cohort was 180 per 1000 person-years (95% CI 164 to 198 per 1000 person-years). In both studies mortality rates were higher in men and in older populations.

Prognosis and progression

The prognosis for individuals with IPF is poor.[2] Mean survival with IPF is generally estimated to be between 2 and 5 years from diagnosis.[1,16] In the study by Navaratnam and colleagues,[13] median survival was estimated to be 3.03 years, with an estimated 5-year survival of 37%. The Gribbin and colleagues[11] study showed the 3- and 5-year survival to be 57% and 43%, respectively.

The prognosis for an individual patient remains difficult to define and can be highly variable.[17] Pulmonary assessments can be useful to predict the course of the disease and which patients have a higher likelihood of death within the next year.[16] A decline in per cent predicted forced vital capacity (FVC) of at least 10% in a 6-month period is associated with a nearly fivefold increase in the risk of mortality.[18] From a single time point other pulmonary function tests, such as the diffusing capacity of the lung for carbon monoxide (DL_{CO}), and exercise tests such as the 6-minute walk test (6MWT), have also been shown to be useful prognostic determinants in some studies.[18]

Acute exacerbations of IPF are also known to have an impact on the prognosis of IPF. These periods of rapid deterioration of IPF without infection, pulmonary embolism or heart failure have an impact on the overall survival of patients with IPF, with mortality possibly being as high as 75%.[1,17,19]

Although the evidence on the incidence and risk factors of acute exacerbation of IPF is limited, and varies between studies, the incidence is thought to be in the region of 10–15%.[1,6,17,19] Risk factors for acute exacerbations include a low FVC, DL_{CO} total lung capacity (TLC), and never having smoked.[17]

The presence of other comorbidities may also have an impact on the prognosis of IPF. Some evidence supports the view that coexisting emphysema leads to a poorer prognosis.[12,20] Although the coexistence of emphysema leads to a preserved lung volume, it also leads to an impaired diffusion capacity, and results of studies suggest that the fibrosis is the dominant prognostic factor.[20] Rates of coexisting emphysema may be as high as 30% in IPF.[12] Pulmonary hypertension has also been shown to be an important determinant of disease outcomes in IPF.[1,21] Pulmonary hypertension is characterised by increased pressure in the pulmonary arteries, and most patients with end-stage disease have this.

While there is no optimal marker for disease severity, recent evidence supports the view that the per cent predicted FVC can be useful to categorise people into three severity groups: mild, moderate and severe. Caution is required in using these cut-offs because they do not take into account other patient factors such as the existence of comorbidities; however, a crude categorisation is normal FVC (> 80% predicted), mild (≥ 70%), moderate (55–69%), and severe (< 55%).[22,23]

Impact of idiopathic pulmonary fibrosis

Presentation of an individual with IPF is often as a result of a gradual onset of shortness of breath on exertion. This non-specific symptom can be wrongly attributed to the ageing process, emphysema, heart disease or a respiratory tract infection, and therefore diagnosis can often be made some time after initial presentation.[1,5] In others, IPF is an incidental finding on a routine chest examination. Key symptoms of IPF include breathlessness (dyspnoea), a non-productive (dry) cough, which can be paroxysmal (spasmodic) in nature, reduced exercise tolerance and anxiety. In a retrospective review of 45 case notes of deceased patients, > 90% suffered from breathlessness and 60% from cough.[24] Other, less frequently documented symptoms were fatigue, depression/anxiety and chest pain.[24] Clark and colleagues[25] described a high prevalence of nocturnal hypoxaemia with decreased energy levels and impaired daytime social and physical functioning.

Symptoms become progressively worse over time.[1] The irritating dry cough associated with IPF has a significant impact on a patient's life, leading to a reduced quality of life (QoL).[4] Finger clubbing is found in approximately 50% of patients.[1] A progressive worsening of these symptoms, and deterioration in lung function, increasingly limits normal physical activity, and the individual becomes more debilitated and disabled.[3,18] The result of this is often an incremental shift to becoming housebound, dependent on others for completion of normal activities, and dependent on oxygen therapy.[2] This leads to death from respiratory failure or a complicating comorbidity.[2]

Breathlessness is one of the main distressing symptoms; however, no treatments aiming to modify the disease process have been shown to improve breathlessness in IPF.[26] Therefore, symptomatic treatment is essential. Some IPF patients may need only short-burst oxygen initially for episodic breathlessness, but many will need long-term oxygen therapy and ambulatory oxygen to maximise their QoL.[27] IPF is recognised as a clinical indication for referral for home oxygen.[27] If breathlessness continues to burden a patient, opioids and benzodiazepines, often in combination, can be prescribed as for other patients with breathlessness in advanced disease.[28]

In a retrospective review of ILD patients, Bajwah and colleagues[24] noted that, beyond symptoms, there was a lack of documentation of spiritual needs and little documentation of assessment for depression and anxiety, or documentation of preferred place of care or preferred place of death. The authors commented that it is likely that these issues occur in these patients.[24]

Many patients with IPF are admitted to hospital and hospices, although accurate data are scarce. Many studies have small sample sizes and data available predominantly relate to hospitals. One recent study undertaken in the USA followed 168 patients over a 76-week period and found that 23% of these were

hospitalised for respiratory-related illnesses on a total of 57 occasions.[29] The mean number of days stay in hospital was 15 days, with the most common reason for hospitalisation being suspected infection. Data from the UK suggest that in 2008–9 there were 9500 finished consultant episodes for people categorised as 'other interstitial pulmonary diseases with fibrosis', with around 600 hospital admissions.[30] The mean length of stay for these people was 9 days.

Idiopathic pulmonary fibrosis is a difficult condition to manage, particularly in the later stages. Early and accurate diagnosis is important to maximise the potential for a better outcome.

Diagnosis

The 2011 ATS/ERS consensus statement for IPF[10] states that 'IPF should be considered in all adult patients with unexplained slowly progressing exertional dyspnoea, and commonly presents with cough, bibasilar inspiratory crackles, and finger clubbing' (p. 792). Treatments for IPF predominantly aim to reduce the decline in lung function and therefore early diagnosis is important to improve prognosis. Despite this, in many cases, a diagnosis is made at the point where lung function abnormalities are more severe, with the individual often having had asymptomatic disease progression for some years before.[2] One reason for this is that the diagnosis of IPF is a challenge because there are no specific abnormalities on laboratory tests. Tests which suggest an inflammatory response in an individual, such as erythrocyte sedimentation rate, and rheumatoid factor, may be abnormal, but these are not specific to IPF. For the general practitioner, the tests available, and the symptoms shown by the patient, could be indicative of a number of other conditions. Guidelines for the diagnosis of IPF have, however, recently been produced in the UK by the National Institute for Health and Care Excellence (NICE).[31]

The gold-standard diagnosis of IPF requires precision and a multidisciplinary approach involving ILD specialists, radiologists and pathologists.[1,2,31] Correlation of the clinical, radiological and histopathological features increases the accuracy of an IPF diagnosis and is often made on an individual basis. Clinical symptoms, findings on HRCT, and results of lung biopsy are taken in context with one another, especially as results on each factor can sometimes be discordant. Although surgical lung biopsy is recommended to confirm all suspected cases of IPF, in some settings surgical lung biopsy may not be required and the diagnosis is made on clinical and HRCT findings alone.[1] This is especially likely when the risks of a surgical procedure are weighed up with the potential delay in diagnosis while awaiting a surgical lung biopsy.[2] If surgical lung biopsy is used, biopsies from two locations are usually taken.

High-resolution computed tomography allows a detailed examination of the pulmonary parenchyma, and estimates of diagnostic accuracy are thought to be high (specificity exceeding 90%). The primary role of HRCT is to discriminate between typical IPF and other HRCT appearances, which may be indicative of other ILDs but may also represent IPF with atypical HRCT appearances.[1]

In the recent guidelines,[10,31] the recommendation for diagnosis requires exclusion of other known causes of ILD (e.g. domestic and occupational environmental exposures, connective tissue disorders and drug toxicity); the presence of a UIP pattern on HRCT in patients not undergoing surgical lung biopsy; and specific combinations of HRCT and surgical lung biopsy pattern in patients undergoing surgical lung biopsy.[10]

The ATS/ERS consensus guideline provides details of the criteria for UIP based on HRCT and histopathology (from surgical lung biopsy) in tabular form. The details for a definite UIP pattern on HRCT include subpleural, basal predominance; reticular abnormality; honeycombing with or without traction bronchiectasis. Findings for a judgement of 'possible UIP pattern' and an 'inconsistent with UIP pattern' are also presented. Histopathological criteria for UIP pattern include evidence of marked fibrosis; presence of patchy involvement of lung parenchyma by fibrosis; and presence of fibroblast foci. Findings for a judgement of 'probably UIP pattern', 'possible UIP pattern' and 'inconsistent with UIP pattern' are also presented. In addition, the guideline presents the specific combinations of HRCT and surgical lung biopsy patterns expected to make a diagnosis of IPF, using the criteria set out for each method as summarised above. These criteria are presented in *Table 2*.

TABLE 2 Combination of HRCT and surgical lung biopsy for the diagnosis of IPF

HRCT pattern						
UIP		Possible UIP			Inconsistent with UIP	
Surgical lung biopsy pattern (if performed)						
UIP, probable UIP, possible UIP, nonclassifiable fibrosis[a]	Not UIP	UIP, probable UIP	Possible UIP, nonclassifiable fibrosis[a]	Not UIP	UIP	Probable UIP, possible UIP, nonclassifiable fibrosis,[a] not UIP
Diagnosis of IPF						
Yes	No	Yes	Probable[b]	No	Possible[b]	No

a Some biopsies may reveal a pattern of fibrosis that does not meet the criteria for UIP pattern and the other IIPs.
b Multidisciplinary discussion should include discussions of the potential for sampling error and a re-evaluation of adequacy of technique of HRCT.
Represented from American Thoracic Society.[10]

Current service provision

As discussed above, IPF is often misdiagnosed because the early manifestations of the disease are non-specific. This can lead to significant delays between symptom presentation and the correct diagnosis in a disease where diagnosing IPF at an early stage is important to maximise the potential for treatment.[32,33] In addition, the incorrect diagnosis can lead to initiation of ineffective or potentially harmful treatments and delay the opportunity for possible lung transplant (discussed in *Lung transplant*, below).[33] In some cases, the delay to diagnosis may be patient related, where individuals can be reluctant to acknowledge the symptoms that they have and to seek help.[34]

Evidence suggests that longer delays in accessing specialist care in IPF are associated with a higher risk of death.[34] A US-based cohort study of 418 adults referred to a tertiary centre found that a median delay from symptoms to initial evaluation at the centre was 2.2 years.[34] Longer delay was associated with an increased risk of death, independent of age, sex and baseline FVC per cent predicted, with an adjusted hazard ratio (HR) per doubling of delay to access of 1.3 (95% CI 1.03 to 1.6). Results of a survey of 1251 IPF patients and 197 caregivers conducted in the USA found that 55% reported at least a 1-year delay between initial presentation and diagnosis, with some 38% having been seen by three or more physicians.[35] Incorrect diagnoses included bronchitis, asthma, chronic obstructive pulmonary disease (COPD), emphysema and heart disease. The study also found that 64% of responders agreed that there was a lack of information and/or resources on IPF when they were diagnosed. Only half felt generally, or well, informed about the treatment options available to them.

A European qualitative study of 45 IPF patients from five European countries found that the median reported time from initial presentation with symptoms to diagnosis of IPF was 1.5 years.[36] This ranged from < 1 week to 12 years, and in 58% of cases the delay was > 1 year. Similar to the larger US-based survey, a proportion of patients had consulted more than three physicians; in this study the rate was 55%. The study also investigated feelings of satisfaction with medical care and information about IPF, and found this was highest in those who received care at centres of excellence.

While these figures are based on studies which have small samples, and may not have participants who are generalisable to those seen in the UK, the results suggest that a large proportion of patients experience delays in obtaining a diagnosis, and it is not anticipated that these are significantly different from what is experienced in the UK.

The 2013 clinical guideline produced by NICE[31] outlines the key clinical features to be used in primary care to identify and assess possible IPF. The guideline recognises the fact that the initial assessment of these individuals needs to be improved, to reduce the risk of delays in diagnosis and initiation of treatments, including monitoring and best supportive care (BSC). The clinical features may include age > 45 years, persistent breathlessness on exertion, persistent cough, bilateral inspiratory crackles and finger clubbing. Spirometry may or may not be impaired. In cases of suspected IPF, people can then be referred to secondary care for diagnosis and clinical management. The guideline states that the diagnosis should be multidisciplinary at each stage of the diagnostic care pathway to include a consultant respiratory physician, consultant radiologist and ILD specialist nurse. All patients should be given BSC from the point of diagnosis, which includes information and support, symptom relief, management of comorbidities, withdrawal of ineffective therapies and end-of-life care. In addition, individuals should be assessed for pulmonary rehabilitation, if appropriate, and have a clinical nurse specialist available to them. The guideline also recommends that lung transplantation in those without contraindications should be considered between 3 and 6 months after diagnosis (or sooner if indicated clinically), with referral to regional transplant units for assessment if appropriate.

Pulmonary rehabilitation

Individuals with ILD experience many of the same symptoms as those with COPD, commonly shortness of breath, fatigue and reduced exercise tolerance. It would, therefore, seem a priority to address these symptoms. Pulmonary rehabilitation is an established evidence-based intervention for individuals with COPD, with the precise aims of reducing the physical and emotional impact of the disease on the individual.[37] It is well documented that rehabilitation improves exercise capacity, dyspnoea and QoL.[37] The delivery of rehabilitation to individuals with COPD is advocated in national guidelines; however, until the 2013 NICE guidelines this had not been extended to IPF.[31] Many individuals are, therefore, currently not offered pulmonary rehabilitation. Further details about the intervention itself can be found in *Description of technologies under assessment*.

The referral to rehabilitation can be instigated by primary or secondary care physicians; in most cases, this would be after a detailed assessment and optimisation of treatment from a physician with a special interest in IPF. Prior to commencing a course of rehabilitation, medical management should be optimised; this may include the provision of supplemental oxygen. The most advantageous time to refer is unclear in the evidence, although it is anticipated that a prompt referral is more likely to be beneficial before an individual is too disabled by the disease.

Lung transplant

Lung transplantation is the only treatment that has been shown to improve survival in patients with IPF. The 5-year survival of IPF patients post transplant is in the region of 40–50%, which compares with an overall 52% 5-year survival rate.[38] As indicated in the NICE guidelines,[31] treating clinicians should carefully consider whether or not their patient might be eligible and refer potentially suitable patients to a recognised transplant centre for a full evaluation. IPF is the most common ILD referred for lung transplantation throughout the world. Data from the International Society for Heart and Lung Transplantation (ISHLT) collected from 1 January 2011 to 30 June 2012 show IPF to be the third most common diagnosis leading to lung transplant throughout Europe, representing 17% of all lung transplants.[38] Data from the NHS Blood and Transplant service reveals that 17% (numbering between 23 and 33 patients per year) of all lung transplants in the UK between April 2009 and March 2013 were for fibrosing lung disease (NHS Blood and Transplant, 26 February 2013, personal communication). It is anticipated that the majority of patients in this category had IPF. This differs significantly from the current experience in the USA, where IPF is now the most common indication for lung transplantation (35% of transplants).[39] In the USA, the allocation of organs is governed by the United Network for Organ Sharing (UNOS) system which calculates a lung allocation score (LAS). This calculates a post-transplant survival score and a waitlist survival score to give an overall LAS. Patients with a higher score are prioritised on the basis of urgency.

As IPF is a disease of older adults, many patients may not be considered for lung transplantation on the basis of their age. Most centres use age > 65 years as a relative, but not an absolute, contraindication to lung transplant. Data from ISHLT from 2011 to the first half of 2012 reveals that, worldwide, 15.4% of patients who received a transplant were > 65 years of age.[38] However, again, there was a considerable difference between European and American practice, with 4.4% of European recipients being > 65 years, compared with 25.3% of North American recipients. Recent data from the USA suggest that post-transplant mortality is no higher in patients > 70 years when compared with those in the 60- to 69-year-old age group.[40]

People considered for transplant referral should be free of malignancy for at least 2 years and have no evidence of significant coronary artery disease, heart failure or any significant disease in a vital organ. They should be not be infected with human immunodeficiency virus (HIV), chronic hepatitis B or C infection or be dependent on tobacco and other substances. Similarly, a lack of social support and significant psychological problems are contraindications to lung transplant. Poor nutritional status and being significantly overweight [body mass index (BMI) > 30 kg/m^2] or underweight is another important contraindication.[41]

Data from the UNOS show that overall mortality post lung transplant is 22% higher in patients with a BMI > 30 kg/m^2.[42] Low-dose prednisolone use (< 0.3 mg/kg/day) is not considered a contraindication.

In patients who have no contraindications to lung transplant, the following specific criteria have been proposed to guide a timely referral in IPF:[41]

1. radiological or pathological evidence of UIP
2. diffusion capacity/transfer factor for carbon monoxide < 40% predicted
3. a reduction of 10% in the FVC over a 6-month follow-up period
4. a reduction in oxygen saturation below 88% on 6MWT
5. the presence of honeycombing on a HRCT scan (fibrosis score > 2).[43]

However, as more patients with IPF than any other condition die on lung transplant waiting lists, the clinician should have a low threshold to refer a deteriorating patient even if they do not yet meet the above criteria.[44] Similarly, many experts would advocate early referral for a deteriorating patient even if the radiological and/or pathological findings are not entirely consistent with a UIP pattern, as evidence of deterioration is a powerful marker of a poor prognosis.

Traditionally, single-lung transplant has been preferred over double-lung transplant in IPF, with the latter being preferred for patients with cystic fibrosis or conditions associated with chronic infections. Some centres have suggested that long-term survival may be better post double-lung transplant,[38,45] while others report little difference.[46,47] This remains a point of considerable debate and given the lack of availability of organs for donation it is likely that single-lung transplant will continue to be widely used. Exceptions to this may be in younger patients or those with a nidus of infection in the native lung.

Palliative care

As stated, patients with IPF have a poor prognosis, although determining individual prognosis is more difficult than in, for instance, lung cancer, and treatment options are limited.[27,48] Lee and colleagues[49] describe three 'pillars' of care: disease-centred management, symptom-centred management, and education and self-management.

American Thoracic Society guidelines suggest that palliative care principles should be initiated when a patient with respiratory disease becomes symptomatic.[48] This may be particularly helpful in IPF as diagnosis can be variable and there may be rapid and unexpected deterioration resulting in death.[27] The guideline of the British Thoracic Society recommends that BSC should be considered a specific and important treatment strategy in all patients with IPF.[50] The recent NICE guideline[31] states that BSC should be initiated

at the point of diagnosis. Best supportive, or palliative, care is a proactive approach to symptomatic treatment and may include oxygen therapy, pulmonary rehabilitation, opiates, antireflux therapy, the withdrawal of steroids and other immunosuppressants, early recognition of terminal decline and liaison with palliative care specialists.[50] This should also allow patients with IPF the opportunity to consider their expectations, goals and preferences, but there is a reluctance of professionals to broach these difficult topics.[27] This might be one explanation as to why the majority of IPF patients in a recent study had no palliative care input in their last year of life, despite these recommendations regarding the early integration of palliative care.[24]

The NICE 2013 guideline suggests that if a patient with IPF has breathlessness on exertion, ambulatory oxygen should be considered.[31] Where there is breathlessness at rest, consideration should also be given to benzodiazepines and/or opioid therapy.

Description of technologies under assessment

There is no universally accepted best treatment for IPF, with several treatment options available to clinicians. Disease-modifying treatments include immunosuppressants, antifibrotics and antioxidants, alone or in combination. Symptomatic treatments available include oxygen therapy, opioids, corticosteroids, antireflux therapy and pulmonary rehabilitation. Other treatments may be used to treat specific symptoms such as intractable cough or pulmonary hypertension. Finally, in some cases, lung transplantation is considered.

Disease-modifying treatments

Cyclophosphamide and azathioprine are immunosuppressant treatments which may also suppress inflammation and have been used in some patients with IPF. Both cyclophosphamide and azathioprine are used, in some cases in combination with the corticosteroid prednisolone. The ATS/ERS 2011 consensus guidance suggests that these are not used routinely in IPF.[10] The NICE guideline states that where the combination of azathioprine and prednisolone is already being used, there should be consideration of gradual withdrawal of these treatments.[31] In the UK, the use of cyclophosphamide and prednisolone is mostly restricted to the treatment of acute exacerbations, although not widely used.

N-*acetylcysteine (NAC)* is a precursor to the antioxidant glutathione, which may be reduced in the lungs of patients with IPF. It is an antioxidant therapy that does not have specific marketing authorisation for use in IPF, and it is readily available in health-food shops and over the internet. It can be used alone or, as has been done in the past, in combination with prednisolone and azathioprine (referred to as triple therapy). The ATS/ERS 2011 consensus guidance[10] suggests that NAC (alone or in triple therapy) is not used in the majority of patients but it may be a reasonable choice in a minority. The 2013 NICE guideline[31] suggests that it may be used alone after discussion of the risks of the treatment, and that if used in combination with azathioprine and prednisolone, these latter two treatments should be gradually withdrawn.

Pirfenidone is an orally bioavailable synthetic molecule with antifibrotic and anti-inflammatory properties. The mechanism of action at the molecular level is not fully understood; however, pirfenidone has marketing authorisation for use in mild to moderate IPF in Europe. Pirfenidone was not included in the scope of the NICE 2013 clinical guideline;[31] however, it has been assessed by NICE under the technology appraisals programme. Final guidance to the NHS issued in 2013[23] states that pirfenidone is recommended as an option where an individual patient has a FVC between 50% and 80% predicted. This is the case only if the manufacturer of pirfenidone provides the treatment at a previously agreed (undisclosed) discount price.

Symptomatic treatments

Prednisolone is a corticosteroid which can be used to reduce inflammation. Most widely used in the past before the nature of IPF was fully understood and the diagnosis more specific, prednisolone is now used predominantly for symptom control, including for acute exacerbations.[10]

Pulmonary rehabilitation is conventionally offered as a package of supervised exercise and education over a 6- to 8-week period by a multidisciplinary team. The ATS/ERS 2011 consensus guidance[10] suggests that this is used in the majority of patients but this may not be a reasonable choice for some. In the UK, the clinical guideline from NICE[31] recommends that individuals should be assessed for pulmonary rehabilitation where appropriate.

Most experience with pulmonary rehabilitation comes from treating COPD, although the adoption of pulmonary rehabilitation in the UK for IPF is becoming more widespread. The process of rehabilitation commences with a detailed assessment; it is most likely that a detailed study of lung function will have already characterised the patient. The rehabilitation assessment should include an evaluation of health-related quality of life (HRQoL), levels of anxiety and depression, and exercise capacity as a minimum. Additional measures that might be considered could be measures of peripheral muscle strength, self-efficacy and physical activity, but these are not widely reported for IPF. HRQoL measures are those commonly employed for individuals with COPD, for example the St George's Hospital Respiratory Questionnaire (SGRQ)[51] or the chronic respiratory disease questionnaire.[52] The exercise test is usually conducted outside the laboratory and in the UK is commonly the 6MWT[53] or the incremental shuttle walking test.[54] During and immediately after an exercise test it is important to record levels of oxygen saturation, and, in discussion with the patient, make a decision about the use of supplemental oxygen while exercising. Repeat testing with oxygen might be expected to correct the desaturation to some degree and/or improve the level of perceived breathlessness.

The usual format of rehabilitation is one of educational and supervised exercise. The groups are not specifically designed for individuals with IPF and, therefore, the education sessions may not all be entirely appropriate. Generally, the educational topics include medicines management, disease pathology, chest clearance techniques, breathing control, healthy eating, travelling with a lung disease, inhaler techniques, relaxation, stress management and energy conservation. The exercise programme is a combination of strength and aerobic exercises, prescribed and progressed on an individual basis. Bouts of purposeful exercise should also be completed at home, to accumulate five sessions per week of aerobic activity. On completion of a course of rehabilitation, it is recommended that a repeat assessment be conducted to reinforce the benefits of rehabilitation.

Maintaining the benefit of rehabilitation is challenging; however, there are opportunities to continue with regular supervised exercise in the community with exercise specialists trained to manage participants with chronic respiratory disease.

Oxygen therapy can either be used as a long-term therapy or be taken as a supplement after exertion. The ATS/ERS 2011 consensus guideline[10] suggests that long-term oxygen therapy is used in patients with IPF and clinically significant resting hypoxaemia, and the NICE 2013 guideline[31] also refers to the use of oxygen for breathlessness on exertion as discussed above.

Palliative care/BSC should be considered from the outset as an adjunct to disease-focused care. This may include corticosteroids, oxygen therapy, treatments for other symptoms such as cough, advanced directives and end-of-life care issues including hospice care.

Treatments for specific symptoms
Sildenafil is a phosphodiesterase-5 inhibitor with marketing authorisation for use in pulmonary hypertension and it has been used as a treatment for IPF. There are currently no recommendations from the ATS/ERS consensus committee of the 2011 guidance[10] and in the NICE guideline sildenafil was not recommended as a treatment for modifying IPF owing to uncertainty around both benefits and harms.[31]

Bosentan is a dual endothelin receptor A and B antagonist that has marketing authorisation for pulmonary hypertension. Endothelin is a vasoconstrictor and growth factor that is involved in the pathogenesis of pulmonary hypertension. The 2011 ATS/ERS[10] consensus guidance suggests that these are not used based on the potential risks and costs of therapy. The NICE clinical guideline concurs with this.[31]

Thalidomide is a potent immunomodulatory, anti-inflammatory and antiangiogenic drug, and may be a potential therapy in IPF for cough.[55] Thalidomide does not have a UK marketing authorisation for IPF; however, the NICE clinical guideline[31] states that thalidomide may be considered in a patient with intractable cough.

Treatments undergoing assessment in UK participants

BIBF 1120 or *nintedanib* is an intracellular inhibitor of tyrosine kinases and is currently under evaluation in phase III randomised controlled trials (RCTs) as a potential therapy for IPF. It has received orphan status from the European Medicines Agency in 2013 (orphan status provides some favourable marketing incentives to manufacturers).[56]

Overall aims and objectives of assessment

This project will evaluate the clinical effectiveness and cost-effectiveness of the different treatment strategies used within the NHS for IPF. It will review systematically the evidence on those interventions that are currently available and used to treat IPF. It will construct a new economic model relevant to the UK setting to estimate the cost-effectiveness of the different treatments. Deficiencies in current knowledge will be identified and recommendations for future research generated.

Chapter 2 Methods for the systematic review of clinical effectiveness and cost-effectiveness

The a priori methods for systematically reviewing the evidence of clinical effectiveness and cost-effectiveness are described in the research protocol, which was sent to our expert advisory group for comment. Although helpful comments were received relating to the general content of the research protocol, there were none that identified specific problems with the methodology of the review. The methods outlined in the protocol are briefly summarised below.

Identification of studies

A comprehensive search strategy was developed, tested and refined by an experienced information scientist. Separate searches were conducted to identify studies of clinical effectiveness, cost-effectiveness, QoL and epidemiology. Sources of information and search terms are provided in *Appendix 1*. The most recent searches were undertaken in July 2013.

Literature was sourced from 11 electronic databases, the bibliographies of articles, and grey literature sources, and our expert advisory group were contacted to identify any additional studies. All databases were searched from inception with no language restrictions. The following electronic databases were searched: MEDLINE (Ovid); MEDLINE In Process & Other Non-Indexed Citations; EMBASE; The Cochrane Library including Central Register of Controlled Trials and Cochrane Database of Systematic Reviews; Centre for Reviews and Dissemination (CRD) including Health Technology Assessment Database, Database of Abstracts of Reviews of Effects and NHS Economic Evaluation Database; Web of Science (Science Citation Index Expanded) and Conference Proceedings Citation Index; and Bioscience Information Service Previews. A comprehensive database of relevant published and unpublished articles was constructed using Reference Manager software (Thomson ResearchSoft, San Francisco, CA, USA). Research-in-progress databases (UK Clinical Research Network website, Current Controlled Trials and Clinical trials.gov, World Health Organization-International Clinical Trials Registry Platform) were searched for any ongoing studies of relevance.

In addition, professional society websites and conferences were searched for recent abstracts and ongoing studies (see *Appendix 1*).

Study selection and data extraction

Studies were selected for inclusion in the systematic review of clinical effectiveness through a two-stage process using predefined and explicit criteria (see *Inclusion and exclusion criteria*). Titles and abstracts from the literature search results were independently screened by two reviewers to identify all citations that possibly met the inclusion criteria. Full papers of relevant studies were retrieved and assessed by one reviewer and checked by a second reviewer using a standardised eligibility form. As far as possible, full papers or abstracts describing the same study were linked together, with the article reporting key outcomes designated as the primary publication.

A two-stage approach was used to establish the relevance of each individual treatment for inclusion in the evidence synthesis. The advisory group individually and independently commented on the relevance of each treatment. The outcomes from this exercise were tabulated and those with complete consensus were either included or excluded as appropriate. Where consensus on any particular intervention was not reached, these were sent to the advisory group for a second time. At this stage, either consensus was

confirmed or a decision was taken to include treatments where at least one expert had indicated that the treatment was used in their clinical practice.

Titles and abstracts identified by the search strategy for the systematic review of cost-effectiveness and HRQoL were assessed for potential eligibility by two reviewers using predetermined inclusion criteria (see *Inclusion and exclusion criteria*). Full papers were formally assessed for inclusion by one health economist with respect to their potential relevance to the research question and this was checked by one reviewer.

Data were extracted by one reviewer using a standard data extraction form (see *Appendix 2*) and checked by a second reviewer. At each stage, any disagreements between reviewers were resolved by consensus or, if necessary, by arbitration by a third reviewer.

Critical appraisal strategy

The methodological quality and the quality of reporting of the included clinical effectiveness studies were assessed using criteria based on those recommended by the CRD[57] and the Cochrane Collaboration[58] (see *Appendix 2*). The risk of bias within each study was summarised according to the risk of selection bias. Quality criteria were applied by one reviewer and checked by a second reviewer, with any differences in opinion resolved by consensus or by arbitration by a third reviewer.

Quality assessment for the systematic review of cost-effectiveness was based on a checklist for economic evaluation publications[59] and guidelines for good practice in decision-analytic modelling in health technology assessment.[60]

Inclusion and exclusion criteria

Participants

- People with a confirmed diagnosis of IPF.

As there have been changes in the diagnostic criteria for IPF, particular attention was paid to the inclusion criteria used by studies to ensure that the results were not influenced by mixed populations with differing prognoses. Studies of mixed populations were included only if the study reported outcomes for those with IPF separately.

Where the definition of IPF was uncertain, this was checked by our advisory group. Any studies remaining uncertain would have been included but discussed separately; no such instances occurred.

The presence of pulmonary hypertension in participants in included studies was not a reason for exclusion. Any included studies with known participants with pulmonary hypertension are discussed as appropriate in the results.

Interventions

- Any available and currently used (in the NHS) interventions which aim to manage symptoms or modify IPF were eligible.

As described above, clinical experts and the expert advisory group for the review were asked to identify the current treatments used in the UK to ensure that the key treatments in use were included in the review. For discussion of the included interventions see *Chapter 3*.

Comparators

- Any of the included interventions.
- BSC.
- Placebo interventions.

Outcomes

- Measures of survival.
- Measures of symptoms (breathlessness, cough).
- QoL/HRQoL.
- Lung function/capacity.
- Exercise performance.
- Adverse events.
- Costs and cost-effectiveness.

Patient-assessed subjective outcome measures were included if assessed by validated tools (for descriptions see *Chapter 3* and *Appendix 3*).

A variety of surrogate end points have been used in trials of interventions for IPF. *Appendix 3* describes the key outcomes used, including evidence of the relationship between these and the final patient outcome.

Design

- RCTs.
- Where no RCTs were identified for a particular intervention, controlled clinical trials (CCTs) with a concurrent control group were eligible.
- Economic evaluations (i.e. costs and consequences), including cost-effectiveness, cost–utility or cost–benefit analyses.

Studies published as abstracts or conference presentations were included only if sufficient details were presented to allow an appraisal of the methodology and the assessment of results to be undertaken.

Systematic reviews identified by the search were used as a source for identifying primary studies and are summarised in *Chapter 3* (see *Existing systematic reviews*).

Method of data synthesis

Studies of clinical effectiveness and cost-effectiveness were synthesised through a narrative review with tabulation of results of included studies.

Where data were of sufficient quality and homogeneity, meta-analyses of the clinical effectiveness studies were performed to estimate the mean difference (for continuous data) and risk ratio (RR) (for dichotomous data) with 95% CIs. The random-effects method was used where statistical heterogeneity was observed. The standardised mean difference was used to combine per cent predicted FVC/VC and mean change in FVC/vital capacity (VC) in a meta-analysis. The standardised mean difference converts all outcomes to a common scale. It is the difference in means between the treatment arms divided by the pooled standard deviation (SD) of outcomes from all participants in each trial. Meta-analysis was performed by using Cochrane Review Manager 5 (RevMan, The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) and statistical heterogeneity was assessed using chi-squared test and the I^2 statistic.[58] Where SDs were not presented in the published papers, these were calculated from the available statistics (CIs, standard errors or *p*-values).

In addition, a network meta-analysis (NMA) was undertaken to allow ranking of the effectiveness of the range of treatments being evaluated. NMA is an extension of traditional, pairwise meta-analysis where a statistical analysis of the network of trial evidence is used to produce comparable estimates of benefits and harms for a range of treatments. The NMA was conducted using current guidance on good practice (for further details *see Chapter 3, Network meta-analysis*).

The methods for the economic model are described in *Chapter 4* (see *Methods for economic analysis*).

Chapter 3 Clinical effectiveness

Research identified and included

Searches identified a total of 905 references after deduplication, and full texts of 64 references were retrieved after screening titles and abstracts. Initial screening of the full texts identified 47 publications of possible relevance. The interventions and comparators used in these 47 studies were tabulated and sent to the advisory group to determine their relevance. The final list of included interventions can be seen in *Table 3*. One new treatment, BIBF 1120 (nintedanib), was included as it will possibly be used in future practice. After this exercise, 13 RCTs and one CCT[61] were included (two of the RCTs were reported in one publication).[62]

One non-English (Polish) publication was considered to be eligible for inclusion.[61] The publication contained an English translation of the abstract and results tables. A translation of the methods was requested from the author; however, no response was received. The methods were translated using Google Translate (http://translate.google.com) and results were extracted from the tables but not the text to avoid error.

The number of references excluded at each stage of the systematic review is shown in *Figure 1*. References that were retrieved as full papers but subsequently excluded are listed in *Appendix 4*, with reasons for exclusion. Seventeen references were excluded owing to participants (eight studies), study design (seven studies) or irrelevant outcomes (three studies), with one study being excluded for more than one reason. Thirty-one studies were eligible in the first full-paper screen but were then excluded as the interventions were not considered relevant by the advisory group. The level of agreement between reviewers for screening was high, although this was not formally assessed.

Searches identified 25 relevant ongoing studies, a summary of which can be seen later in this chapter (see *Ongoing randomised controlled trials*). One conference abstract was also identified but the abstract did not present sufficient information for the study to be formally assessed for inclusion. This study was a CCT of a 12-week pulmonary rehabilitation compared with a control group (undefined).[64] No full publication for this study has been identified on updated searches.

Four RCTs evaluated the use of pirfenidone,[62,65,66] three the use of NAC (alone or in combination),[67–69] one azathioprine,[70] one BIBF 1120,[71] one sildenafil (in severe IPF),[72] one thalidomide,[55] one a pulmonary rehabilitation programme,[73] and one a disease management programme.[74] In addition, one CCT of pulmonary rehabilitation was included.[61] No studies of bosentan, cyclophosphamide and prednisolone, methyprednisolone (alone or in combination with prednisolone) or co-trimoxazole (Septrin®, Aspen) were included because studies of these interventions had ineligible comparator treatments. No studies of palliative care interventions were identified that met the inclusion criteria.

Description of included studies

Eleven of the included RCTs investigated the use of pharmacological interventions for IPF. One was a single-centre randomised crossover trial,[55] and the rest were multicentre trials, with the number of centres ranging from two[70] to 110.[62] The majority of centres were in either the USA or European countries, although some included countries such as Mexico and Republic of Korea, and three studies were undertaken solely in Japan.[65,66,69] Sample sizes ranged from 24[55] to 435,[62] although all except two trials[55,70] had sample sizes of 100 or more participants. Nine of the 11 trials investigating pharmaceutical agents received some funding from the drug manufacturer. Across the 11 trials, participants would likely be

TABLE 3 Status of interventions identified at the initial screening stage

Treatment	Decision
Ambrisentan	Exclude
Azathioprine and prednisolone[a]	**Include**
BIBF 1120 (nintedanib)	**Include as possible future use**
Bosentan	**Include if pulmonary hypertension present**
Bromhexine hydrochloride	Exclude
Colchicine	Exclude
Colchicine and prednisolone	Exclude
Colchicine and d-penicillamine	Exclude
Colchicine and cyclophosphamide	Exclude
Cyclophosphamide and prednisolone	**Include if used for acute exacerbations**
Ciclosporin	Exclude
Disease management programmes	**Include**
Etanercept	Exclude
Everolimus	Exclude
Iloprost	Exclude
Imatinib	Exclude
Interferon-γ–1b	Exclude
Interferon-γ–1b and prednisolone	Exclude
Macitentan	Exclude
Methylprednisolone	**Include if used for acute exacerbations**
NAC	**Include**
NAC (inhaled)	**Include**
NAC, azathioprine, prednisolone	**Include**
Pirfenidone	**Include**
Pulmonary rehabilitation	**Include**
Prednisone (prednisolone)	**Include**
Prednisone and d-penicillamine	Exclude
Prednisone, d-penicillamine, colchicine	Exclude
Prednisone and warfarin	Exclude
Prednisone and methylprednisolone	**Include if used for acute exacerbations**
Septrin	**Include**
Sildenafil	**Include if pulmonary hypertension present or in severe IPF**
Thalidomide	**Include if for cough**

a Although this is no longer indicated in IPF, it is still in use.
Bold used to denote included treatments.

```
                    ┌─────────────────────────┐
                    │  Identified on searching │
                    │  (after duplicate removal)│
                    │        (n = 905)          │
                    └─────────────────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐        ┌──────────────┐
          │ Titles and abstracts     │───────▶│   Excluded   │
          │      inspected           │        │  (n = 841)   │
          └─────────────────────────┘        └──────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐
          │  Full papers retrieved   │
          │        (n = 64)          │
          └─────────────────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐        ┌──────────────┐
          │  Full papers inspected   │───────▶│   Excluded   │
          │                          │        │  (n = 17)    │
          └─────────────────────────┘        └──────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐
          │   Potentially eligible   │
          │        (n = 47)          │
          └─────────────────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐        ┌──────────────┐
          │ Interventions sent to    │───────▶│   Excluded   │
          │    advisory group        │        │  (n = 31)    │
          └─────────────────────────┘        └──────────────┘
                                 │
                                 ▼
          ┌─────────────────────────┐
          │ Included (n = 15 publications)ᵃ│
          │        (14 trials)       │
          │    (n = 1 protocol)ᵇ     │
          └─────────────────────────┘
```
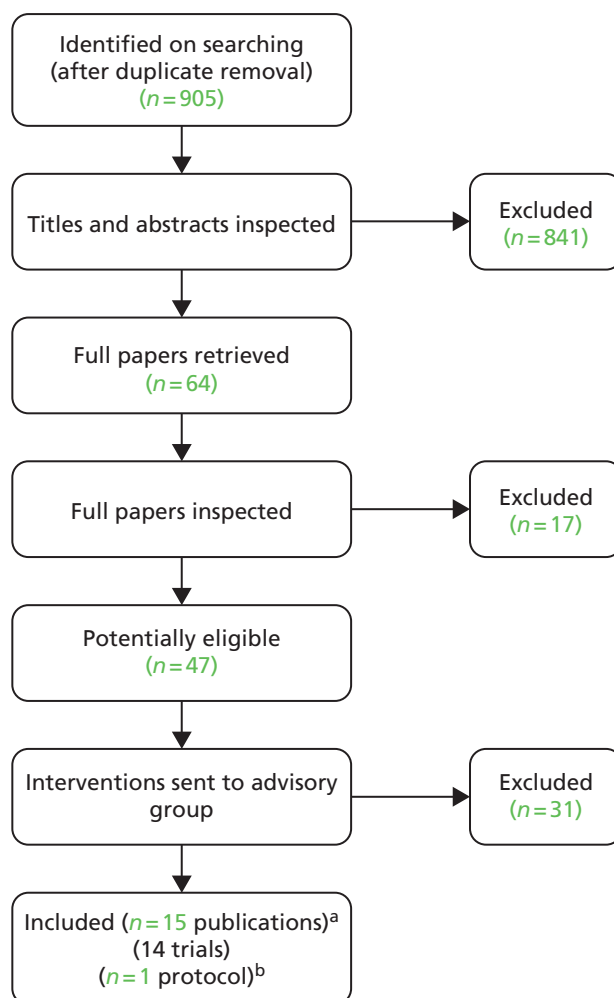
FIGURE 1 Flow chart for the identification of studies. a, two RCTs were reported in one publication; two further trials each had two linked publication; b, one trial was published as a protocol only:[63] see *Ongoing randomised controlled trials*.

classified as mild to moderate in 10, with one study looking at participants in the more severe stages of IPF.[72] In the three studies investigating non-pharmacological interventions, one RCT was undertaken in Japan and investigated a pulmonary rehabilitation programme,[73] one CCT was undertaken in Poland and investigated inspiratory muscle training in addition to pulmonary rehabilitation,[61] and the other RCT investigated a disease management programme in the USA.[74] These trials were single-centre studies and funding was from public or charitable funds where data were available. The sample size was 30 in both of the pulmonary rehabilitation programme studies,[61,73] and 21 in the disease management programme trial.[74] The populations in these trials were at the more moderate to severe end of the spectrum of IPF. Further detail of the characteristics of these studies and their participants can be found in subsections below and in *Table 4*, which summarises the key attributes of the fourteen included studies, and *Table 5*, which summarises the key participant characteristics.

The following sections are subdivided into pharmacological and non-pharmacological agents, and individual interventions are discussed in alphabetical order within severity categories.

TABLE 4 Summary of included studies

Study	Intervention details	Key eligibility criteria[a]	Outcomes
Pharmacological agents			
Azathioprine			
Raghu _et al._ 1991[70] _Country:_ USA _Design:_ RCT _Number of centres:_ two _Funding:_ Grant from Virginia Mason Research Centre, Seattle, WA, USA	1. Prednisone and placebo, $n = 13$ 2. Prednisone and azathioprine, $n = 14$ _Dose details:_ prednisolone decreased from 1.5 mg/kg/day (not to exceed 100 mg/day) for 2 weeks, decrease of 20 mg/day until a dose of 40 mg/day was reached for 2 weeks, decrease of 5 or 10 mg/day every 2 weeks according to patient tolerance. Maintenance dose of ≤ 20 mg/day Azathioprine 3 mg/kg/day (not to exceed 200 mg/day) to the nearest 25-mg dose increment for the duration _Duration of treatment:_ assume 112 months	_Inclusion criteria:_ a diagnosis of IPF supported by lung biopsy. Previously untreated and available for routine follow-up. Fulfilled criteria for progressive clinical disease [one or more of (1) progressive dyspnoea from day of onset, (2) progressive roentgenographic parenchymal abnormality, (3) $\geq 10\%$ decrease in FVC or TLC compared with previous values, (4) $\geq 20\%$ reduction in DL_{CO} compared with previous values] _Excluded:_ collagen vascular disease, hypersensitivity pneumonitis, pneumoconiosis, drug-induced diffuse pulmonary injury, or irradiation fibrosis	_Primary outcomes:_ not stated as primary or secondary – measurable change in lung function (FVC, DL_{CO}, $PA - aO_2$) at 12 months; survival _Length of follow-up:_ 12 months
BIBF 1120			
Richeldi _et al._ 2011[71] _Country:_ 25 countries including Italy, Mexico, Germany, USA, Republic of Korea, UK, France _Design:_ RCT (dose-finding phase II study) _Number of centres:_ 92 _Funding:_ supported by Boehringer Ingelheim	1. BIBF 1120 50 mg/day, $n = 86$ 2. BIBF 1120 50 mg twice per day (100 mg/day), $n = 86$ 3. BIBF 1120 100 mg twice per day (200 mg/day), $n = 86$ 4. BIBF 1120 150 mg twice per day (300 mg/day), $n = 85$ 5. Placebo, $n = 85$ _Dose details:_ a group-wise dose escalation was used with stepwise increases in dose for serial cohorts _Duration of treatment:_ 52 weeks	_Inclusion criteria:_ ≥ 40 years, IPF consistent with the ATS/ERS 2000 criteria, diagnosis < 5 years, HRCT < 1 year, FVC that was $\geq 50\%$ predicted, DL_{CO} 30–79% predicted, and PaO_2 when breathing ambient air ≥ 55 mmHg at altitudes up to 1500 m, or ≥ 50 mmHg at altitudes above 1500 m _Excluded:_ medical conditions or concomitant medications that might interfere with the performance of the study, other diseases that might interfere with testing procedures (stated), continuous oxygen supplementation, predisposition to bleeding or thrombosis, concomitant anticoagulation medication, elevated liver enzymes, likelihood of lung transplantation during the study or life expectancy < 2.5 years for a disease other than IPF	_Primary outcomes:_ annual rate of decline in FVC _Secondary outcomes:_ per cent predicted FVC; DL_{CO}; SpO_2; TLC; 6MWT, SGRQ, decrease in FVC of > 10% or > 200 ml; SpO_2 decrease of > 4%; acute exacerbations; survival; death from a respiratory cause; adverse events _Length of follow-up:_ 54 weeks

TABLE 4 Summary of included studies (*continued*)

Study	Intervention details	Key eligibility criteria[a]	Outcomes
NAC (alone or in combination)			
Demedts *et al.* 2005[67] *Country:* Belgium, France, Germany, Italy, Spain, the Netherlands *Design:* RCT *Number of centres:* 36 *Funding:* sponsored by the Zambon group	1. NAC, prednisolone, azathioprine, $n = 92$ (80 analysed) 2. Placebo, prednisolone, azathioprine, $n = 90$ (75 analysed) *Dose details:* NAC at 600 mg three times per day. Azathioprine at 2 mg/kg per day. Prednisone dose decreases from 0.5 mg/kg body weight per day; 0.4 mg/kg per day at month 2; 0.3 mg/kg at month 3; until 10 mg per day in months 4–6 and then maintained until month 12 *Duration of treatment:* not stated, assume 12 months	*Inclusion criteria:* 18–75 years, histological or radiologic pattern typical of UIP. HRCT suggestive or consistent with a probable diagnosis of UIP. Open or thoracoscopic lung biopsy in those aged < 50 years. In the absence of lung biopsy, a transbronchial biopsy was advocated to exclude alternative diagnoses. Duration > 3 months, bibasilar inspiratory crackles, dyspnoea score at least 2 on a scale of 0 (minimum) to 20 (maximum), VC ≤ 80% predicted or TLC < 90% predicted, and single-breath DL_{CO} < 80% predicted *Excluded:* contraindication/intolerance to study treatments. Prednisone at least 0.5 mg/kg/day or azathioprine at least 2 mg/kg/day during the month before inclusion, treatment with NAC of > 600 mg/day for > 3 months in the previous 3 years. Concomitant or pre-existing diseases, or treatments that interfere with IPF	*Primary outcomes:* absolute changes in VC and DL_{CO} at 12 months *Secondary outcomes:* per cent predicted VC, per cent predicted DL_{CO}, alveolar volume change and per cent predicted, CRP score, dyspnoea score, maximum exercise indexes, HRCT outcomes, SGRQ, adverse events, withdrawals, and mortality *Length of follow-up:* 12 months
Raghu *et al.* IPFCRN 2012[68] *Country:* USA *Design:* RCT (PANTHER study) *Number of centres:* 25 *Funding:* grants from the NHLBI; the Cowlin Family fund. NAC and placebo donated by Zambon	1. NAC and placebo (data not presented in article as 'ongoing' data collection), $n = 81$ 2. NAC/prednisolone/azathioprine, $n = 77$ 3. Placebo, $n = 78$ *Dose details:* NAC at 600 mg orally three times per day. Prednisone decreased from 0.5 mg/kg of ideal body weight to 0.15 mg/kg during 25 weeks. Azathioprine (maximum 150 mg/day) based on ideal weight, concurrent use of allopurinol, and thiopurine methyltransferase activity *Duration of treatment:* up to 60 weeks	*Inclusion criteria:* IPF aged between 35–85 years with mild to moderate lung-function impairment (FVC ≥ 50% and DL_{CO} ≥ 30% of predicted values), modified criteria of the ATS/ERS (2011) for diagnosis of IPF, received diagnosis on the basis of HRCT or biopsy ≤ 48 months before enrolment *Excluded:* FEV_1–FVC ratio < 0.65; PaO_2 on room air < 55 mmHg; residual volume > 120% predicted; evidence of active infection; significant bronchodilator response; post-bronchodilator FVC differing by > 11%; listed for lung transplantation; history of cardiac disease or hypertension; HIV or hepatitis C; liver cirrhosis and chronic active hepatitis	*Primary outcomes:* change in FVC at 60 weeks *Secondary outcomes:* rate of death, time until death, frequency of acute exacerbation, frequency of maintained FVC response, time to disease progression, clinical and physiological measures including DL_{CO}, 6MWT, CPI, UCSDSBQ, SGRQ, SF-36, EQ-5D. Adverse events *Length of follow-up:* 60 weeks in the planned analysis. The study was stopped early. The mean follow-up was 32 weeks

TABLE 4 Summary of included studies (*continued*)

Study	Intervention details	Key eligibility criteria[a]	Outcomes
Homma *et al.* 2012[69] *Country:* Japan *Design:* RCT *Number of centres:* 27 *Funding:* grant from Ministry of Health, Labour and Welfare	1. NAC inhaled, $n = 51$ (38 analysed) 2. Control, $n = 49$ (38 analysed) *Dose details:* NAC inhalation of 352.4 mg diluted with saline to a total volume of 4 ml twice a day, using microair nebulisers and vibration mesh technology Control: states 'no treatment or placebo' *Duration of treatment:* 48 weeks	*Inclusion criteria:* well-defined IPF (by ATS/ERS guidance and Japanese criteria). Histological evidence of UIP was not mandatory, but HRCT evidence was required as were presence of other typical clinical features. Aged between 50 and 79 years, severity of disease classified as stage I (partial O_2 concentration ≥ 80 mmHg at rest), or stage II (partial O_2 concentration 70–80 mmHg at rest), a lowest arterial oxygen saturation of $> 90\%$ during the 6MWT *Excluded:* an improvement in symptoms during the preceding 3 months; use of NAC, immunosuppressive agents, oral prednisolone or pirfenidone; clinical suspicion of IIP other than IPF	*Primary outcomes:* absolute change in FVC at 48 weeks *Secondary outcomes:* changes in lowest arterial O_2 saturation, 6MWT distance, PFT parameters (VC, per cent predicted VC, TLC, per cent predicted TLC, DL_{CO}, predicted DL_{CO}), serum markers of pneumocyte injury; disease progression as determined by HRCT; subjective changes in symptoms such as dyspnoea, adverse events *Length of follow-up:* 48 weeks
Pirfenidone			
Noble *et al.* 2011[62] Capacity study 006 *Country:* Australia, Belgium, Canada, France, Germany, Ireland, Italy, Mexico, Poland, Spain, Switzerland, UK, USA *Design:* RCT *Number of centres:* 110 centres *Funding:* InterMune	1. Pirfenidone 2403 mg/day, $n = 171$ 2. Placebo, $n = 173$ *Dose details:* the dose was increased to the full dose over 2 weeks *Duration of treatment:* 72 weeks	*Inclusion criteria:* aged 40–80 years, diagnosis of IPF in the previous 48 months, no evidence of improvement in measures of disease severity over the preceding year. Predicted FVC of $\geq 50\%$, predicted DL_{CO} of $\geq 35\%$, either predicted FVC or $DL_{CO} \leq 90\%$, 6MWT distance of ≥ 150 m *Excluded:* obstructive airway disease, connective tissue disease, alternative explanation for ILD, those on a waiting list for a lung transplant	*Primary outcomes:* change in per cent predicted FVC *Secondary outcomes:* categorical FVC (5-point scale), PFS, worsening IPF, dyspnoea, 6MWT distance, worst peripheral oxygen saturation (SpO_2) during the 6MWT, per cent predicted DL_{CO}, fibrosis, mortality *Length of follow-up:* 72 weeks from the date the last patient was enrolled
Noble *et al.* 2011[62] Capacity study 004 *Country:* Australia, Belgium, Canada, France, Germany, Ireland, Italy, Mexico, Poland, Spain, Switzerland, UK, USA *Design:* RCT *Number of centres:* 110 centres *Funding:* InterMune	1. Pirfenidone 2403 mg/day, $n = 174$ 2. Pirfenidone 1197 mg/day, $n = 87$ 3. Placebo, $n = 174$ *Dose details:* the dose was increased to the full dose over 2 weeks *Duration of treatment:* 72 weeks	*Inclusion criteria:* aged 40–80 years, diagnosis of IPF in the previous 48 months, no evidence of improvement in measures of disease severity over the preceding year. Predicted FVC of $\geq 50\%$, predicted DL_{CO} of $\geq 35\%$, either predicted FVC or $DL_{CO} \leq 90\%$, 6MWT distance of ≥ 150 m *Excluded:* obstructive airway disease, connective tissue disease, alternative explanation for ILD, those on a waiting list for a lung transplant	*Primary outcomes:* change in per cent predicted FVC *Secondary outcomes:* categorical FVC (5-point scale), PFS, worsening IPF, dyspnoea, 6MWT distance, worst peripheral oxygen saturation (SpO_2) during the 6MWT, per cent predicted DL_{CO}, mortality *Length of follow-up:* 72 weeks from the date the last patient was enrolled

TABLE 4 Summary of included studies (*continued*)

Study	Intervention details	Key eligibility criteria[a]	Outcomes
Taniguchi *et al.* 2010[65] *Country:* Japan *Design:* RCT *Number of centres:* 73 *Funding:* public sector grants. Drug and placebo from Shionogi & Co., Ltd	*1.* Pirfenidone 1800 mg/day, $n = 108$ *2.* Pirfenidone 1200 mg/day, $n = 55$ *3.* Placebo, $n = 104$ *Dose details:* doses increased in a stepwise manner over 4 weeks *Duration of treatment:* 52 weeks	*Inclusion criteria:* aged 20–75 years, IPF diagnosed using ATS/ERS and clinical diagnostic criteria guidelines for IIP in Japan. Independently evaluated HRCT scans. Probably UIP by surgical lung biopsy. Arterial oxygen saturation criteria of (1) oxygen desaturation of $\geq 5\%$ difference between resting SpO_2 and the lowest SpO_2 during a 6MET; and (2) the lowest SpO_2 during the 6MET of $\geq 85\%$ while breathing air *Excluded:* decrease in symptoms (preceding 6 months); use of immune-suppressants and/or oral corticosteroids at a dose of > 10 mg/day^{-1} (preceding 3 months); clinical features of IIP other than IPF; coexisting lung conditions (stated) or severe respiratory infection	*Primary outcomes:* change in VC to week 52 *Secondary outcomes:* PFS time, change in lowest SpO_2 during the 6MET. PaO_2, PA – aO_2, TLC and DL_{CO}, acute exacerbation, markers of interstitial pneumonias, symptoms *Length of follow-up:* 52 weeks
Azuma *et al.* 2005[66] *Country:* Japan *Design:* RCT *Number of centres:* 25 *Funding:* Shionogi & Co., Ltd	*1.* Pirfenidone 1800 mg/day, $n = 73$ *2.* Placebo, $n = 36$ *Dose details:* doses increased in a stepwise manner over 1 week *Duration of treatment:* 9 months	*Inclusion criteria:* histological evidence of UIP not mandatory, HRCT evidence of definite or probable UIP required. Presence of bibasilar inspiratory crackles, abnormal PFTs, and increased serum levels of damaged-pneumocyte markers. Aged 20–75 years with $PaO_2 \geq 70$ mmHg at rest and demonstrated SpO_2 of $\leq 90\%$ during exertion while breathing air, 1 month before enrolment *Excluded*: a decrease in symptoms (preceding 6 months), use of immunosuppressive and/or oral prednisolone > 10 mg/day (preceding 3 months), clinical suspicion of IIP other than IPF, coexisting lung conditions (defined); uncontrolled diabetes, comorbid conditions	*Primary outcomes:* change in the lowest SpO_2 during the 6MET *Secondary outcomes:* resting PFTs while breathing air (VC, TLC, DL_{CO} PaO_2), disease progression by HRCT patterns, acute exacerbation, serum markers of pneumocyte damage, QoL *Length of follow-up:* minimum of 9 months

TABLE 4 Summary of included studies (*continued*)

Study	Intervention details	Key eligibility criteria[a]	Outcomes
Thalidomide			
Horton *et al.* 2012[55,75] *Country:* USA *Design:* randomised crossover trial *Number of centres:* one *Funding:* Celgene Corporation	1. Thalidomide, n = 23 2. Placebo, n = 23 *Dose details:* 50 mg oral daily, increased to 100 mg if no improvement in cough occurred after 2 weeks (21 of 22 receiving thalidomide and 23 of 23 receiving placebo) *Duration of treatment:* 12 weeks each treatment with a 2-week washout period between treatments	*Inclusion criteria:* aged > 50 years, clinical history consistent with IPF (symptom duration $\geq$ 3 months and $\leq$ 5 years) and chronic cough (defined as > 8 weeks' duration, that adversely affected QoL and not due to other identifiable causes). HRCT scans consistent with IPF or surgical lung biopsy demonstrating UIP, FVC between 40% and 90% predicted, TLC between 40% and 80% predicted, and DL_{CO} between 30% and 90% predicted *Exclusion criteria:* pregnancy, female with childbearing potential, toxic or environmental exposure to respiratory irritants, collagen vascular disease, airflow obstruction, active narcotic antitussive use, peripheral vascular disease or neuropathy, inability to give informed consent, allergy or intolerance to thalidomide, life expectancy < 6 months (opinion of investigators)	*Primary outcomes:* cough-specific quality of life *Secondary outcomes:* cough, respiratory QoL *Method of assessing outcome:* cough-specific QoL measured by CQLQ. Cough measured by a 10-cm VAS. Respiratory QoL measured by SGRQ *Length of follow-up:* 12 weeks
Sildenafil (severe IPF)			
Zisman and colleagues IPFCRN 2010[72] *Country:* USA *Design:* RCT *Number of centres:* 14 *Funding:* NHLBI; the Cowlin Fund (Chicago Community trust); Pfizer; Masimo	1. Sildenafil, *n* = 89 2. Placebo, *n* = 91 *Dose details :* sildenafil, oral, 20 mg, three times daily *Duration of treatment:* 12 weeks	*Inclusion criteria:* diagnosis of IPF as defined by the 2000 ATS/ERS consensus criteria, in an advanced stage (a DL_{CO} of < 35% predicted) *Excluded:* 6MWT distance < 50 m; a difference > 15% in the 6MWT distance between two pre-randomisation walks; emphysema extent greater fibrotic change; treatment with medications containing nitrates; aortic- or idiopathic hypertrophic subaortic stenosis; within 30 days of screening: pulmonary rehabilitation; initiation or change in dose of treatment for IPF; treatments for pulmonary hypertension with prostaglandins, endothelin-1 antagonists, or other phosphodiesterase inhibitors; a resting oxygen saturation of < 92% while breathing 6 litres of supplemental oxygen; waiting list for lung transplantation	*Primary outcomes:* presence or absence of an improvement of at least 20% in the 6MWT distance at 12 weeks *Secondary outcomes:* changes in the 6MWT distance, degree of dyspnoea, QoL, FVC, DL_{CO}, arterial partial pressure of oxygen and arterial oxygen saturation, and the alveolar–arterial oxygen gradient while breathing ambient air, adverse events, hospitalisations, death *Length of follow-up:* 12 weeks

TABLE 4 Summary of included studies (*continued*)

Study	Intervention details	Key eligibility criteria[a]	Outcomes
Non-pharmacological interventions			
Disease management programme/pulmonary rehabilitation			
Lindell *et al.* 2010[74] *Country:* USA *Design:* RCT *Number of centres:* one *Funding:* Fairbanks-Horix Foundation	1. Program to Reduce IPF Symptoms and Improve Management (PRISIM), *n* = 10 pairs 2. Usual care, *n* = 11 pairs *Description of interventions:* 6-weekly group sessions attended by patients and care partners. Each session 2 hours. Also group exercises Usual care participants were seen by members of the clinical care team at every 3–6 months *Duration of treatment:* 6 weeks	*Inclusion criteria:* aged > 21 years, be able to read and understand English, to be diagnosed with IPF, to have a FVC reflecting moderate IPF (FVC 55–70% predicted) or severe IPF (< 55% predicted). Care partners were required to be aged > 21 years, be able to read and understand English, to live with or care for the patient with IPF Diagnostic criteria for IPF not stated *Excluded:* not stated	*Primary outcomes:* not specified as primary or secondary outcomes. Dyspnoea (UCSDSBQ); anxiety (BAI); depression (Beck Depression Inventory-II); stress (PSS); QoL (SF-36) *Length of follow-up:* unclear
Jastrzebski *et al.* 2008[61] *Country:* Poland *Design:* CCT *Number of centres:* one *Funding:* not translated	1. Inspiratory muscle training, *n* = 16 2. Control, *n* = 14 *Description of intervention:* not translated in full. Appears that both groups received pulmonary rehabilitation, described as general body conditioning exercises *Duration of treatment:* 12 weeks (two 6-week cycles)	*Inclusion criteria:* diagnosis by ATS/ERS criteria in patients aged > 50 years, and by lung biopsy in patients < 50 years. At least 2 years' duration of disease and in remission Excluded: > 20 mg prednisolone per day, duration of treatment > 2 years, use of home oxygen therapy	*Primary outcomes:* not specified as primary or secondary. Dyspnoea (oxygen cost diagram, BDI). QoL (SF-36), 6MWT (distance, dyspnoea in Borg's scale), maximal inspiratory pressure, lung function tests (IC, TLC, VC, FEV$_1$, DL$_{COSB}$, DL$_{CO}$/VA) *Length of follow-up:* 12 weeks
Nishiyama *et al.* 2008[73] *Country:* Japan *Design:* RCT *Number of centres:* one *Funding:* Japanese Ministry of Health, Labour and Welfare	1. Pulmonary rehabilitation programme (PRP), *n* = 15 (13 analysed) 2. Control, *n* = 15 *Description of intervention:* general programme of pulmonary rehabilitation, not specific to IPF. Twice-weekly programme of exercise and peripheral muscle training. Some educational lectures were also held No details of control group provided *Duration of treatment:* 10-week programme	*Inclusion criteria:* < 75 years, diagnosis using ATS/ERS consensus criteria, stable IPF, shortness of breath on effort, no infection or exacerbations in the previous 3 months *Excluded:* severe comorbid illnesses, collagen vascular diseases, the need for long-term oxygen therapy and previous treatment with corticosteroids or immunosuppressives	*Primary outcomes:* not specified as primary or secondary. Pulmonary function tests (FVC, FEV$_1$, TLC, *P*aO$_2$, *P*aCO$_2$,) DL$_{CO}$, 6MWT; BDI; SGRQ *Length of follow-up:* 10 weeks after the start of the programme

6MET, 6-minute exercise test; BAI, Beck Anxiety Inventory; BDI, Baseline Dyspnoea Index; CPI, Composite Physiologic Index; CQLQ, Cough Quality of Life Questionnaire; CRP, clinical, radiological, physiological; DL$_{COSB}$, diffusing capacity of the lung for carbon monoxide, single breath; EQ-5D, European Quality of Life-5 Dimensions; FEV$_1$, forced expiratory volume in 1 second; IPFCRN, Idiopathic Pulmonary Fibrosis Clinical Research Network; NHLBI, National Heart, Lung and Blood Institute; *P*A – aO$_2$, alveolar – arterial oxygen partial pressure; PANTHER, Prednisone, Azathioprine, and *N*-acetylcysteine: a study THat Evaluates Response in idiopathic pulmonary fibrosis; *P*aO$_2$, partial pressure of oxygen in arterial blood; PFT pulmonary function test; PSS, Perceived Stress Scale; SF-36, Short-Form questionnaire-36 items; SpO$_2$, peripheral oxygen saturation; UCSDSBQ, University of California, San Diego, Shortness-of-Breath Questionnaire; VAS, visual analogue scale.
a Further details of inclusion and exclusion criteria for each included study can be seen in *Appendix 5*.

TABLE 5 Baseline characteristics of participants

Study	Arm	Age (years)	Sex (M/F), %	Time since diagnosis, years	Surgical lung biopsy, n (%)	FVC,[a] % predicted	DL$_{CO}$, % predicted or mmol/minute/Pa	Smoking history, %
Pharmacological agents								
Azathioprine								
Raghu et al. 1991[70]	Prednisone and azathioprine	58 (SE 2)	36/64	2.2 (SE 0.5)	Assumed all	70 (SE 4)	48 (SE 5)[a]	NR
	Prednisone and placebo	54 (SE 3)	54/46	1.9 (SE 0.5)	Assumed all	65 (SE 4)	40 (SE 4)[a]	NR
BIBF 1120								
[b]Richeldi et al. 2011[71]	BIBF 1120 300 mg/day	65.4 (7.8)	77/23	1.0 (1.2)	29 (34.1)	79.1 (18.5)	3.7 (1.0) mmol/minute/Pa	NR
	Placebo	64.8 (8.6)	74/26	1.4 (1.5)	19 (22.4)	81.7 (17.6)	3.8 (1.1) mmol/minute/Pa	NR
NAC (alone or in combination)								
Demedts et al. 2005[67]	NAC, prednisolone, azathioprine	62 (9)	69/31	1.7 (2.4)	38 (48)	[c]64.76 (15.41)	43.04 (13.10)	Never: 39 Former: 58 Current: 4
	Placebo, prednisolone, azathioprine	64 (9)	75/25	1.6 (2.8)	35 (47)	[c]66.57 (14.42)	44.79 (15.15)	Never: 31 Former: 63 Current: 7
Raghu et al. IPFCRN 2012[68]	NAC, prednisolone, azathioprine	68.8 (7.3)	77/23	0.9 (1.1)	38 (49)	69.3 (15.1)	42.1 (10.2)[a]	Never: 30 Former: 66 Current: 4
	Placebo	67.9 (8.1)	73/27	1.1 (1.0)	37 (47)	72.1 (14.1)	45.3 (12.4)[a]	Never: 26 Former: 69 Current: 5

Study	Arm	Age (years)	Sex (M/F), %	Time since diagnosis, years	Surgical lung biopsy, n (%)	FVC,[a] % predicted	DL$_{CO}$, % predicted or mmol/minute/Pa	Smoking history, %
Homma et al. 2012[69]	NAC	67.6 (6.4)	76/24	3.0 (3.4)	NR	89.2 (17.8)	72.3 (25.3)	Never: 26 Former: 67 Current: 8
	Control	68.2 (7.7)	76/24	3.2 (2.5)	NR	88.7 (15.5)	64.4 (20.1)	Never: 24 Former: 68 Current: 8
Pirfenidone								
Noble et al. 2011[62] Capacity study 006	Pirfenidone 2403 mg	66.8 (7.9)	72/28	≤1 year: 58%	94 (55)	74.9 (13.2)	47.8 (9.8)	Never: 35 Former: 65 Current: 0
	Placebo	67.0 (7.8)	72/28	≤1 year: 62%	94 (54)	73.1 (14.2)	47.4 (9.2)	Never: 37 Former: 58 Current: 5
Noble et al. 2011[62] Capacity study 004	Pirfenidone 2403 mg	65.7 (8.2)	68/32	≤1 year: 48%	86 (49)	74.5 (14.5)	46.4 (9.5)	Never: 32 Former: 63 Current: 5
	Placebo	66.3 (7.5)	74/26	≤1 year: 47%	85 (49)	76.2 (15.5)	46.1 (10.2)	Never: 29 Former: 66 Current: 5

TABLE 5 Baseline characteristics of participants (*continued*)

Study	Arm	Age (years)	Sex (M/F), %	Time since diagnosis, years	Surgical lung biopsy, *n* (%)	FVC,[a] % predicted	DL$_{CO}$, % predicted or mmol/minute/Pa	Smoking history, %
Taniguchi *et al.* 2010[65]	Pirfenidone 1800 mg	65.4 (6.2)	79/21	< 1 year: 35% 1–3 year: 27% ≥ 3 year: 38%	26 (24)	[c]77.3 (16.8)	52.1 (16.8)	Never: 20 Former: 75 Current: 5
	Placebo	64.7 (7.3)	78/22	< 1 year: 39% 1–3 year: 24% ≥ 3 year: 37%	28 (27)	[c]79.1 (17.4)	55.2 (18.2)	Never: 20 Former: 67 Current: 13
Azuma *et al.* 2005[66]	Pirfenidone 1800 mg	64.0 (7.1)	86/14	< 1 year: 28% 1–3 year: 24% ≥ 3 year: 49%	15 (21)	[c]81.6 (20.3)	57.6 (17.2)	Never: 11 Former: 79 Current: 10
	Placebo	64.3 (7.6)	94/6	< 1 year: 17% 1–3 year: 29% ≥ 3 year: 54%	8 (23)	[c]78.4 (17.2)	57.7 (13.8)	Never: 6 Former: 86 Current: 9
Thalidomide								
Horton *et al.* 2012[55,75]	Thalidomide and placebo (crossover trial)	67.6 (7.8)	78/22	1.7 (0.25–4.9)	5 (21.7)	70.4 (13.7)	57.4 (14.4)	NR

Study	Arm	Age (years)	Sex (M/F), %	Time since diagnosis, years	Surgical lung biopsy, n (%)	FVC,[a] % predicted	DL_{CO}, % predicted or mmol/minute/Pa	Smoking history, %
Sildenafil (severe IPF)								
Zisman et al. IPFCRN 2010[72]	Sildenafil	69.7 (8.7)	84/16	2.03 (1.94)	NR	54.89 (14.00)	25.81 (6.03)	History of smoking: 76
	Placebo	68.2 (9.3)	82/18	1.87 (1.93)	NR	58.73 (14.12)	26.73 (6.16)	History of smoking: 76
Non-pharmacological interventions								
Pulmonary rehabilitation/disease management programme								
Lindell et al. 2010[74]	Disease management (PRISIM)	65.2 (10.3)	70/30	NR	3 (30)	> 55: 8 (80)	NR	NR
						50–55: 1 (10)		
						< 50: 1 (10)		
	Usual care control	67.09 (11.90)	82/18	NR	9 (81)	> 55: 6 (60)	NR	NR
						50–55: 2 (20)		
						< 50: 2 (20)		
Jastrzebski et al. 2008[61]	Inspiratory muscle training	56.5 (6.5)	63/37	> 2 years	NR	67.3 (14.3)	39.5 (15.9)	NR
	Control	56.2 (7.2)	64/36	> 2 years	NR	69.2 (14.6)	38.1 (18.9)	NR
Nishiyama et al. 2008[73]	Pulmonary Rehabilitation	68.1 (8.9)	92/8	NR	NR	66.1 (13.2)	59.4 (16.7)	NR
	Control	64.5 (9.1)	60/40	NR	NR	68.7 (19.5)	48.6 (16.7)	NR

F, female; IPFCRN, Idiopathic Pulmonary Fibrosis Clinical Research Network; M, male; NR, not reported; PRISIM, Program to Reduce IPF Symptoms and Improve Management; SE, standard error.
a Corrected for haemoglobin.
b For lower-dose groups, see *Appendix 5*.
c Vital capacity reported.
Mean (SD) unless otherwise stated, % rounded to whole numbers.

Pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis

Azathioprine

One RCT compared azathioprine and prednisolone with placebo and prednisolone in 27 participants.[70] Azathioprine and prednisolone use in the UK is limited; however, these interventions were included as they are still occasionally used. Azathioprine was prescribed at a rate of 3 mg/kg/day (to not exceed 200 mg/day) for the 12 months of the study. Prednisolone was commenced at 1.5 mg/kg/day (to not exceed 100 mg/day) for 2 weeks, and was then decreased at a rate of 20 mg/day until 40 mg/day was reached, which was then prescribed for 2 weeks. After this time, and depending on tolerance, this was subject to further decreases of 5 or 10 mg/day every 2 weeks. The maintenance dose was no more than 20 mg/day for the remaining months of the study. The primary outcome was not stated in this study, which measured FVC and survival at follow-up (at least 12 months). The study did not provide details of any power calculation and, with the small sample size, may be underpowered to detect an effect in the outcome measures reported.

Participants were included in this study if they had a diagnosis of IPF supported by lung biopsy and were previously untreated and newly diagnosed. Criteria for progressive clinical disease were set as seen in *Table 4*, with participants being required to meet one or more of these to be eligible for the trial. Exclusion criteria can be seen in *Table 4*. However, prior to 2000, NSIP and IPF were classified together, and it is possible that this study included patients who would be diagnosed with NSIP according to current criteria. The mean age of participants was 58 years in the treatment group and 54 years in the comparison group. There were a greater proportion of males in the comparator group (53% vs. 36% in the intervention group). Participants had been diagnosed with IPF for approximately 23–26 months and their baseline per cent predicted FVC was in the region of 65–70%. Baseline per cent predicted DL_{CO} ranged from 40% in the control group to 48% in the treatment group. No details were provided of the participants' smoking histories.

Participants were able to cross over to the other treatment arm for a number of stated reasons (see *Appendix 5*), including adverse events unresponsive to treatment, disease progression and participants request. During the study, three participants crossed over because of clinical deterioration: two from the placebo and prednisolone group, and one from the azathioprine and prednisolone group. Four participants from each group died before completion. Results from this trial are reported in *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *Azathioprine*.

BIBF 1120 (nintedanib)

One RCT[71] investigated the use of a new agent for IPF, BIBF 1120 (referred to in the RCT as BIBF 1120, also known as nintedanib). This trial was a dose-finding study assessing four doses of BIBF 1120 against a placebo comparator. Currently, the treatment is unlicensed in IPF. It has been included in the current review as it may become a relevant treatment in the future. Caution is, therefore, recommended in the interpretation of these data as no confirmatory phase III trials are currently published.

Four randomised groups received doses of BIBF 1120 of 50 mg/day, 100 mg/day, 200 mg/day and 300 mg/day, respectively. A fifth group received matched placebo. Doses of BIBF 1120 were escalated using a stepwise approach, beginning with the lowest dose of BIBF (or placebo) and each stepwise increase being reviewed by a data monitoring committee before proceeding. For the purpose of the current review, the 300 mg/day group versus placebo group is described as this is the treatment group that the trial authors concentrated on. Details of the lower-dose groups have been data extracted, however, and can be seen in *Appendix 5*. Treatment was for 52 weeks. Four hundred and thirty-two participants were randomised (four did not participate) to the five groups, with each group having 85 or 86 participants (see *Table 4*), meeting the requirement of the power calculation for this study. The primary outcome in this study was the annual rate of decline in FVC. Secondary outcomes included lung function tests [FVC, DL_{CO}, peripheral oxygen saturation (SpO_2) and TLC]; the 6MWT; QoL assessed by the SGRQ; acute exacerbations; survival; and adverse events. End-point assessment was at 54 weeks.

Individuals with a diagnosis of IPF that was consistent with the ATS/ERS 2000 criteria, and were diagnosed within the previous 5 years, were eligible for inclusion. Participants were also required to be at least 40 years old and have a FVC that was at least 50% and a DL_{CO} that was between 30% and 79% predicted. Those with continuous oxygen supplementation, the likelihood of lung transplant or a life expectancy of < 2.5 years were excluded (see *Table 4* for further exclusion criteria that were specified in this study). Baseline characteristics of the participants were observed to be similar between the treatment groups. The mean age at baseline was approximately 65 years in all groups, and the proportion of males ranged from 72% to 77%. Approximately 22–34% of participants had their diagnosis made on the basis of results from a surgical lung biopsy and the time since diagnosis was in the region of 1.0–1.4 years. Baseline lung function parameters ranged from 79.1% to 85.5% for per cent predicted FVC, and 3.7 to 3.9 mmol/minute/kPa for DL_{CO}. Smoking history of participants was not reported and the 6MWT was similarly not measured. The proportion of participants discontinuing study medication ranged from 16.3% (200 mg/day) to 37.6% (300 mg/day) in the BIBF 1120-treated participants, and was 28.2% in the placebo-treated participants. In addition, four individuals who were randomised in the study did not participate, although the reasons were not provided. The study reports that 85.7% of discontinuations were due to adverse events. The study also reports the numbers of participants who had their doses reduced (see *Appendix 5* for details). Results from this trial are reported in *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, BIBF 1120 (nintedanib).

N-acetylcysteine (alone or in combination)

Three RCTs investigated the use of NAC for IPF. In two RCTs, NAC was used in combination with prednisolone and azathioprine,[67,68] and in the other RCT NAC was given as a monotherapy in an inhaled (nebulised) form.[69] The combination of NAC, prednisolone and azathioprine is hereafter referred to as 'triple therapy'. The use of triple therapy or inhaled NAC is limited in the UK; however, these were included as they are occasionally used in clinical practice.

Demedts and colleagues[67] randomised participants to two groups: triple therapy (*n* = 92), or placebo, prednisolone and azathioprine (referred to here as 'dual therapy') (*n* = 90). The dose of NAC in this study was 600 mg, three times per day. Azathioprine was given at a rate of 2 mg/kg/day and prednisolone was given in a reducing dose from 0.5 mg/kg/day for month 1, 0.4 mg/kg/day for month 2, 0.3 mg/kg/day for month 3, until 10 mg per day during months 4–6. This maintenance dose was then continued until study completion after month 12. The Raghu and colleagues Idiopathic Pulmonary Fibrosis Clinical Research Network (IPFCRN) 2012 trial[68] randomised participants to three groups: NAC monotherapy, triple therapy and placebo. The trial publication reports data for the triple-therapy (*n* = 77) and placebo (*n* = 78) groups only, which were analysed at an interim time point (mean follow-up of 32 weeks). The triple-therapy arm was discontinued at this point by the data safety and monitoring committee, with the comparison of NAC monotherapy and placebo continuing to study end (anticipated publication was 2014). NAC was prescribed at a dose of 600 mg orally, three times each day, in this study. Azathioprine was prescribed at a maximum of 150 mg/day based on ideal weight, concurrent use of allopurinol, and thiopurine methyltransferase (TPMT) activity. Prednisone was provided in a decreasing dose from 0.5 mg/kg per day of ideal body weight to 0.15 mg/kg during 25 weeks. The duration of treatment was up to 60 weeks. In the Homma and colleagues trial,[69] NAC was given as an inhalation of 352.4 mg diluted with saline to a total volume of 4 ml, twice per day via a nebuliser, to 51 participants. The comparison was made with a no-treatment group, which the authors also described as a placebo group. The duration of treatment was 48 weeks in this trial.

The primary outcomes reported in these studies were similar, although different lengths of follow-up were used. In the Demedts and colleagues trial,[67] the primary outcome was the change in absolute VC and DL_{CO}. In the Raghu and colleagues IPFCRN 2012 trial,[68] the primary outcome was change in FVC. The Homma and colleagues trial[69] has absolute change in FVC as their primary outcome. Secondary outcomes in the Demedts and colleagues[67] trial included per cent predicted VC, per cent predicted DL_{CO}, alveolar volume, dyspnoea, QoL and adverse events. The length of follow-up was 12 months. Secondary outcomes in the Raghu and colleagues IPFCRN 2012 trial[68] included rate of death, acute exacerbation rate, time to disease progression, DL_{CO}, 6MWT distance, QoL, dyspnoea and adverse events. The analysis was

undertaken after 60 weeks. Secondary outcomes in the Homma and colleagues trial[69] included 6MWT distance, per cent predicted VC, TLC, DL_{CO}, disease progression, dyspnoea and adverse events. End-point assessment was at the end of treatment, 48 weeks. Of the three RCTs, only Demedts and colleagues[67] had an adequate sample size based on power calculations (Homma and colleagues[69] did not report a calculation).

In the Demedts and colleagues[67] trial comparing triple therapy with dual therapy, participants were eligible for the study if they were aged between 18 and 75 years and had a histological or radiologic pattern typical of UIP. In those aged < 50 years they also required evidence of IPF from either a surgical or a thoracoscopic lung biopsy. The duration of IPF had to be > 3 months and the VC no greater than 80% predicted or a TLC < 90% predicted, with a single breath DL_{CO} of < 80% predicted. Other specific inclusion criteria can be found in *Table 4* and *Appendix 5*. Key exclusion criteria were contraindication or intolerance to the study treatments, use of prednisolone or azathioprine in the month prior to randomisation, coexistent diseases, or treatments that interfere with IPF. *Table 4* and *Appendix 5* provide further detail of exclusion criteria for this trial. The Raghu and colleagues IPFCRN 2012 trial[68] included only those aged between 35 and 85 years with mild to moderate lung function impairment, defined as FVC at least 50% and DL_{CO} at least 30% of predicted values. Diagnosis of IPF was required using the ATS/ERS 2011 consensus criteria (full details of inclusion criteria are provided in *Table 4*). Those with a forced expiratory volume in 1 second (FEV_1)–FVC ratio of < 0.65, a partial pressure of oxygen in arterial blood (PaO_2) of < 55 mmHg or residual volume > 120% predicted were excluded. Other exclusion criteria included listed for lung transplantation, history of other cardiac or liver diseases (see *Table 4* for further details). In the Homma and colleagues[69] trial, participants were required to have well-defined IPF by the ATS/ERS 2011 criteria and Japanese criteria, aged between 50 and 79 years, and classed as having stage I or stage II disease (details provided in *Table 4*). Exclusion criteria can be seen in *Table 4* and include improvement in symptoms during the preceding 3-month period, and use of a number of treatments for IPF.

Baseline characteristics of the participants in the Demedts and colleagues trial[67] and the Raghu and colleagues IPFCRN 2012 trial[68] appear to be similar. Participants in the Homma and colleagues trial[69] differ from the other two trials on some factors. Mean age of participants was similar across all three trials, with a range of 62–69 years. Similar proportions of male participants were also seen across the three trials, with a range of 69–77%. The mean time since diagnosis of IPF was approximately 3 years for participants in both arms of the Homma and colleagues[69] trial. In the Raghu and colleagues IPFCRN 2012 trial,[68] participants had been diagnosed for approximately 1 year (0.9 years in the intervention arm vs. 1.1 years in the placebo arm). The participants had been diagnosed with IPF for approximately 1.5 years in the Demedts and colleagues trial[67] (19.9 months in the intervention group and 18.9 months in the placebo group). Approximately 50% of participants in the trial by Demedts and colleagues[67] and the 2012 trial by Raghu and colleagues (IPFCRN)[68] had been diagnosed on the basis of a lung biopsy, with similar rates between groups. The rates of lung biopsy was not reported in the Homma and colleagues trial.[69] The baseline per cent predicted VC was 64.76% in the intervention group and 66.57% in the placebo group of the Demedts and colleagues trial,[67] and 69.3% in the intervention group and 72.1% in the placebo group of the Raghu and colleagues IPFCRN 2012 trial.[68] The baseline per cent predicted FVC was higher in the Homma and colleagues trial,[69] at 89.2% in the intervention group and 88.7% in the control group. Predicted DL_{CO} per cent at baseline was also similar between participants in the two triple-therapy trials,[67,68] with this ranging from 42.1% to 45.3%. In the inhaled NAC study,[69] the per cent predicted DL_{CO} ranged between 64.4% and 72.3% in the control and treated groups, respectively. Only the Raghu and colleagues IPFCRN 2012 trial[68] reported baseline 6MWT, which was 362.0 m in the triple-therapy arm and 368.9 m in the placebo arm. Previous and current smokers made up the largest proportion of participants in the three trials, with those having never smoked being in the region of 31–39% in the Demedts and colleagues trial,[67] 26–30% in the Raghu and colleagues IPFCRN 2012 trial[68] and 24–26% in the Homma and colleagues trial.[69]

In the Demedts and colleagues trial,[67] 20% of participants randomised to treatment with triple therapy, and 21% of participants randomised to treatment with dual therapy withdrew from the study (reasons provided in *Appendix 5*). Twenty-seven randomised participants were not treated (12 in the triple-therapy

arm and 15 in the double-therapy arm). In the majority of cases this was because the diagnosis was not confirmed. Treatment modifications were permitted, with treatments being adapted according to the preference of each centre. Standard therapy was changed for 13 participants in the triple-therapy arm and 11 in the double-therapy arm. The withdrawal rate in the Raghu and colleagues IPFCRN 2012 trial[68] was reported to be six (8%) participants from the triple-therapy arm and three (4%) from the placebo arm (reasons provided for both groups in *Appendix 5*). Treatment discontinuations were also reported and were seen to be statistically significantly higher in the triple-therapy arm [23 (30%) discontinued prednisolone, 31 (40%) azathioprine and 24 (31%) NAC, with 20 (26%) discontinuing all three] than in the placebo arm [three (4%) discontinued placebo prednisolone, four (5%) placebo azathioprine and four (5%) placebo NAC, with three discontinuing all three placebos]; $p < 0.001$. In the Homma and colleagues[69] trial, six participants in the inhaled NAC arm were excluded from the efficacy data set and eight in the control group (reasons provided in *Appendix 5*). For the safety analysis, seven participants in the NAC treated group and three in the control group were excluded (reasons given). In total, 38 (74.5%) completed the study in the NAC group and 38 (77.6%) completed the study in the placebo group. No details of dose discontinuations or modifications were provided. Results from these trials are reported in *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, N-*acetylcysteine* (*alone or in combination*).

Pirfenidone

Four RCTs compared pirfenidone with placebo. Pirfenidone is currently available in the UK for patients meeting certain criteria (see *Chapter 1, Description of technologies under assessment*). Two RCTs in Japan used pirfenidone at a dose of 1800 mg per day,[65,66] and two RCTs in Europe used pirfenidone at a dose of 2403 mg per day.[62] These European trials were reported within one publication, and are referred to as the Capacity 004 and Capacity 006 trials. The doses are understood to be different because the participants' average body compositions are different between Japanese and European populations. In two of the RCTs, participants were randomised to a third comparison arm: this was of pirfenidone 1197 mg per day in Capacity 004,[62] and 1200 mg per day in the trial by Taniguchi and colleagues.[65] On advice from our advisory group, the present review will concentrate on the outcomes for the higher doses (1800 mg and 2403 mg per day) from Taniguchi and colleagues[65] and Capacity 004,[62] respectively, as these are aligned with the licensed indications for pirfenidone in these respective populations. Full details of the outcomes for the lower doses can be seen in *Appendix 5*. Therapy was for 72 weeks in the two Capacity trials,[62] 52 weeks in the trial by Taniguchi and colleagues[65] and 9 months in the trial by Azuma and colleagues.[66]

The primary outcome in the two Capacity trials was the change in per cent predicted FVC; in the Taniguchi and colleagues trial, this was change in VC litres, and in the Azuma and colleagues trial the primary outcome was the change in the lowest SpO_2 during the 6MWT. Secondary outcomes include progression-free survival (PFS), dyspnoea, 6MWT distance, and per cent predicted DL_{CO} in the Capacity trials,[62] and PFS, TLC, DL_{CO} and acute exacerbations in the Taniguchi and colleagues trial.[65] Secondary outcomes in the Azuma and colleagues[66] trial include VC, TLC, DL_{CO}, disease progression, acute exacerbations and QoL. All the included trials report data on adverse events and follow-up was undertaken at the end of the treatment period in all four studies (72 weeks in the Capacity trials,[62] 52 weeks in the Taniguchi and colleagues trial[65] and 9 months in the Azuma and colleagues trial[66]). All four studies had reasonable sample sizes; only the two Japanese RCTs reported details of a sample size calculation with the sample size being adequately powered.

Eligibility criteria for the two Capacity trials were identical. Participants were eligible if they were aged between 40 and 80 years, had their diagnosis of IPF within 48 months and had no evidence of improvement in measures of disease severity over the preceding year. Participants were also required to have a FVC of at least 50%, a predicted DL_{CO} of at least 35%, either predicted FVC or DL_{CO} of $\leq 90\%$, and the distance walked on the 6MWT of at least 150 m. Key exclusion criteria included evidence of obstructive airway disease, connective tissue disease, alternative explanation for ILD, and those on the waiting list for a lung transplant. In the Taniguchi and colleagues trial,[65] participants were eligible for inclusion if they were aged between 20 and 75 years and had IPF diagnosed using ATS/ERS guidelines and diagnostic criteria from Japan.

Lung function criteria included oxygen desaturation with at least a 5% difference between resting SpO_2 and the lowest SpO_2 during a 6MWT, and the lowest SpO_2 during the 6MWT of at least 85% while breathing air. Participants were excluded if they had a decrease in symptoms during the preceding 6 months, had used any immunosuppressants or corticosteroids at a dose of > 10 mg/day in the preceding 3 months, or had coexisting lung conditions, a severe respiratory infection or clinical features suggestive of another form of an interstitial pneumonia. Participants were eligible for inclusion in the Azuma and colleagues[66] trial if they had evidence on HRCT of definite or probable UIP, abnormal pulmonary function tests, bibasilar inspiratory crackles, and increased serum levels of damaged-pneumocyte markers. Participants were required to be aged between 20 and 75 years with a partial pressure of oxygen of at least 70 mmHg at rest and SpO_2 of ≤ 90% during exertion. Exclusion criteria were similar to those in the Taniguchi and colleagues[65] trial, i.e. a decrease in symptoms during the preceding 6 months, had used in the preceding 3 months any immunosuppressant agents or corticosteroids at a dose of > 10 mg/day, or had coexisting lung conditions, a severe respiratory infection or clinical features suggestive of another form of an interstitial pneumonia. In addition, participants with uncontrolled diabetes or other comorbidities were excluded.

Some participants' characteristics appear to differ between the four included studies (see *Table 5*). Some 47–58% of participants in the Capacity trials[62] were diagnosed within 1 year. In the Taniguchi and colleagues[65] trial, the proportion of participants diagnosed within the last year ranged between 35% and 39%, and in the Azuma and colleagues[66] trial this ranged between 17% and 28%. Larger proportions of participants in these two trials were diagnosed 3 years or more before inclusion into the trials (36–40% in Taniguchi and colleagues[65] and 49–54% in the Azuma and colleagues trials,[66] respectively). As per the inclusion criteria for the Capacity trials,[62] all participants would have been diagnosed within 4 years but further details of the breakdown of time since diagnosis were not reported. Surgical lung biopsy rates were lower in the two Japanese trials, in the region of 21–29% having had a lung biopsy. In the Capacity trials, the proportion having had a lung biopsy ranged from 37% to 55%. Demographic characteristics of the participants in the four trials, in terms of age and sex, appear to be reasonably similar. In the Capacity trials,[62] the mean age of participants was in the region of 66–68 years, in the Taniguchi and colleagues[65] trial the mean age ranged from 64–65 years, and in the Azuma and colleagues trial[66] the mean age was approximately 64 years in both arms. The proportion of participants who were male was 72% in the Capacity 006,[62] between 68% and 75% in the Capacity 004,[62] between 78% and 86% in the Taniguchi and colleagues trial,[65] and between 86% and 94% in the Azuma and colleagues[66] trial.

Lung function tests at baseline indicated that participants in the Japanese trials were likely to have more 'mild' IPF, with per cent predicted FVC ranging from 76% to 82% across the two studies[65,66] and per cent predicted DL_{CO} ranging between 52% and 58%. These parameters ranged from 73% to 76% for per cent predicted FVC, and 46% to 48% for per cent predicted DL_{CO} in the two Capacity trials,[62] although the differences between these two trials are not clinically significant and are within the error of the tests. The distance walked on the 6MWT was in the region of 378–417 m in the Capacity trials,[62] but was not recorded at baseline in the two Japanese studies. The ratio of former or current smokers to those who had never smoked was in the range of approximately 65% : 35% in the Capacity trials and 90% : 10% in the two Japanese trials (see *Table 5* for full details).

In the two Capacity trials,[62] numbers and reasons for withdrawal from the study and numbers and reasons for discontinuation of treatment were provided. In Capacity 006, there were 13 (7.6%) withdrawals in the pirfenidone group and nine (5.2%) in the placebo group. In Capacity 004, the numbers were 13 (7.5%) in the pirfenidone 2403 mg/day group, five (5.7%) in the pirfenidone 1197 mg/day group and eight (4.6%) in the placebo group. Reasons for withdrawal can be seen in *Appendix 5*. Treatments were discontinued in 34 (19.9%) of the participants treated with pirfenidone in the Capacity 006 study, compared with 31 (17.9%) of the participants treated with placebo (most were for adverse events, but for full details of reasons provided see *Appendix 5*). In Capacity 004, these numbers were 38 (21.8%) in the pirfenidone 2403 mg/day group, 17 (19.5%) in the pirfenidone 1197 mg/day group and 31 (17.8%) in the placebo group. In both studies, dose modifications could be made but no details were reported of the number of participants having modifications to the treatment dose. The Azuma and colleagues[66] study report only

that one participant from each group was excluded for violation of inclusion criteria (1.4% pirfenidone and 2.8% placebo); no details of the number of participants with a dose modification were reported. Treatments were discontinued in 16 (22.2%) of the participants treated with pirfenidone group and eight (22.8%) in the placebo group (for reasons, see *Appendix 5*). In the high-dose pirfenidone arm of the study by Taniguchi and colleagues,[65] 40 (37%) withdrew. The rates of withdrawal in the low-dose group and the placebo group were 15 (27%) and 31 (30%), respectively. Reasons for withdrawal were provided and can be seen in *Appendix 5*. In addition, eight participants were excluded from the study after randomisation. Results from these trials are reported in *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *Pirfenidone*.

Thalidomide

One randomised crossover trial compared thalidomide and placebo in 24 participants.[55] Participants were randomised to receive either thalidomide or placebo for 12 weeks, followed by a 2-week washout period and then the alternative treatment for a further 12 weeks. Thalidomide was prescribed at a dose of 50 mg by mouth at bedtime, with the dose increased to 100 mg if no improvement in cough occurred after 2 weeks. This occurred in 21 of 22 participants receiving thalidomide and 23 of 23 participants receiving placebo.

The trial aimed to determine the efficacy of thalidomide in supressing cough. The primary outcome measure was cough-specific QoL, measured by the Cough Quality of Life Questionnaire (CQLQ), and secondary outcomes were cough measured by a visual analogue scale (VAS) and the SGRQ. No lung function tests were reported. Sample size calculations were based on the CQLQ. No data on the minimum clinically important difference in CQLQ or data on the variance in IPF patients were available; therefore, calculations were based on the ability to recruit 20 participants, which would provide 80% power to detect a difference of 4.67 units with a two-sided alpha error level of 5% in a two-treatment crossover study, based on the assumption that the within-patient SD of the response variance would be 5.0 units.

Participants were included in this study if they had a clinical history consistent with IPF (symptom duration ≥ 3 months and ≤ 5 years) and chronic cough (defined as > 8 weeks' duration, that adversely affected QoL and was not due to other identifiable causes) and were aged > 50 years. Participants were also required to have HRCT scans consistent with IPF or surgical lung biopsy demonstrating UIP, FVC between 40% and 90% predicted, TLC between 40% and 80% predicted, and DL_{CO} between 30% and 90% predicted. Exclusion criteria can be seen in *Table 4*. The mean age of participants was 68 years, and 78% were male. The mean time from diagnosis was 1.7 years (range 0.25–4.9 years). Mean baseline per cent predicted FVC was 70% and per cent predicted DL_{CO} was 57%. No details of smoking history were provided.

Of the 24 participants randomised, 23 were treated and 20 completed both treatment periods. Incomplete data on three participants who received placebo first but withdrew before completion of the thalidomide period were also included in the analysis. Results from this trial are reported in *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *Thalidomide*.

Pharmacological interventions for severe idiopathic pulmonary fibrosis

Sildenafil

One RCT[72] compared the use of sildenafil with a matched placebo in 180 participants (89 and 91 in the two groups, respectively). Sildenafil was given orally at a dose of 20 mg, three times per day, for a 12-week period. The primary outcome of this study was the presence or absence of an improvement of at least 20% in the 6MWT distance at 12 weeks. Secondary outcomes included the degree of dyspnoea, QoL, FVC, DL_{CO}, and adverse events. The study was adequately powered to detect a difference in the primary outcome.

Individuals with a diagnosis of IPF that was consistent with the ATS/ERS 2000 consensus criteria, and with an advanced stage of IPF (as defined by a DL_{CO} of < 35% predicted), were eligible for inclusion. The study excluded those with a 6MWT distance of < 50 m at baseline, those receiving treatments for pulmonary hypertension, those on the waiting list for a lung transplant or with a resting oxygen saturation of < 92% while on supplemental oxygen therapy (see *Table 4* for further exclusion criteria that were specified in this study). Baseline characteristics of the participants were similar between the treatment groups (although arterial oxygen saturation was reported to be statistically significant, the values reported appear to be similar). The mean age or participants ranged between 68 and 70 years and the proportion of males was 75% in both groups. The time since diagnosis was in the region of 2 years. No details were provided about the proportion who had received their diagnosis on the basis of results from a surgical lung biopsy. Per cent predicted FVC was 55% in the sildenafil group and 59% in the placebo group and DL_{CO} was between 26% and 27% predicted. The distance walked on the 6MWT on the second test at baseline was 246 m in the sildenafil-treated group and 270 m in the placebo group. Some 68–69% of participants had a history of smoking (no further details provided).

There were eight (8.9%) participant dropouts from the sildenafil group (reasons provided in *Appendix 5*) and six (6.6%) from the placebo group (see *Appendix 5* for details). No details of treatment discontinuations or dose modifications were reported. Results from this trial are reported in *Assessment of effectiveness of pharmacological interventions for severe idiopathic pulmonary fibrosis, Sildenafil*.

Non-pharmacological interventions

Disease management programmes

One RCT[74] investigated a disease management programme with a usual-care comparator group. This study randomised 10 participant pairs (patients and carers) to the intervention and 11 to the comparison group. The intervention consisted of a group programme which was held for 2 hours once a week for 6 weeks. Participants and their caregivers attended the programme, which also included group exercises. The content for the sessions was developed collaboratively by a pulmonary clinical nurse specialist, a psychiatric clinical specialist and an advanced care planning instructor. The sessions included background information about IPF, causes and treatments; controlling moods and feelings; planning for uncertainty; communication; and symptom management. No details of the background of the group facilitator were provided. Usual-care participants were seen by their clinical team at an interval of every 3–6 months with an optional monthly support group and a nurse specialist available by telephone if required. Psychological counselling was provided if indicated but not offered on a routine basis. Both groups received a book entitled '*Feeling Good: The New Mood Therapy*', which the treatment group referred to in their group exercises and the usual-care control participants could read at their leisure. It is unclear when the follow-up occurred.

Outcomes measured in the RCT were measures of dyspnoea, anxiety, depression, stress and QoL. The study is described as a pilot study and, therefore, no sample size calculation was undertaken.

To be eligible for the study, participants were required to be aged > 21 years and have a diagnosis of IPF reflecting moderate IPF (stated as a per cent predicted FVC of 55–70%) or severe (stated as a per cent predicted FVC of < 55%). The diagnostic criteria for IPF were not stated; however, all participants had received either a surgical lung biopsy or HRCT. Participants' ages ranged from 65 years in the intervention arm to 67 years in the comparator arm. The duration of diagnosis of IPF was not reported. Seventy per cent of participants in the intervention arm were male, compared with 82% of participants in the comparator arm. There was a difference in the proportion of participants who had received a surgical lung biopsy between groups, with 30% in the intervention group and 81% in the comparator group. Baseline per cent predicted FVC was categorised into three groups: > 55%, 50–55% and < 50%. The proportions of participants in these three categories were 80%, 10% and 10%, respectively, in the disease management group, compared with 60%, 20% and 20%, respectively, in the usual-care controls. Baseline DL_{CO}, 6MWT distance and smoking history were not reported. Two (18%) participants in the control group discontinued

the programme having either died or received a lung transplant. Results from this trial are reported in *Assessment of effectiveness of non-pharmacological interventions for severe idiopathic pulmonary fibrosis*, *Disease management programmes*.

Pulmonary rehabilitation

One RCT evaluated pulmonary rehabilitation, assessing this against a control group,[73] and one CCT compared inspiratory muscle training in addition to pulmonary rehabilitation (which was described as being general body conditioning exercises) with the pulmonary rehabilitation alone (control).[61] In the RCT by Nishiyama and colleagues,[73] 30 participants were randomised to the two groups: 15 in the intervention group and 15 in the control. The intervention lasted for 10 weeks and consisted of two sessions each week focusing on exercise and peripheral muscle training, together with some educational lectures in one RCT.[73] The exercises included use of a treadmill, with supplemental oxygen given to maintain oxygen saturation > 90%, strength training using elastic bands, arm raising and knee extensions. No details were provided about the trainer. Details of the control intervention were not reported. In the Jastrzebski and colleagues CCT,[61] 30 participants were randomised, with 16 in the intervention group and 14 in the control group, and the intervention lasted for 12 weeks and was over two 6-week cycles; no other details were translated.

The primary outcome was not described in these studies, which between them measured lung function (including FVC, TLC and DL_{CO}), 6MWT distance, dyspnoea and QoL.[61,73] End-point assessment was at the end of the 10- and 12-week programmes, respectively. No sample size calculation was reported in the study publication for the Nishiyama and colleagues RCT.[73] Owing to limitations of the translation of the Jastrzebski and colleagues[61] study, it is unclear whether or not this was reported.

Individuals < 75 years of age with a diagnosis of IPF using the major criteria of the ATS/ERS 2002 consensus criteria – stable IPF (no infection of exacerbations in the previous 3 months) and shortness of breath on effort – were eligible for inclusion in the Nishiyama and colleagues RCT.[73] Having severe comorbid illnesses, being on long-term oxygen supplementation and having been previously treated with corticosteroids or immunosuppressents were reasons for exclusion (see *Table 4* for further exclusion criteria specified in this study). Jastrzebski and colleagues[61] included those with IPF diagnosed by ATS/ERS criteria if aged over 50 years, and by lung biopsy in those below 50 years. Participants had to have had IPF for at least 2 years to be included in the trial, but those who had a duration of treatment of more than 2 years, were receiving more than 20 mg of prednisolone per day, or were using home oxygen therapy were excluded. The mean age at baseline across the two studies[61,73] ranged from 56 to 68 years, and the proportion of males ranged from 60% to 92%. Surgical lung biopsy rates and the time since diagnosis were not reported. Per cent predicted FVC ranged from 66.1% to 69.2% for the two trials.[61,73] Per cent predicted DL_{CO} ranged from 38.1% to 59.4%, and the 6MWT distance from 385 m to 487.4 m. In the Jastrzebski and colleagues CCT,[61] there were no differences between the groups at baseline. In the Nishiyama and colleagues RCT,[73] the groups were reported to be statistically significantly different at baseline on the $PaCO_2$ mmHg and the Baseline Dyspnoea Index (BDI) score. Two (13.4%) participants from the pulmonary rehabilitation group withdrew prior to the start of the programme in the Nishiyama and colleagues study[73] and four participants did not complete the study in the Jastrzebski and colleagues study.[61] Results from these trials are reported in *Assessment of effectiveness of non-pharmacological interventions for severe idiopathic pulmonary fibrosis*, *Pulmonary rehabilitation*.

Quality of included studies

The methodological quality of reporting varied across studies (*Table 6*). Six RCTs[55,62,67,68,71] described an adequate randomisation procedure which ensured both true random assignment to treatment groups and adequate concealment of allocation. Five RCTs provided details of the methods of generating the randomisation sequence, but details demonstrating adequate concealment of allocation were not presented.[65,66,70,72,74] The remaining two RCTs[69,73] provided no details of the methods of generating the

TABLE 6 Quality assessment of included trials

Study	Random sequence generation	Allocation concealment	Similarity of groups	Blinding of			Dropout imbalance Explained (if yes)	Selective reporting	ITT defined (if yes)	Missing data appropriate (if yes)
				Care provider	Participant	Outcome assessor				
Capacity study 006[62] 2011	Yes	Yes	Yes	Yes	Yes	Yes	No	Unclear	Yes / Yes	Yes / Yes
Capacity study 004[62] 2011	Yes	Yes	Yes	Yes	Yes	Yes	No	Unclear	Yes / Yes	Yes / Yes
Taniguchi et al. 2010[65]	Yes	Unclear	Yes	Unclear	Unclear	Unclear	No	Unclear	No	Yes / Yes
Azuma et al. 2005[66]	Yes	Unclear	Yes	Unclear	Unclear	Unclear	No	Unclear	No	Yes / Yes
Raghu et al. IPFCRN 2012[68]	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes / Yes	Unclear	Yes / No	Yes / No
Homma et al. 2012[69]	Unclear	Unclear	Undear	No	No	No	No	Yes	No	No
Demedts et al. 2005[67]	Yes	Yes	Yes	Unclear	Yes	Unclear	No	Yes	No	Yes / Yes
Raghu et al. 1991[70]	Yes	Unclear	Yes	Unclear	Yes	Unclear	No	Unclear	Yes / Yes	Unclear
Richeldi et al. 2011[71]	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes / *Unclear*	No	Yes / Yes	Yes / Yes

| Study | Random sequence generation | Allocation concealment | Similarity of groups | Blinding of | | | Dropout imbalance Explained (if yes) | Selective reporting | ITT defined (if yes) | Missing data appropriate (if yes) |
				Care provider	Participant	Outcome assessor				
Horton et al. 2012[55,75]	Yes	Yes	NA	Yes	Yes	Yes	Yes	Unclear	Yes	Yes
							Yes		Yes	Yes
Zisman et al. IPFCRN 2010[72]	Yes	Unclear	Yes	Unclear	Yes	Unclear	No	Unclear	Yes	Yes
									Yes	Yes
Lindell et al. 2010[74]	Yes	Unclear	No	No	No	No	Yes	Yes	No	No
							Yes			
Jastrzebski et al.[61] 2008	No	No	Yes	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear
Nishiyama et al. 2008[73]	Unclear	Unclear	No	No	No	Unclear	Yes	Unclear	No	No
							Yes			

ITT, intention to treat; NA, not applicable.

randomisation sequence or the allocation procedure used, and, consequently, are rated as unclear on these quality factors. Without adequate published information, it is not possible to assess whether or not there is a risk of selection bias in these studies, with the allocation sequence being open to possible manipulation. As a non-RCT, the CCT was assessed as 'no' on these two factors.[61]

In 10 studies, the participants appeared to be similar across groups at baseline on key demographic and prognostic characteristics.[61,62,65–68,70–72] This question was not applicable to the crossover trial as all participants received both interventions.[55] In one RCT,[69] baseline characteristics were presented for a subset of participants only (those who provided end-point data) and therefore it was not possible to judge this factor. In two RCTs,[73,74] the groups were seen to have some differences in characteristics at baseline. Three of the RCTs reported details suggesting adequate blinding of both care providers, participants and outcome assessors;[55,62] however, the authors of the crossover trial noted that thalidomide has characteristic effects that might have alerted participants to the identity of the drug. Many of the studies were described as double blind, but methods undertaken to ensure that this occurred were not presented. For the majority of studies, therefore, these factors were marked as unclear. In two non-pharmacological RCTs,[73,74] there was no double blinding, which given the disparity in the treatment and comparator interventions is not surprising; however, these trials do not discuss this. In the CCT, no details were provided in the translated methods; however, as the translation of the paper was not in full this has been marked as unclear on these criteria.[61] Details of blinding for outcome assessors were not reported in 9 of the 14 studies,[61,65–68,70–73] which may lead to detection bias.

There was no imbalance in the rate of dropouts between treatment arms in eight of the RCTs. In four RCTs, an imbalance was observed; however, the studies adequately explained the cause of the imbalance.[55,68,73,74] One further RCT[71] also had an imbalanced dropout rate but in this trial the information provided did not allow an assessment of the reason for this. The CCT was considered unclear on this factor owing to limitations in the translation. Information to judge the likelihood of selective reporting of outcomes was not clear in nine studies. In one study[71] the protocol was available and there was no evidence of selective reporting of outcomes. Selective reporting of outcomes was evident in three RCTs[67,69,74] and one CCT.[61]

Six RCTs[62,68,70–72] reported using an intention-to-treat (ITT) analysis, with only one not providing the definition used.[68] In addition, one RCT did not use the term ITT but appeared to follow ITT principles.[55] Six RCTs did not report an ITT analysis and the CCT was considered unclear on this factor owing to limitations in the translation. Reasons for missing data were adequately explained by nine RCTs,[55,62,65–68,71,72,75] unclear in one RCT[70] and classed as inadequate in three RCTs as there was no discussion of numbers and reasons for any attrition.[69,73,74] The CCT was considered unclear on this factor owing to limitations in the translation.

One RCT[55] used a crossover design and an assessment was made of the appropriateness of the design, which can be seen in *Appendix 5*. A crossover design in this case was deemed to be appropriate despite IPF being a disease that deteriorates over time because the RCT was of a short duration only.

Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis

This section reports the results from the included studies of pharmacological interventions. Results are described narratively and tabulated, and where appropriate have been displayed in forest plots. In most cases these forest plots present the results from the different comparisons only. Where meta-analysis of data was undertaken, this is indicated. Where BIBF 1120 is included the comparison is between the 300 mg/day dose with placebo only. Likewise, pirfenidone includes the comparison between the high dose and placebo only. Figures present treatments for mild IPF and severe IPF in some cases.

Meta-analyses that were undertaken are also presented within these figures where it was appropriate to statistically pool data.

Azathioprine

Survival

At 12 months, four participants in each treatment group of the Raghu and colleagues[70] trial had died. Survival estimates based on an additional observational period (presumed duration of 9 years) were compared between those treated with azathioprine and prednisolone, and those treated with placebo and prednisolone, and were reported in an unadjusted and age-adjusted data set (*Table 7*). The HR of survival in the unadjusted set was 0.48 (95% CI 0.17 to 1.38), a non-statistically significant effect ($p = 0.16$). In the age-adjusted analysis, the azathioprine group were seen to show a survival advantage over those treated with placebo (HR 0.26, 95% CI 0.08 to 0.88), suggesting a 64% reduction in the risk of death with azathioprine; however, it is unclear if this is statistically significant (two *p*-values are reported: $p = 0.02$ using a large sample approximation and $p = 0.05$ using the randomisation test). Moreover, the likely inclusion of participants with NSIP may explain this treatment effect, and the small sample size and potential risk of bias in this study should be considered when interpreting these data.

Forced vital capacity

There was no significant difference in FVC per cent predicted at 12 months (difference of 4.8%; $p = 0.87$) or in the proportion of participants categorised as 'improved' (*Table 8* and *Figure 2*).

TABLE 7 Overall survival: azathioprine

Study details	Estimated overall survival			Mortality
	Outcome	Arm	Mean	At 12 months
Raghu et al.[70]	Survival, all patients	Prednisone and placebo (*n* = 13)	NR	4
		Prednisolone and azathioprine (*n* = 14)	NR	4
		HR (95% CI); *p*-value	0.48 (0.17 to 1.38); $p = 0.16$	
	Survival, all patients adjusted for age	Prednisone and placebo (*n* = 13)		
		Prednisolone and azathioprine (*n* = 14)		
		HR (95% CI); *p*-value	0.26 (0.08 to 0.88); $p = 0.02$	

NR, not reported.

TABLE 8 Forced vital capacity outcomes: azathioprine

Study	Outcome	Interventions		Difference between groups
		Prednisone/azathioprine (*n* = 10)	Prednisone/placebo (*n* = 9)	
Raghu et al.[70]	Change in per cent predicted FVC (SE) at 12 months	6.5 (5.3)	1.7 (7.4)	4.8%;[a] $p = 0.87$
	Per cent predicted FVC (%)[b]			
	Improved	5 (35.7)	3 (23.1)	% improved; $p = 0.68$
	Unchanged	3 (21.4)	3 (23.1)	
	Deteriorated	6 (42.9)	7 (53.8)	

SE, standard error.
a Calculated by reviewer.
b Improvement was defined as at least a 10% improvement in FVC, deteriorated was defined as at least a 10% decrement, and any lesser degree of change were categorised as unchanged.

Study or subgroup	Treatment			Control			Weight	Mean difference IV, random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
Azathioprine vs. placebo at 12 months								
Raghu 1991[70]	6.5	16.8	10	1.7	22.2	9	100.0%	4.80 (−13.05 to 22.65)
Subtotal (95% CI)			**10**			**9**	**100.0%**	**4.80 (−13.05 to 22.65)**
Heterogeneity: not applicable								
Test for overall effect: z=0.53 (p=0.60)								
BIBF 1120 vs. placebo at 54 weeks								
Richeldi 2011[71]	−1.04	9.1	85	−6	9.4	85	100.0%	4.96 (2.18 to 7.74)
Subtotal (95% CI)			**85**			**85**	**100.0%**	**4.96 (2.18 to 7.74)**
Heterogeneity: not applicable								
Test for overall effect: z=3.50 (p=0.0005)								
NAC triple vs. double (final value at 12 months)								
Demedts 2005[67]	63.14	19.98	80	61.59	15.17	75	100.0%	1.55 (−4.01 to 7.11)
Subtotal (95% CI)			**80**			**75**	**100.0%**	**1.55 (−4.01 to 7.11)**
Heterogeneity: not applicable								
Test for overall effect: z=0.55 (p=0.59)								
Pirfenidone vs. placebo at study end								
Capacity 004[62]	−8	16.5	174	−12.4	18.5	174	52.9%	4.40 (0.72 to 8.08)
Capacity 006[62]	−9	19.6	171	−9.6	19.1	173	47.1%	0.60 (−3.49 to 4.69)
Subtotal (95% CI)			**345**			**347**	**100.0%**	**2.61 (−1.11 to 6.33)**
Heterogeneity: $\tau^2=3.28$; $\chi^2=1.83$, df=1 (p=0.18); $I^2=45\%$								
Test for overall effect: z=1.38 (p=0.17)								
Sildenafil vs. placebo at 12 weeks								
IPFCRN 2010[72]	−0.97	4.9	89	−1.29	4.8	91	100.0%	0.32 (−1.10 to 1.74)
Subtotal (95% CI)			**89**			**91**	**100.0%**	**0.32 (−1.10 to 1.74)**
Heterogeneity: not applicable								
Test for overall effect: z=0.44 (p=0.66)								

Favours control   Favours treatment

FIGURE 2 Forest plot of mean change in per cent predicted FVC/VC from trials of pharmacological interventions.

Diffusing capacity of the lung for carbon monoxide

There was no significant difference in DL_{CO} per cent predicted at 12 months (difference 6.4%; $p = 0.70$), or in the proportion of participants categorised as 'improved' (*Table 9* and *Figure 3*).

Other secondary outcomes

Raghu and colleagues[70] also report the partial pressure of oxygen as an outcome measure, the results of which can be found in *Appendix 5*.

Adverse events

Adverse events reported in the Raghu and colleagues[70] trial can be seen in *Table 10*. Overall, there were 28 adverse events in the azathioprine-treated group and 25 in the placebo-treated group. More participants treated with placebo reported a subjective adverse event (gastrointestinal or neuropsychiatric) than those treated with azathioprine, and more azathioprine-treated participants had vertebral fractures. All other adverse events were reasonably well matched between the two groups, although this is based on observation of the data only.

Summary

One small study with an unclear risk of bias provided data on the effectiveness of azathioprine and prednisolone compared with placebo and prednisolone. This trial may include participants who would be diagnosed with NSIP according to current criteria. Treatment with azathioprine and prednisolone led to an improvement in survival compared with placebo and prednisolone when this was age-adjusted, but did not lead to improvements in measures of lung function. Azathioprine did not appear to lead to significant adverse events when compared with placebo over the 12-month duration of this study. Caution is required in the interpretation of these data in the light of the small sample size, potential risk of bias and possible inclusion of NSIP.

TABLE 9 Diffusing capacity of the lung for carbon monoxide outcomes: azathioprine

| Study | Outcome | Interventions | | Difference between groups |
		Prednisone/azathioprine ($n = 10$)	Prednisone/placebo ($n = 9$)	
Raghu *et al.*[70]	Change in per cent predicted DL_{CO} (SE)	7.3 (5.3)	0.9 (5.7)	6.4%;[a] $p = 0.70$
	Per cent predicted DL_{CO} (%)			
	Improved[b]	3 (21.4)	2 (15.4)	% improved; $p = 1.00$
	Unchanged[b]	6 (42.9)	5 (38.5)	
	Deteriorated[b]	5 (35.7)	6 (46.2)	

SE, standard error.
a Calculated by reviewer.
b Improvement defined as at least a 10% improvement in DL_{CO}, deteriorated as at least a 10% decrement in DL_{CO}, and any lesser degree of change as unchanged.

Study or subgroup	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Mean difference IV, random, 95% CI
Azathioprine vs. placebo at 12 months								
Raghu 1991[70]	7.3	16.8	10	0.9	17.1	9	100.0%	6.40 (−8.87 to 21.67)
Subtotal (95% CI)			**10**			**9**	**100.0%**	**6.40 (−8.87 to 21.67)**
Heterogeneity: not applicable								
Test for overall effect: z=0.82 (p=0.41)								
NAC triple vs. double (final value at 12 months)								
Demedts 2005[67]	40.85	14.85	68	38.75	14.75	63	100.0%	2.10 (−2.97 to 7.17)
Subtotal (95% CI)			**68**			**63**	**100.0%**	**2.10 (−2.97 to 7.17)**
Heterogeneity: not applicable								
Test for overall effect: z=0.81 (p=0.42)								
Pirfenidone vs. placebo at study end								
Capacity 004[62]	−7.9	12.8	174	−9.9	12.8	174	49.1%	2.00 (−0.69 to 4.69)
Capacity 006[62]	−9.8	12.3	171	−9.2	12.3	173	50.9%	−0.60 (−3.20 to 2.00)
Subtotal (95% CI)			**345**			**347**	**100.0%**	**0.68 (−1.87 to 3.22)**
Heterogeneity: $\tau^2=1.56$; $\chi^2=1.86$, df=1 (p=0.17); $I^2=46\%$								
Test for overall effect: z=0.52 (p=0.60)								
Sildenafil vs. placebo at 12 weeks								
IPFCRN 2010[72]	−0.33	4.9	89	−1.87	5	91	100.0%	1.54 (0.09 to 2.99)
Subtotal (95% CI)			**89**			**91**	**100.0%**	**1.54 (0.09 to 2.99)**
Heterogeneity: not applicable								
Test for overall effect: z=2.09 (p=0.04)								

Favours control   Favours treatment

FIGURE 3 Forest plot of mean change in per cent predicted DL_{CO} from trials of pharmacological interventions.

TABLE 10 Adverse events: azathioprine

Adverse events	Prednisone and azathioprine (*n* = 14)	Prednisone and placebo (*n* = 13)
Any	28	25
Subjective		
Gastrointestinal	3	6
Neuropsychiatric	1	4
Objective		
Elevated liver enzymes	1	0
Vertebral fractures	3	0
Acne	0	1
Cushingoid features	5	4
Hypertension	2	1
Diabetes treatment (oral)	2	3
Diabetes treatment (insulin)	2	0
Congestive heart failure	0	2
Myocardial infarction	1	1
Urosepsis	1	1
Bacterial pneumonia	2	0
Herpes zoster	1	0
Cataracts	1	1
Myopathy	2	1
Peptic ulcer disease	1	1
Pancytopenia	0	0

BIBF 1120 (nintedanib)

As discussed above, the narrative synthesis in the present report concentrates on the BIBF 1120 300 mg/day treatment group, compared with placebo, reported in the Richeldi and colleagues[71] publication. Full details of outcomes from the lower-dose groups and their comparisons with placebo can be found in *Appendix 5*.

Forced vital capacity

The annual rate of decline in FVC was the primary outcome in this study. After 54 weeks this was lower (0.06 l) in the BIBF 1120 300 mg/day-treated group than in the placebo-treated group (0.19 l). The difference-between-groups (0.13 l) comparison was not statistically significant ($p = 0.06$) using the prespecified closed testing procedure for multiplicity. With hierarchical testing this was statistically significant ($p = 0.01$). A statistically significant difference was seen between the absolute change in FVC at 54 weeks between BIBF 1120 300 mg/day and placebo ($p = 0.001$), and the absolute change in per cent predicted FVC at 54 weeks between the two groups ($p < 0.001$), as seen in *Table 11*. *Figures 2* and *4* show, graphically, the mean change in per cent predicted FVC and in FVC litres, respectively. The proportion of participants with a reduction in mean FVC of > 10% or 200 ml was lower in the BIBF 1120 300 mg/day group than that seen in the placebo group (23.8% vs. 44%, respectively). This was shown to be statistically significantly different between groups ($p = 0.004$) and is displayed in *Figure 5*. It is unclear whether this was calculated as an absolute or relative change in FVC.

TABLE 11 Forced vital capacity: BIBF 1120

Study	Outcome	Interventions		Difference between groups
		BIBF 1120 300 mg/day (*n* = 85)	Placebo (*n* = 85)	
Richeldi *et al.* 2011[71]	Annual rate of decline in FVC, mean (SE) (95% CI), l, at 54 weeks	−0.06 (0.04) (−0.14 to 0.02)	−0.19 (0.04) (−0.26 to −0.12)	0.13 l;[a] *p* = 0.06[b]
	Absolute change mean (SE) (95% CI) in FVC, l, at 54 weeks	−0.06 (0.04) (−0.13 to 0.01)	−0.23 (0.04) (−0.30 to −0.16)	0.17 l;[a] *p* = 0.001
	Absolute change in FVC, per cent predicted (SE) (95% CI)	−1.04 (0.99) (−2.98 to 0.91)	−6.00 (1.02) (−8.01 to −4.00)	4.96;[a] *p* < 0.001
	Participants with a reduction in mean FVC of > 10% or 200 ml, *n* (%)	20 (23.8)	37 (44.0)	20.2%;[a] *p* = 0.004

SE, standard error.
a Calculated by reviewer.
b Between 300 mg/day group and placebo using the closed testing procedure for multiplicity; *p* = 0.01 with hierarchical testing.
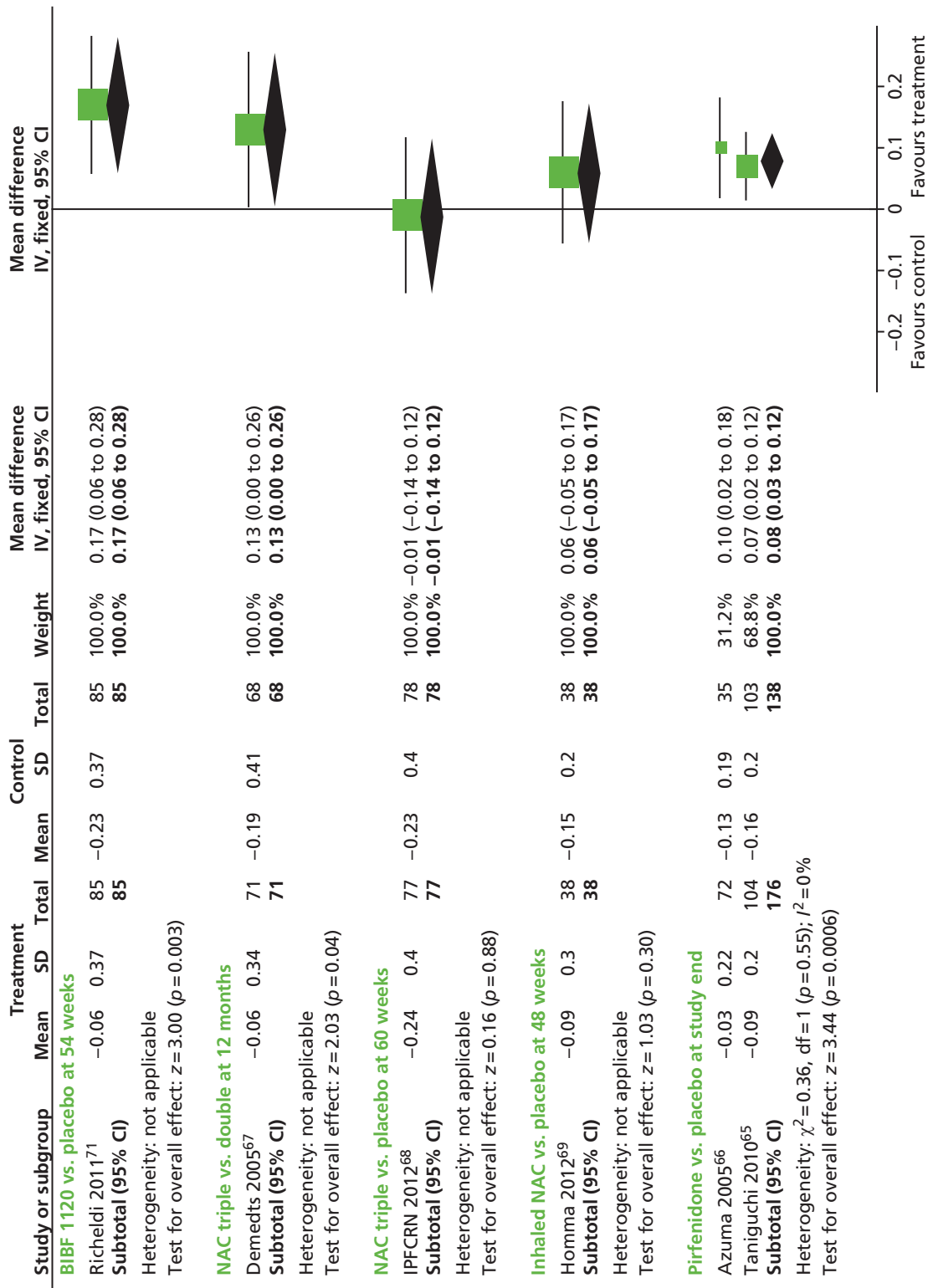
FIGURE 4 Forest plot of mean change in FVC/VC from trials of pharmacological interventions.

| Study or subgroup | Treatment | | Control | | | | |
	Events	Total	Events	Total	Weight	RR M–H, random, 95% CI	RR M–H, random, 95% CI
BIBF 1120 vs. placebo at 54 weeks							
Richeldi 2011[71]	20	85	37	85	100.0%	0.54 (0.34 to 0.85)	
Subtotal (95% CI)		**85**		**85**	**100.0%**	**0.54 (0.34 to 0.85)**	
Total events	20		37				
Heterogeneity: not applicable							
Test for overall effect: $z = 2.66$ ($p = 0.008$)							
Pirfenidone vs. placebo at study end							
Azuma 2005[66]	9	72	12	35	19.2%	0.36 (0.17 to 0.78)	
Capacity 004[62]	35	174	60	174	40.8%	0.58 (0.41 to 0.84)	
Capacity 006[62]	39	171	46	173	40.0%	0.86 (0.59 to 1.24)	
Subtotal (95% CI)		**417**		**382**	**100.0%**	**0.62 (0.41 to 0.93)**	
Total events	83		118				
Heterogeneity: $\tau^2 = 0.07$; $\chi^2 = 4.69$, df = 2 ($p = 0.10$); $I^2 = 57\%$							
Test for overall effect: $z = 2.29$ ($p = 0.02$)							

FIGURE 5 Forest plot of FVC/VC decline of ≥ 10% from trials of pharmacological interventions. M–H, Mantel–Haenszel.

Diffusing capacity of the lung for carbon monoxide

The Richeldi and colleagues[71] publication stated that it measured DL_{CO} as an outcome measure; however, no data were reported at follow-up for this outcome.

Six-minute walk test

The Richeldi and colleagues[71] publication stated that it measured the 6MWT as an outcome measure; however, no data were reported at follow-up for this outcome.

Other secondary outcomes

Richeldi and colleagues[71] also reported the oxygen saturation and the TLC as outcome measures, the results of which can be found in *Appendix 5*.

Acute exacerbations

The incidence of acute exacerbations was statistically significantly lower in the BIBF 1120 300 mg/day-treated participants than in the placebo-treated participants in the Richeldi and colleagues trial (*Table 12*).[71] The rates were 15.7 per 1000 patient-years in the placebo group and 2.4 per 1000 patient-years in the BIBF 1120 300 mg/day-treated group ($p = 0.02$). This equated to a RR of 0.16 (95% CI 0.03 to 0.70) in favour of treatment.

St George's Hospital Respiratory Questionnaire

The total score, symptoms score, activity score, and impacts score on the SGRQ were reported as change from baseline scores at follow-up (54 weeks) in the Richeldi and colleagues[71] trial. As can be seen in *Table 13*, the change scores were statistically significantly different between the BIBF 1120 300 mg/day group and placebo group for the change in total score, symptom score and activity score. No statistically significant effect of BIBF 1120 was seen on the impacts score of the SGRQ.

The authors stated in the publication that the minimally clinical important difference (MCID) in total score on the SGRQ is 4 points, and they analysed the proportion of participants improving by at least 4 points. The proportions improving by at least four points on the SGRQ total score was 29.1% in the BIBF 1120 300 mg/day-treated group compared with 16.1% in the placebo-treated group ($p = 0.03$). It should be noted that elsewhere in the publication it states that the MCID is 5–8 points. For a fuller description of the SGRQ, see *Chapter 4, Systematic review of health-related quality of life studies*.

TABLE 12 Acute exacerbation rate: BIBF 1120

| Study | Outcome | Interventions | | Difference between groups |
		BIBF 1120 300mg/day ($n = 85$)	Placebo ($n = 85$)	
Richeldi *et al.* 2011[71]	Incidence of acute exacerbations, *n* per 100 patient-years	2.4	15.7	13.3;[a] $p = 0.02$
	Incidence of acute exacerbations, RR (95% CI) compared with placebo	0.16 (0.03 to 0.70)	NA	

NA, not applicable.
a Calculated by reviewer.

TABLE 13 St George's Hospital Respiratory Questionnaire: BIBF 1120

Study	Outcome	Interventions		Difference between groups
	Mean (SD or SE)	BIBF 1120 300 mg/day (*n* = 85)	Placebo (*n* = 85)	
Richeldi *et al.* 2011[71]	SGRQ,[a] total score (SD) at baseline	40.1 (18.3)	41.2 (17.9)	
	SGRQ change in total score (SE) (95% CI) at 54 weeks	−0.66 (1.71) (−4.02 to 2.71)	5.46 (1.73) (2.06 to 8.86)	*p* = 0.007
	SGRQ symptoms score (SD) at baseline	43.1 (25.2)	42.2 (21.6)	
	SGRQ change in symptoms score (SE) (95% CI) at 54 weeks	−3.14 (2.40) (−7.86 to 1.58)	6.45 (2.45) (1.65 to 11.26)	*p* = 0.03
	SGRQ activity score (SD) at baseline	53.9 (21.6)	54.2 (22.2)	
	SGRQ change in activity score (SE) (95% CI) at 54 weeks	0.32 (1.89) (−3.39 to 4.03)	7.48 (1.91) (3.73 to 11.24)	*p* = 0.004
	SGRQ impacts score (SD) at baseline	30.8 (19.0)	33.1 (19.7)	
	SGRQ change in impacts score (SE) (95% CI) at 54 weeks	−0.14 (1.97) (−4.00 to 3.73)	4.21 (1.99) (0.31 to 8.12)	NS
	SGRQ % improving ≥ 4 points	29.1%	16.1%	*p* = 0.03

NS, not significant; SE, standard error.
a The scores for the subgroups of the SGRQ all range from 0 to 100, with lower scores indicating better QoL.

Deaths

The number of deaths from respiratory causes was reported in the Richeldi and colleagues[71] trial. These were two (2.4%) in the BIBF 1120 300 mg/day arm and eight (9.2%) in the placebo arm (*Table 14*). The paper reports that the difference between groups was not statistically significant ($p = 0.06$).

Deaths from any cause were more similar between the two groups, with seven (8.2%) participants dying from any cause in the BIBF 1120 300 mg/day treatment arm and nine (10.3%) in the placebo-treated arm (*p*-value not reported, paper stated not significant).

TABLE 14 Adverse events: BIBF 1120

Adverse events	BIBF 1120 300 mg/day (*n* = 85)	Placebo (*n* = 85)
Deaths from respiratory causes, *n*	2[a]	8
Deaths from any cause, *n*	7	9
Adverse events, any, *n* (%)	80 (94.1)	77 (90.6)
Adverse events occurring in > 10% in any study arm, *n* (%)		
Diarrhoea	47 (55.3)	13 (15.3)
Cough	8 (9.4)	17 (20.0)
Nausea	20 (23.5)	8 (9.4)
Bronchitis	9 (10.6)	11 (12.9)
Dyspnoea	6 (7.1)	11 (12.9)
Progression of IPF	4 (4.7)	11 (12.9)
Vomiting	11 (12.9)	4 (4.7)
Upper abdominal pain	10 (11.8)	3 (3.5)
Nasopharyngitis	6 (7.1)	11 (12.9)
URTI	7 (8.2)	13 (15.3)
Headache	11 (12.9)	5 (5.9)
Fatigue	9 (10.6)	7 (8.2)
Decreased appetite	13 (15.3)	0
Severe adverse events	19 (22.4)	20 (23.5)
Serious adverse events	23 (27.1)	26 (30.6)
Fatal adverse events	1 (1.2)	12 (14.1)
Adverse events requiring hospitalisation	23 (27.1)	22 (25.9)
Drug-related adverse event	55 (64.7)	25 (29.4)
Adverse events leading to discontinuation	26 (30.6)	22 (25.9)
Respiratory, thoracic and mediastinal disorders	4 (4.7)	10 (11.8)
GI disorders (see also below)	14 (16.5)	2 (2.4)
Infections and infestations	0	6 (7.1)
Cardiac disorders	0	6 (7.1)

continued

TABLE 14 Adverse events: BIBF 1120 (*continued*)

Adverse events	BIBF 1120 300 mg/day (*n* = 85)	Placebo (*n* = 85)
GI adverse events, *n* (%)		
Any	63 (74.1)	27 (31.8)
Severe diarrhoea	4 (4.7)	0
Serious diarrhoea	3 (3.5)	0
Serious GI, %	4.7%	0
Severe GI, %	5.9%	0
Related to treatment (any GI)	48 (56.5)	11 (12.9)
Related to treatment (diarrhoea)	36 (42.4)	5 (5.9)
Reduced dose due to GI adverse event	9 (10.6)	0
Discontinuation due to any GI adverse event	14 (16.5)	2 (2.4)
Discontinuation due to diarrhoea	10 (11.8)	0
Mean number days with diarrhoea, *n*	85.4	6.5
Discontinuation due to nausea, %	4.7%	0%
Discontinuation due to vomiting, %	2.4%	1.2%

GI, gastrointestinal; URTI, upper respiratory tract infection.
a $p = 0.06$.

Adverse events

The study reports on a range of adverse events (see *Table 14*). The proportion with any adverse event was similar across both groups, with 94.1% of those treated with BIBF 1120 300 mg/day and 90.6% of the placebo-treated group having an adverse event recorded. On observation of the data, more participants treated with BIBF 1120 300 mg/day had diarrhoea, nausea, vomiting, abdominal pain, headache or decreased appetite (see also specific gastrointestinal events in *Table 14*), and more participants treated with placebo had upper respiratory tract infection (URTI) and nasopharyngitis. The rate of severe adverse events and serious adverse events appears to be similar between the two groups, as is the rate of hospitalisation due to adverse events (see *Table 14*). More fatal adverse events occurred in the placebo group than the BIBF 1120 300 mg/day-treated group. Adverse events categorised as drug-related occurred in 64.7% of BIBF 1120 300 mg/day-treated participants, compared with 29.4% of placebo-treated participants.

Summary

BIBF 1120 at 300 mg/day led to more favourable outcomes than placebo on some measures of the FVC, rate of acute exacerbations and the number of deaths. The primary outcome in this study was the annual rate of decline in FVC, and the observed benefit did not reach statistical significance. QoL as reported on the SGRQ was generally favourable in those treated with BIBF 1120 300 mg/day. Serious adverse event rates and hospitalisation rates appear to be similar between the two groups. BIBF 1120 300 mg/day led to more gastrointestinal events than placebo. Treatment with placebo led to more respiratory events. Risk of bias was low in this study. This study also randomised participants to three other treatment groups according to the dose of BIBF 1120 and results for these can be seen in *Appendix 5*.

N-*acetylcysteine alone or in combination*

Progression-free survival

In the Raghu and colleagues IPFCRN 2012 trial,[68] the outcome of time to death or disease progression, defined as a relative drop in FVC of > 10%, was reported. Although not labelled as such, this has been considered as PFS as this definition has been applied in other included studies (see *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *Pirfenidone*) to mean PFS. The HR for the estimate of PFS between the triple-therapy group and the placebo group was not reported in the trial. The publication states that there was no statistically significant difference in the risk of death or disease progression between groups; however, no *p*-value was reported to support this. The Kaplan–Meier plot was published in the study article.

Progression-free survival was not reported as an outcome in the Demedts and colleagues[67] trial comparing triple therapy with double therapy, or the trial by Homma and colleagues[69] comparing inhaled NAC with control.

Time to death/time to death or hospitalisation

The HR for time to death or the HR for time to death or disease progression was not reported in the Raghu and colleagues IPFCRN 2012 trial.[68] The Kaplan–Meier plots were published in the study article.

In the same study,[68] the HR for time to death or hospitalisation was reported to be 12.11 (95% CI 2.83 to 51.85, *p* < 0.0001), indicating a shorter time to death or hospitalisation in the triple-therapy-treated population.

Outcomes relating to time to death or hospitalisation were not reported in the Demedts and colleagues[67] trial comparing triple therapy with double therapy, or the trial by Homma and colleagues[69] comparing inhaled NAC with control.

Discussion of all-cause mortality, respiratory mortality and any-cause mortality or hospitalisation is provided below.

Forced vital capacity

Demedts and colleagues[67] compared triple therapy with double therapy and assessed the VC in litres and the VC per cent predicted. At 12 months it can be observed that the change in VC was lower in the triple-therapy group than the double-therapy group; however, no statistical analysis of the difference between groups was undertaken (*Table 15*). *Figure 4* presents results graphically where it can be seen that treatment with triple therapy is approaching statistical significance on this outcome (95% CI touches the line of no effect). A statistically significant difference in the mean VC litres between groups in favour of triple therapy at 12 months was observed (difference 0.18; *p* = 0.02) when using the least squares (LS) mean method. The final value per cent predicted VC was also shown to be statistically significantly different between groups at 12 months when using the LS mean method (with the LS mean the difference was 4.79%; *p* = 0.02), but not when calculating the mean difference using the unadjusted data (see *Table 15* and *Figure 2*, calculated by reviewer). The mean change from baseline in VC per cent predicted was not reported. All of the analyses were undertaken using the last observation carried forward (LOCF) approach.

After a mean follow-up of 32 weeks of treatment, no difference in mean change in FVC litres between triple therapy and placebo in the Raghu and colleagues IPFCRN 2012[68] trial was observed (*p* = 0.85) (see *Table 15* and *Figure 4*).

TABLE 15 Forced vital capacity outcomes: NAC therapies

Study	Outcome	Interventions		Difference between groups (95% CI)
		NAC, azathioprine, prednisolone (*n* = 80)	Placebo, azathioprine, prednisolone (*n* = 75)	
Demedts *et al.* 2005[67]	VC, mean (SD), l, at 12 months, LOCF	*n* = 71 2.22 (0.77)	*n* = 68 2.17 (0.71)	
	VC, l, mean change at 12 months, LOCF (95% CI)	−0.06 (−0.14 to 0.02)	−0.19 (−0.29 to −0.09)	0.13[a]
	VC, l, LS mean (SD), LOCF at 12 months	*n* = 71 2.27 (0.05)	*n* = 68 2.10 (0.05)	0.18 (0.03 to 0.32); *p* = 0.02
	VC, % predicted, mean (SD) at 12 months, LOCF	63.14 (19.98)	61.59 (15.17)	Not reported
	VC, % predicted, LS mean (SD) at 12 months, LOCF	65.13 (1.85)	60.34 (1.85)	4.79 (0.80 to 8.77); *p* = 0.02
		NAC/prednisolone/azathioprine (*n* = 77)	Placebo (*n* = 78)	
Raghu *et al.* IPFCRN 2012[68]	FVC change (95% CI), l, at interim analysis	−0.24 (−0.33 to −0.15)	−0.23 (−0.32 to −0.14)	−0.01 (−0.14 to 0.11); *p* = 0.85
		NAC (*n* = 38)	Control (*n* = 38)	
Homma *et al.* 2012[69]	Mean FVC (SD), l, at 48 weeks	2.67 (0.84)	2.51 (0.68)	Not stated
	Mean change (SD) in FVC, ml, at 48 weeks	−90 ml (300)	−150 ml (200)	60 ml; *p* = 0.2661
	% with a decline of > 10% FVC at 48 weeks	Data not shown	Data not shown	36.4% lower in the NAC group than control
	% with a decline of < 10% in FVC at 48 weeks	Data not shown	Data not shown	14.8% greater in the NAC group than control; *p* = 0.42

a Calculated by reviewer.

Inhaled NAC therapy did not confer a statistically significant benefit over no-treatment control on FVC litres in the Homma and colleagues[69] trial (difference 60 ml; *p* = 0.27) (see *Table 15* and *Figure 4*). In this study, data were dichotomised into two categories: those with a decline of < 10% FVC litres, and those with a decline of > 10% FVC litres. It is uncertain if these were a priori groups or post hoc. Data were not presented for the proportions within the treated and placebo groups; however, the differences between groups in these two categories were presented (see *Table 15*), but these did not reach the threshold for statistical significance. The results of this study need to be interpreted cautiously given the uncertain risk of bias identified when assessing the study quality.

Diffusing capacity of the lung for carbon monoxide

All three included trials using NAC therapy alone or in combination recorded DL_{CO} (*Table 16*, see *Figures 3* and *6*). This was reported as the mean change in mmol/minute/mmHg and per cent predicted DL_{CO} in the Demedts and colleagues[67] trial of triple therapy versus double therapy and the Homma and colleagues[69] trial of inhaled NAC versus control, and DL_{CO} mmol/minute/mmHg only in the Raghu and colleagues IPFCRN 2012[68] trial of triple therapy versus placebo. Triple therapy led to a smaller decline in DL_{CO} mmol/minute/mmHg after 12 months than double therapy in the Demedts and colleagues[67] study; however, statistical analysis was not undertaken on the difference between groups on these change scores. The mean value of DL_{CO} (see *Table 16*) was statistically compared between groups and showed a statistically significant difference in favour of triple therapy (difference 0.75, 95% CI 0.27 to 1.23; $p = 0.003$) when using the LS mean method. On the per cent predicted DL_{CO}, a similar pattern was observed in the mean values, where there was a difference between groups of 5.08% (95% CI 1.17% to 8.99%, $p = 0.01$).

In the Raghu and colleagues IPFCRN 2012[68] study, the difference in DL_{CO} change from baseline (corrected for haemoglobin) was not statistically significant (difference 0.06; $p = 0.93$) between triple therapy and placebo. Homma and colleagues[69] did not report data on these outcomes; they reported only that there were no statistically significant differences between groups at 48 weeks.

Six-minute walk test

The 6MWT was reported in two of the RCTs evaluating NAC[68,69] (*Table 17*). The difference between groups in the change in the distance walked on the 6MWT favoured placebo in the Raghu and colleagues IPFCRN 2012[68] trial, although the difference was not statistically significant between the two groups.

TABLE 16 Diffusing capacity of the lung for carbon monoxide: NAC therapy studies

Study	Outcome	Interventions		Difference between groups
		NAC, azathioprine, prednisolone ($n = 80$)	Placebo, azathioprine, prednisolone ($n = 75$)	
Demedts *et al.* 2005[67]	DL_{CO} mean (SD) mmol/minute/kPa at 12 months, LOCF	$n = 68$ 3.74 (1.99)	$n = 63$ 3.20 (1.26)	
	DL_{CO} mmol/minute/kPa mean change at 12 months, LOCF (95% CI)	−0.11 (−0.47 to 0.25)	−0.70 (−0.95 to −0.45)	
	DL_{CO} mmol/minute/kPa, LS mean (SD), LOCF	$n = 68$ 3.85 (0.17)	$n = 63$ 3.10 (0.18)	0.75 (0.27 to 1.23); $p = 0.003$
	DL_{CO} per cent predicted, mean at 12 months, LOCF	40.85 (14.85)	38.75 (14.75)	Not reported
	DL_{CO} per cent predicted, LS mean at 12 months, LOCF	41.6 (1.35)	36.52 (1.45)	5.08 (1.17 to 8.99); $p = 0.01$
		NAC/prednisolone/ azathioprine ($n = 77$)	Placebo ($n = 78$)	
Raghu *et al.* IPFCRN 2012[68]	DL_{CO} corrected, change from baseline	−1.72 (−2.73 to −0.71)	−1.66 (−2.65 to −0.67)	−0.06 (−1.48 to 1.35); $p = 0.93$
		NAC ($n = 38$)	Control ($n = 38$)	
Homma *et al.* 2012[69]	DL_{CO}	Data not shown	Data not shown	States NS
	DL_{CO} per cent predicted	Data not shown	Data not shown	States NS

NS, not significant.

TABLE 17 Six-minute walk test: NAC interventions

| Study | Outcome | Interventions | | Difference between groups |
		NAC/prednisolone/ azathioprine ($n = 77$)	Placebo ($n = 78$)	
Raghu *et al.* IPFCRN 2012[68]	6MWT at baseline	362.0 (113.0)	368.9 (117.3)	
	6MWT distance change from baseline, m	–93.0 (–142.0 to –44.1)	–73.6 (–118.4 to –28.7)	–19.5 (–85.9 to 46.9); $p = 0.56$
		NAC ($n = 38$)	**Control ($n = 38$)**	
Homma *et al.* 2012[69]	6MWT distance	Data not presented	Data not presented	States NS

NS, not significant.

The Homma and colleagues[69] study comparing inhaled NAC with placebo reported that the distance walked was not statistically significantly different between groups, although the data were not provided in the study publication.

Lowest SpO_2 during the 6MWT was also stated to be an outcome in the Homma and colleagues[69] trial; however, data were not reported. The publication states there were no statistically significant differences between those treated with inhaled NAC and those treated as controls (*p*-value not reported).

Other secondary outcomes

Other outcomes reported in these studies include TLC; clinical, radiological, physiological (CRP) scoring system; exercise performance; and HRCT outcome. The results can be found in *Appendix 5*.

Dyspnoea

Dyspnoea was reported as an outcome in the three included RCTs investigating NAC therapy alone or in combination (*Table 18*). Different measures of dyspnoea were used. In the Demedts and colleagues trial[67] of triple therapy compared with dual therapy, a dyspnoea score which measured breathlessness on a scale

TABLE 18 Dyspnoea outcomes: NAC

| Study | Outcome | Interventions | | Difference between groups |
		NAC, azathioprine, prednisolone ($n = 80$)	Placebo, azathioprine, prednisolone ($n = 75$)	
Demedts *et al.* 2005[67]	Dyspnoea score[a] at 12 months, LS mean (SD), LOCF	8.88 (0.49)	9.20 (0.51)	–0.32 (–1.72 to 1.09); $p = 0.65$
		NAC/prednisolone/ azathioprine ($n = 77$)	**Placebo ($n = 78$)**	
Raghu *et al.* IPFCRN 2012[68]	USCDSBQ[b]	10.6 (3.60 to 17.6)	8.01 (1.67 to 14.3)	2.57 (–6.87 to 12.0); $p = 0.59$
		NAC ($n = 38$)	**Control ($n = 38$)**	
Homma *et al.* 2012[69]	% improved or stable dyspnoea	86.8% (33/38)	84.2% (32/38)	$p = 1.00$
		Two improved, 31 stable	One improved, 31 stable	

UCSDSBQ, University of California, San Diego, Shortness-of-Breath Questionnaire.
a Dyspnoea scores range from 0 to 20.
b UCSDSBQ total score ranges from 0 to 120, with higher scores indicating greater shortness of breath.

from 0 to 20 was used. On this measure, higher scores indicate worse dyspnoea. In the Raghu and colleagues IPFCRN 2012 trial[68] of triple therapy compared with placebo, the USCDSBQ scale was used. This scale rates shortness of breath on a scale from 0 to 120, with higher scores indicating greater breathlessness. The Homma and colleagues[69] study reported the proportion of participants with improved or stable dyspnoea (definition not provided). Dyspnoea scores were not statistically significantly different between triple therapy and double therapy in the Demedts and colleagues trial ($p = 0.65$)[67] or between triple therapy and placebo in the Raghu and colleagues IPFCRN 2012 trial[68] (see *Table 18*). The proportion of participants with improved or stable dyspnoea were similar in the inhaled NAC therapy arm and the control arm in the Homma and colleagues trial[69] ($p = 1.00$).

Acute exacerbations

The rate of acute exacerbations were reported in the Raghu and colleagues IPFCRN 2012[68] trial comparing triple therapy with placebo (*Table 19*). There were five (6%) acute exacerbation incidents in the triple-therapy group and no incidents in the placebo group. Analysis of the difference between groups was not reported. Acute exacerbations were not reported as an outcome in the Homma and colleagues[69] study; however, the number of participants excluded from the analysis because of acute exacerbations was recorded. These were four in the placebo group and one in the inhaled NAC group.

Quality of life

Demedts and colleagues[67] and the Raghu and colleagues IPFCRN 2012[68] trial assessed QoL using the SGRQ. However, results were not reported by Demedts and colleagues;[67] therefore, no conclusions can be made regarding QoL comparing triple therapy with double therapy. As noted previously, the SGRQ has a total score and three subscales; each scale ranges from 0 to 100, with lower scores indicating better QoL. The Raghu and colleagues IPFCRN 2012[68] group reported all scales of the SGRQ and also assessed QoL by the Short Form questionnaire-36 items (SF-36) and the European Quality of Life-5 Dimensions (EQ-5D). The SF-36 measures functional health and well-being scores on eight scales that correlate with two aggregate scores (physical and mental health). Each score ranges from 0 to 100, with a higher score indicating better function. Scores are normalised to a mean (SD) of 50 (10). The EQ-5D measures general QoL on a self-report questionnaire on a scale of 0 to 1.00 (higher score indicates better QoL) and on a VAS (described in the Raghu and colleagues IPFCRN 2012 trial as a thermometer score) with a range of 0 to 100 (higher indicates better QoL). For a fuller description of the SF-36 and EQ-5D, see *Chapter 4, Systematic review of health-related quality of life studies*. Homma and colleagues did not report QoL as an outcome.

Full results can be seen in *Table 20*. In the Raghu and colleagues IPFCRN 2012[68] trial, the change in SGRQ scores in those treated with triple therapy was lower (better) than in those treated with placebo; however, the differences were not statistically significant on total score or two of the subscales (only the symptoms score was statistically significantly different; $p = 0.014$). On the SF-36 in this study, the difference between groups on the physical score was not statistically significant ($p = 0.58$). On the SF-36 mental score, the change from baseline score was smaller (better) in the triple-therapy group than the placebo group, and the difference was statistically significant ($p = 0.027$). On the two EQ-5D indices, the difference in changes in scores between the two groups was not statistically significant (see *Table 20*).

TABLE 19 Acute exacerbations: NAC therapy

Study	Outcome	Interventions		Difference between groups
		NAC/prednisolone/ azathioprine ($n = 77$)	Placebo ($n = 78$)	
Raghu *et al.* IPFCRN 2012[68]	Acute exacerbation rate	5 (6%)	0	Not reported

TABLE 20 Quality of life outcomes: NAC therapy

Study	Outcome (outcomes all assessed at interim analysis, up to 60 weeks)	Interventions		Difference between groups
		NAC/prednisolone/ azathioprine (n = 77)	Placebo (n = 78)	
Raghu *et al.* IPFCRN 2012[68]	SGRQ[a] total score (SD) at baseline	38.7 (17.4)	39.4 (17.4)	
	SGRQ total score (95% CI) change	4.29 (–1.14 to 9.73)	7.50 (2.57 to 12.4)	–3.20 (–10.5 to 4.13); $p = 0.39$
	SGRQ symptoms score (SD) at baseline	49.4 (21.1)	45.6 (21.8)	
	SGRQ symptoms score (95% CI) change	–4.42 (–11.9 to 3.1)	8.31 (1.47 to 15.2)	–12.7 (–22.9 to –2.61); $p = 0.014$
	SGRQ activity score (SD) at baseline	51.1 (19.0)	52.7 (21.0)	
	SGRQ activity score (95% CI) change	7.33 (1.05 to 13.6)	10.3 (4.66 to 16.0)	–2.99 (–11.4 to 5.46); $p = 0.49$
	SGRQ impacts score (SD) at baseline	27.8 (19.2)	28.8 (17.3)	
	SGRQ impacts score (95% CI) change	5.23 (–0.80 to 11.3)	5.80 (0.34 to 11.27)	–0.57 (–8.71 to 7.57); $p = 0.89$
	SF-36[b] aggregate physical score (SD) at baseline	40.3 (9.8)	40.6 (9.3)	
	SF-36 aggregate physical score (95% CI) change	–4.18 (–7.40 to –0.97)	–2.96 (–5.90 to –0.02)	–1.23 (–5.58 to 3.13); $p = 0.58$
	SF-36 aggregate mental score (95% CI) at baseline	53.9 (9.6)	55.7 (7.4)	
	SF-36 aggregate mental score (95% CI) change	0.96 (–2.51 to 4.44)	–4.35 (–7.50 to –1.20)	5.31 (0.62 to 10.00); $p = 0.027$
	EQ-5D[c] score (SD) at baseline	0.8 (0.2)	0.8 (0.2)	
	EQ-5D score (95% CI) change	–0.07 (–0.14 to –0.00)	–0.02 (–0.09 to 0.04)	–0.05 (–0.14 to 0.05); $p = 0.31$
	EQ-5D thermometer response (SD) at baseline	76.8 (15.5)	78.1 (15.4)	
	EQ-5D thermometer response (95% CI) change	–6.81 (–13.0 to –0.67)	–6.66 (–12.4 to –0.94)	–0.15 (–8.54 to 8.24); $p = 0.93$

a Scores range from 0 to 100, with lower scores indicating better QoL.
b Score ranges from 0 to 100, with a higher score indicating better function.
c Scale of –0.59 to 1.00 (higher score indicates better QoL).

Adverse events and mortality

A range of adverse events were reported by Demedts and colleagues[67] and these are summarised in *Table 21*. The study report that none of the differences between the study groups were significant, except for those relating to bone marrow toxicity, which occurred in 4% of participants receiving triple therapy and in 13% receiving double therapy; $p = 0.03$. Detailed descriptions of the types of events included in these categories were reported but have not been data extracted. There were seven deaths in the triple-therapy arm and eight in the dual-therapy arm (9% vs. 11%, respectively) of the Demedts and colleagues[67] study and this was not statistically significant between the two groups ($p = 0.69$).

The Raghu and colleagues IPFCRN 2012 group report data on the rate of serious adverse events in the triple-therapy and placebo arms of their trial.[68] These can be seen in *Table 22*, where it can be seen that the event rate of any serious adverse event was 31% in the triple-therapy arm, compared with 10% in the placebo arm ($p = 0.001$). Respiratory system serious adverse events contributed the most to the difference observed, and were also seen to be statistically significantly higher in the participants in the triple-therapy arm, compared with the placebo arm ($p = 0.03$). All other serious events recorded in the study were not statistically significantly different between the two groups.

TABLE 21 Adverse events: Demedts *et al.*[67]

Adverse events occurring in at least 5% of participants	Number of patients (%), number of events	Number of patients (%), number of events
	NAC, azathioprine, prednisolone ($N = 80$)	Placebo, azathioprine, prednisolone ($N = 75$)
All adverse events	72 (90), 322	67 (89), 303
Respiratory tract infection	20 (25), 22	24 (32), 27
Dyspnoea	16 (20), 16	19 (25), 21
Fever	15 (19), 17	10 (13), 10
Liver-function test abnormal	14 (18), 15	11 (15), 13
Cough	13 (16), 15	16 (21), 17
Abdominal pain	12 (15), 12	7 (9), 7
URTI	11 (14), 11	13 (17), 15
Blood glucose ↑	9 (11), 9	11 (15), 12
C-reactive protein ↑	6 (8), 7	3 (4), 3
Blood alkaline phosphatase ↑	6 (8), 6	1 (1), 1
Blood lactate dehydrogenase ↑	6 (8), 6	2 (3), 2
Back pain	6 (8), 6	5 (7), 6
Respiratory failure	5 (6), 5	1 (1), 1
Bone marrow toxic effects	3 (4), 3[a]	10 (13), 10[a]
Oedema	3 (4), 3	5 (7), 5
Headache	3 (4), 4	6 (8), 6
Asthenia	3 (4), 3	5 (7), 5
Influenza-like illness	3 (4), 3	5 (7), 5
Muscle cramp	1 (1), 1	4 (5), 4
Deaths	7 (9%)	8 (11%)[b]

a　$p = 0.03$.
b　$p = 0.69$.

TABLE 22 Serious adverse events: IPFCRN 2012[68]

Serious adverse event rate	NAC, azathioprine, prednisolone, n (%) (N = 77)	Placebo, n (%) (N = 78)	p-value
Any	24 (31)	8 (10)	0.001
Respiratory system	12 (16)	4 (5)	0.03
Infectious	5 (6)	1 (1)	0.12
Gastrointestinal	1 (1)	3 (4)	0.62
Cardiac	3 (4)	0	0.12
General disorder[a]	3 (4)	0	0.12
Neoplasm	2 (3)	0	0.25
Metabolism	1 (1)	0	0.50
Musculoskeletal system	0	1 (1)	1.00
Nervous system	1 (1)	0	0.50
Reproductive system	1 (1)	0	0.50

a Includes all serious adverse events that did not fall into another body-system category, including drug reactions and drug fever.

Specific adverse events with a significant between-group difference from the Raghu and colleagues IPFCRN 2012[68] trial can be seen in *Table 23*. Overall, there was a higher proportion of individuals with an adverse event in the triple-therapy arm than in the placebo arm ($p = 0.09$), with rates of general disorder, skin disorder and renal and urinary system disorders also being recorded more frequently in the triple-therapy arm than in the placebo arm. All other adverse events, which were labelled as non-serious and had non-significant between-group differences, were reported by the study publication and can be seen in *Appendix 5*. The hospitalisation rates (all-cause) were also seen to be statistically significantly different between the two groups, with 23 (30%) of the triple-therapy participants and seven (9%) of the placebo participants being hospitalised during the 60 weeks of the study; $p < 0.001$.

The Raghu and colleagues IPFCRN 2012 trial[68] reported mortality outcomes under a number of different categories (*Table 24*). All-cause mortality was statistically significantly greater in the triple-therapy-treated participants (10%) than the placebo-treated participants (1%); $p = 0.01$. A similar pattern was observed in the rates of respiratory mortality (9% triple therapy, 1% placebo; $p = 0.02$). HRs for estimated any-cause mortality and any-cause mortality or hospitalisation were also seen to favour placebo (see *Table 24*). Estimated any-cause mortality or at least a 10% decline in FVC was similar across the two groups.

TABLE 23 Adverse events: IPFCRN 2012[68]

Adverse events[a]	NAC, azathioprine, prednisolone, n (%) (N = 77)	Placebo, n (%) (N = 78)	p-value
Any	68 (88)	61 (78)	0.09
General disorder	34 (44)	21 (27)	0.03
Skin	13 (17)	4 (5)	0.02
Renal and urinary system	10 (13)	1 (1)	0.005

a Specific adverse events with a significant between-group difference. For other adverse events, see *Appendix 5*.

TABLE 24 Mortality and hospitalisation outcomes: IPFCRN 2012[68]

Mortality and hospitalisation outcomes	NAC, azathioprine, prednisolone (*n* = 77)	Placebo (*n* = 78)	*p*-value
All-cause mortality, *n* (%)	8 (10)	1 (1)	0.01
Respiratory mortality, *n* (%)	7 (9)	1 (1)	0.02
Estimated any-cause mortality at 60 weeks, % (95% CI)	19.8 (9.9 to 37.2)	2.0 (0.3 to 13.6)	HR 9.26 (1.16 to 74.1); 0.01
Estimated any-cause mortality or hospitalisation, % (95% CI)	43.6 (30.7 to 59.0)	16.9 (8.7 to 31.5)	HR 3.74 (1.68 to 8.34); < 0.001
Estimated any-cause mortality or ≥ 10% decline in FVC, % (95% CI)	36.3 (23.7 to 53.0)	32.4 (19.7 to 50.3)	HR 1.46, (0.70 to 3.05); 0.30
All-cause hospitalisations, *n* (%)	23 (30)	7 (9)	< 0.001

Frequently recorded adverse events in the study by Homma and colleagues[69] can be seen in *Table 25*. The paper states that there were no significant differences in the number of adverse events reported for the two groups. The severity of the events was less than grade 2. No adverse events were recorded in participants in the control group. There were some grade 1 and grade 2 adverse events (bacterial pneumonia, cough, sore throat and hypercholesterolaemia) in the inhaled NAC group.

Summary

Triple therapy did not confer any improvement on the FVC compared with placebo in one study but did confer an improvement on the VC when compared with double therapy in one other study. Inhaled single-therapy NAC did not confer a statistically significant benefit over a control intervention in a third included study. Secondary outcomes including DL_{CO} and 6MWT were reported with mixed results. There were no differences in rates of dyspnoea in the three studies. QoL did not appear to differ between participants treated with triple therapy when compared with placebo, and acute exacerbations were greater with triple therapy but not significantly. Triple therapy led to more deaths, serious adverse events and adverse events than placebo in one trial. In the other two studies, adverse events did not differ significantly between randomised groups. Of the three studies, two had a low risk of bias; however, the third study by Homma and colleagues[69] had an unclear risk of bias and therefore results should be interpreted with caution.

TABLE 25 Adverse events: Homma *et al.*[69]

Adverse events	NAC (*n* = 44)	Control (*n* = 46)
Bacterial pneumonia		
Grade 1	2	0
Grade 2	2	0
Cough		
Grade 1	1	0
Grade 2	1	0
Sore throat		
Grade 1	2	0
Grade 2	0	0
Hypercholesterolaemia		
Grade 1	2	0
Grade 2	0	0

Pirfenidone

Time to progression

Time to progression (TTP) was calculated in the two Capacity trials[62] (*Table 26*). TTP was defined as the time to 'worsening' of IPF, with no further details provided. The HR for the estimated TTP was not statistically significant in the Capacity study 006 (HR 0.73, 95% CI 0.43 to 1.24; $p = 0.248$) or in the Capacity study 004 (HR 0.84, 95% CI 0.50 to 1.42; $p = 0.515$), with no benefit of pirfenidone seen.[62]

Progression-free survival

Progression-free survival was defined in the two Capacity trials[62] as time to confirmed $\geq 10\%$ decline in per cent predicted FVC, $\geq 15\%$ decline in per cent predicted DL_{CO}, or death. The HR for estimated PFS was not statistically significant in the Capacity study 006 (HR 0.84, 95% CI 0.58 to 1.22; $p = 0.355$) but was statistically significant in the Capacity study 004 [HR 0.64 (95% CI 0.44 to 0.95); $p = 0.023$], indicating a 36% reduction in the risk of death or disease progression with pirfenidone treatment (*Table 27*).[62]

TABLE 26 Time to progression: pirfenidone

Study details	HR (95% CI)
Noble *et al.* 2011[62]	0.73 (0.43 to 1.24); $p = 0.248$
Capacity study 006	
Noble *et al.* 2011[62]	0.84 (0.50 to 1.42); $p = 0.515$
Capacity study 004	
Time to progression defined as time to 'worsening' of IPF – no further details provided.	

TABLE 27 Progression-free survival: pirfenidone

Study details	Estimated PFS	
	Arm	*n/N*
Noble *et al.* 2011[62]	*Pirfenidone*	126/171
Capacity study 006[a]	*Placebo*	123/171
	HR, *p*-value	HR 0.84 (0.58 to 1.22); $p = 0.355$
Noble *et al.* 2011[62]	*Pirfenidone*	138/174
Capacity study 004[a]	*Placebo*	116/174
	HR, *p*-value	HR 0.64 (0.44 to 0.95); $p = 0.023$
[b]Taniguchi *et al.* 2010[65]	*Pirfenidone high dose*	45/106
	Placebo	40/104
	HR, *p*-value	HR not reported, HD $p = 0.0280$

a PFS defined as time to confirmed $\geq 10\%$ decline in per cent predicted FVC, $\geq 15\%$ decline in per cent predicted DL_{CO} or death. Kaplan–Meier plots available.
b PFS defined as death and/or $\geq 10\%$ decline in VC from baseline, or when VC data not obtainable (due to worsening respiratory symptoms). Kaplan–Meier plot available.

In the Taniguchi and colleagues[65] trial, PFS was defined as death or a $\geq 10\%$ decline in VC from baseline. In this study the HRs were not reported. The publication states that the HR for the high-dose pirfenidone group versus placebo was statistically significantly ($p = 0.03$) in favour of treatment (the HR for the low-dose pirfenidone group vs. placebo was not). There is an uncertain risk of bias in this study, which should be taken into account when interpreting these data.

Forced vital capacity

The two Capacity trials[62] report the mean change in per cent predicted FVC at 72 weeks (see *Table 27*). In the Capacity 006 trial, the difference between groups (0.6%) was not statistically significant ($p = 0.501$).[62] In the Capacity 004 trial, the difference of 4.4% between groups was shown to be statistically significantly different ($p = 0.001$).[62] Combining these two trials in a random-effects meta-analysis (see *Figure 2*) demonstrated no statistical difference overall (mean difference 2.61, 95% CI –1.11 to 6.33; $p = 0.17$) and moderate statistical heterogeneity between the trials ($I^2 = 45\%$). Taniguchi and colleagues[65] reported the mean change in VC litres at 52 weeks. In this study the difference in change in VC between the high-dose pirfenidone group (1800 mg/day) and placebo was seen to be statistically significant (difference 0.07 litres; $p = 0.04$). In the RCT by Azuma and colleagues[66] at 6 months the mean change in VC litres was not statistically significantly different between those treated with pirfenidone 1800 mg/day and those treated with placebo (difference 0.07; $p = 0.10$) (see *Table 27* for full details). At 9 months a statistically significant difference between groups was seen, with the mean change in VC between groups being 0.10 litres ($p = 0.04$). Combining these two trials in a fixed-effects meta-analysis (see *Figure 4*) showed a statistically significant benefit of treatment with pirfenidone compared with placebo (mean difference 0.08, 95% CI 0.03 to 0.12; $p = 0.0006$) and no statistical heterogeneity between the trials ($I^2 = 0\%$).

The FVC/VC outcomes (both per cent predicted FVC/VC and mean change in FVC/VC) from the four included pirfenidone studies were combined in a random-effects meta-analysis using the standardised mean difference (*Figure 6*). A statistically significant benefit of treatment with pirfenidone (standardised mean difference 0.24, 95% CI 0.06 to 0.41; $p = 0.008$) was found. Moderate statistical heterogeneity was present, as indicated by the I^2 test (45%). The uncertain risk of bias in the Azuma and colleagues[66] and Taniguchi and colleagues[65] studies should be noted when interpreting the effects shown.

The Capacity trials[62] also reported the proportion of participants with a decline in FVC of at least 10% (*Table 28* and *Figure 5*). It is unclear if this was calculated as an absolute or relative change. In the Capacity 006 study, 23% of participants treated with pirfenidone and 27% of participants treated with placebo showed a decline of at least 10% in the per cent predicted FVC. Statistical analysis was undertaken on five categories of change in FVC per cent predicted in this study (severe decline $\geq 20\%$; moderate decline $< 20\%$ but $\geq 10\%$; mild decline $< 10\%$ but ≥ 0; mild improvement > 0 but $< 10\%$; moderate improvement $\geq 10\%$) and this showed no statistically significant differences between groups in the proportion of participants within these categories. In Capacity 004, 20% of participants in the pirfenidone group and 35% of participants in the placebo group showed a decline on the per cent predicted FVC of at least 10%. In the analysis of all five categories in this study, a statistically significant effect of pirfenidone was seen ($p = 0.01$). It is unclear whether or not these subgroup analyses were adequately powered to detect a difference and, therefore, caution is required in their interpretation.

Azuma and colleagues[66] also report VC litres in a categorical analysis, using three categories: 'improved', 'stable' and 'deteriorated' (see *Figure 5*). Improvement was defined as at least a 10% improvement in VC, deteriorated was defined as at least a 10% decrease in VC, and any other value was categorised as stable. It is unclear whether this was calculated as an absolute or relative change. At 9 months a statistically significant difference between groups was observed on this categorical analysis ($p = 0.003$). The proportion of participants with an improvement in their VC was 9% in the pirfenidone group and 0% in the placebo group. The proportion with a decline in their VC was 13% in the pirfenidone group and 36% in the

Study or subgroup	Pirfenidone			Placebo			Weight	Standardised mean difference IV, random, 95% CI	Standardised mean difference IV, random, 95% CI
	Mean	SD	Total	Mean	SD	Total			
Azuma 2005[66]	−0.03	0.22	72	−0.13	0.19	35	13.9%	0.47 (0.06 to 0.88)	
Capacity 004[62]	−8	16.5	174	−12.4	18.5	174	31.2%	0.25 (0.04 to 0.46)	
Capacity 006[62]	−9	19.6	171	−9.6	19.1	173	31.1%	0.03 (−0.18 to 0.24)	
Taniguchi 2010[65]	−0.09	0.2	104	−0.16	0.2	103	23.8%	0.35 (0.07 to 0.62)	
Total (95% CI)			**521**			**485**	**100.0%**	**0.24 (0.06 to 0.41)**	

Heterogeneity: $\tau^2 = 0.01$; $\chi^2 = 5.44$, df = 3 ($p = 0.14$); $I^2 = 45\%$
Test for overall effect: $z = 2.64$ ($p = 0.008$)

FIGURE 6 Meta-analyses of pirfenidone trials: FVC/VC mean change and per cent predicted mean change outcomes at end of study.

TABLE 28 Forced vital capacity outcomes: pirfenidone

Study	Outcome	Interventions		Difference between groups (95% CI)
		Pirfenidone 2403 mg/day (*n* = 171)	Placebo (*n* = 173)	
Noble *et al.* 2011[62]	Mean change per cent predicted FVC at 72 weeks	–9.0% (SD 19.6)	–9.6% (SD 19.1)	0.6% (–3.5 to 4.7); *p* = 0.501
Capacity study 006	Proportion with a decline in FVC ≥ 10%	39 (23%)	46 (27%)	3.8 (–2.7 to 10.2); *p* = 0.440[a]
		Pirfenidone 2403 mg/day (*n* = 174)	Placebo (*n* = 174)	
Noble *et al.* 2011[62]	Mean change per cent predicted FVC at 72 weeks	–8.0% (SD 16.5)	–12.4% (SD 18.5)	4.4% (0.7 to 9.1); *p* = 0.001
Capacity study 004	Proportion with a decline in FVC ≥ 10%	35 (20%)	60 (35%)	14.4 (7.4 to 21.3); *p* = 0.01[a]
		High-dose 1800 mg pirfenidone (*n* = 108)	Placebo (*n* = 104)	
Taniguchi *et al.* 2010[65]	Adjusted (LOCF) mean change in VC, l (SE) at 52 weeks	*n* = 104 –0.09 (0.02)	*n* = 103 –0.16 (0.02)	0.07 (0.03); *p* = 0.0416
	Unadjusted mean change in VC, l (SE)	*n* = 67 2.36 (0.73)	*n* = 72 2.42 (0.75)	Not tested
		Pirfenidone 1800 mg (*n* = 72)	Placebo (*n* = 35)	
Azuma *et al.* 2005[66]	Mean (SD) change VC, l, at 6 months	–0.01 (0.21)	–0.08 (0.19)	0.07;[b] *p* = 0.0995
	Mean (SD) change VC, l, at 9 months	–0.03 (0.22)	–0.13 (0.19)	0.10;[b] *p* = 0.0366
	Categorical analysis,[c] 9 months	*n* = 67	*n* = 33	
	Improved	6 (9%)	0	*p* = 0.0028
	Stable	52 (78%)	21 (64%)	
	Deteriorated	9 (13%)	12 (36%)	

SE, standard error.
a *p*-value from an analysis of five categories: severe decline (≥ 20%), moderate decline (< 20% but ≥ 10%), mild decline (< 10% but ≥ 0%), mild improvement (> 0% but < 10%) and moderate improvement (≥ 10%).
b Calculated by reviewer.
c Changes of 10%.

placebo group, and those classified as stable were 78% in the pirfenidone group and 64% in the placebo group. It is unclear whether or not the analysis of these data was adequately powered, and in addition, the analysis was not undertaken on the ITT population (*n* = 67 out of 72 pirfenidone and *n* = 33 out of 35 placebo) and caution is, therefore, required in the interpretation of these data.

Combining the three trials in a random-effects meta-analysis demonstrated a significant effect in favour of pirfenidone on the proportion of people with a decline in FVC of ≥ 10% (RR 0.62, 95% CI 0.41 to 0.93; *p* = 0.02 see *Figure 5*). Moderate statistical heterogeneity was present between the trials (I^2 = 57%).

Diffusing capacity of the lung for carbon monoxide

All four included pirfenidone trials report data on the DL_{CO} at end point (*Table 29*). In the Capacity 006 study[62] the mean change from baseline in the per cent predicted DL_{CO} at 72 weeks was not statistically significant between the two groups (–9.8% vs. –9.2%, pirfenidone vs. placebo, respectively; $p = 0.996$). This was also the case in the Capacity 004 study[62] (mean change from baseline –7.9% vs. –9.9%, pirfenidone vs. placebo, respectively; $p = 0.145$), and when combining these two trials in a meta-analysis (mean difference 0.68, 95% CI –0.87 to 3.22; $p = 0.6$; see *Figure 3*). The mean change in DL_{CO} ml/minute/ mmHg in the Taniguchi and colleagues[65] trial was also not statistically significantly different between the high-dose pirfenidone and placebo arms ($p = 0.23$) after 52 weeks. Finally, mean change in DL_{CO} ml/minute/mmHg was not statistically significant between the pirfenidone-treated participants and the placebo-treated participants at either 6 months ($p = 0.49$) or 9 months ($p = 0.21$) in the Azuma and colleagues trial.[66] However, combining these last two trials in a meta-analysis produces a beneficial effect in favour of pirfenidone that approaches statistical significance (mean difference 0.53, 95% CI –0.00 to 1.05; $p = 0.05$; *Figure 7*).

A categorical analysis of DL_{CO} was also reported by Azuma and colleagues.[66] Three categories, 'improved', 'stable', and 'deteriorated', were used, where a change of 15% was used to categorise either improvement (+15%) or deterioration (–15%), and all other values were categorised as stable. At 9 months no statistically significant difference between groups was observed on the categorical analysis ($p = 0.158$). The proportion of participants with an improvement in their DL_{CO} was 16% in the pirfenidone group and 6% in the placebo group. The proportion with a decline in their DL_{CO} was 36% in the pirfenidone group and 47% in the placebo group, and those classified as stable were 50% in the pirfenidone group and 47% in the placebo group. Caution is required in the interpretation of these data.

TABLE 29 Diffusing capacity of the lung for carbon monoxide outcomes: pirfenidone

Study	Outcome	Interventions		Difference between groups
		Pirfenidone 2403 mg/day (*n* = 171)	**Placebo (*n* = 173)**	
Noble *et al.* 2011[62] Capacity study 006	Mean change in DL_{CO} (% predicted)	–9.8	–9.2	–0.5 (–3.2 to 2.2); $p = 0.99$
		Pirfenidone 2403 mg/day (*n* = 174)	**Placebo (*n* = 174)**	
Noble *et al.* 2011[62] Capacity study 004	Mean change in DL_{CO} (% predicted)	–7.9	–9.9	2.0 (–0.4 to 4.4); $p = 0.15$
		High-dose 1800 mg pirfenidone (*n* = 108)	**Placebo (*n* = 104)**	
Taniguchi *et al.* 2010[65]	Change in DL_{CO}	$n = 96$ –0.88	$n = 98$ –1.36	$p = 0.24$
		Pirfenidone 1800 mg (*n* = 72)	**Placebo (*n* = 35)**	
Azuma *et al.* 2005[66]	Mean (SD) change DL_{CO} ml/minute/mmHg at 6 months	–0.50 (2.07)	–0.83 (2.16)	$p = 0.49$
	Mean (SD) change DL_{CO} ml/minute/mmHg at 9 months	–0.57 (2.15)	–1.19 (2.30)	$p = 0.21$
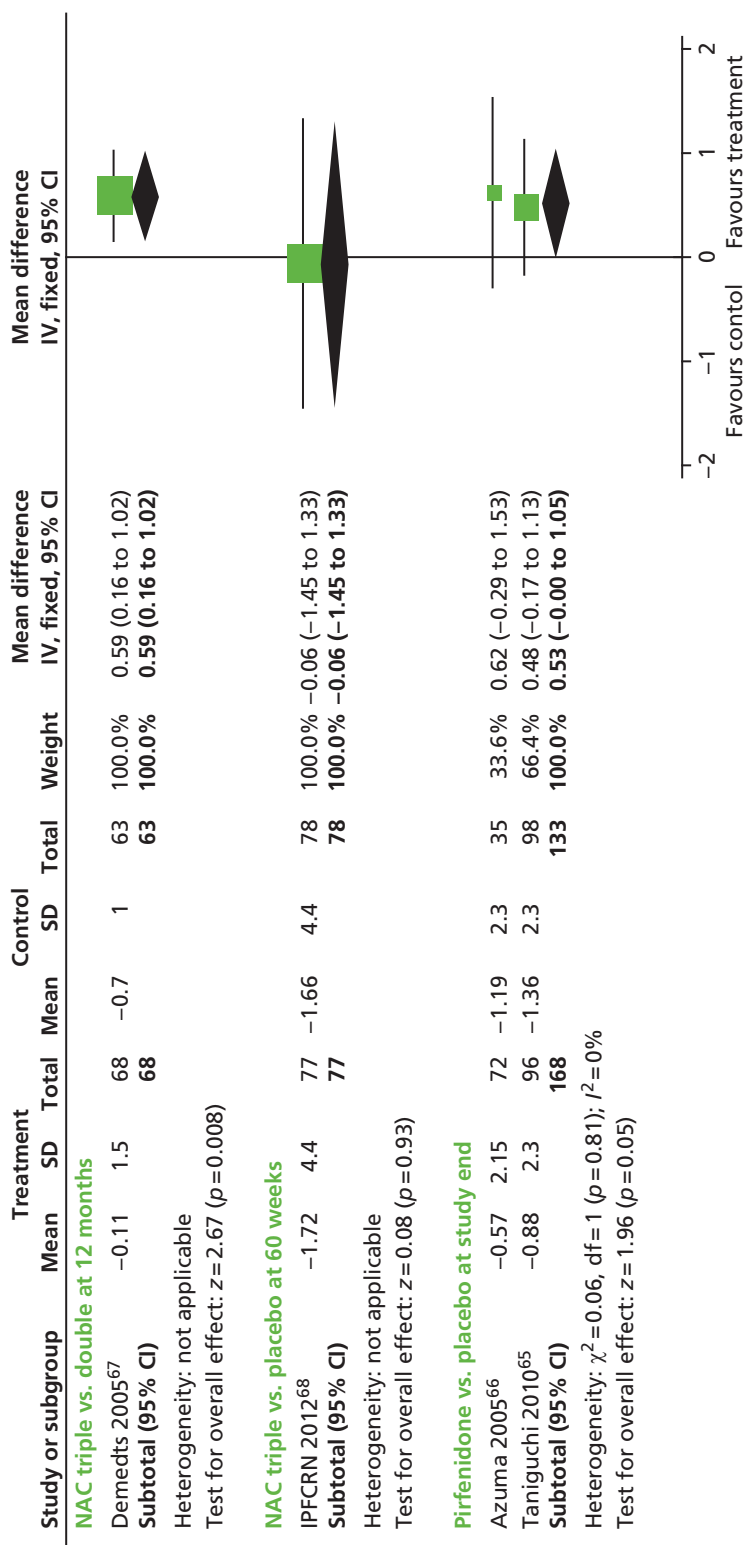
FIGURE 7 Forest plot of mean change in DL_{CO} from trials of pharmacological interventions.

Six-minute walk test

The mean change on the distance walked on the 6MWT was reported in the two Capacity RCTs (*Table 30*). In the Capacity study 006, a statistically significant treatment effect for pirfenidone was seen at 72 weeks, with a 31.8 m difference in the distance walked between the two groups seen ($p = 0.0009$). In the Capacity study 004, there was no statistically significant effect of treatment with pirfenidone (difference in the distance walked 16.4 m; $p = 0.171$).

All four pirfenidone studies report data on the lowest SpO_2 during the 6MWT (*Table 31* and *Figure 8*). In neither of the Capacity studies was a statistically significant difference between groups observed at 72 weeks (Capacity 006: difference –0.5, $p = 0.893$; Capacity 004: difference 0.8, $p = 0.087$).

TABLE 30 Six-minute walk test outcomes: pirfenidone

Study	Outcome	Interventions		Difference between groups
		Pirfenidone 2403 mg/day (*n* = 171)	**Placebo (*n* = 173)**	
Noble *et al.* 2011[62]	6MWT at baseline	378.0 (82.2)	399.1 (89.7)	
Capacity study 006	Mean change in 6MWT distance, m	–45.1	–76.9	31.8 (3.2 to 60.4); $p = 0.0009$
		Pirfenidone 2403 mg/day (*n* = 174)	**Placebo (*n* = 174)**	
Noble *et al.* 2011[62]	6MWT at baseline	411.1 (91.8)	410.0 (90.9)	16.4 (–10.9 to 43.7); $p = 0.17$
Capacity study 004				

TABLE 31 Lowest SpO_2 during the 6MWT, % pirfenidone

Study	Outcome	Interventions		Difference between groups
		Pirfenidone 2403 mg/day (*n* = 171)	**Placebo (*n* = 173)**	
Noble *et al.* 2011[62]	Mean change in worst SpO_2 during 6MWT, %	–1.9	–1.3	–0.5 (–1.7 to 0.7); $p = 0.89$
Capacity study 006				
		Pirfenidone 2403 mg/day (*n* = 174)	**Placebo (*n* = 174)**	
Noble *et al.* 2011[62]	Mean change in worst SpO_2 during 6MWT, %	–1.5	–2.3	0.8 (–0.2 to 1.8); $p = 0.09$
Capacity study 004				
		High dose 1800 mg pirfenidone (*n* = 108)	**Placebo (*n* = 104)**	
Taniguchi *et al.* 2010[65]	Change in lowest SpO_2 during the 6MET (SE)	*n* = 99 –1.70 (0.35)	*n* = 100 –1.53 (0.35)	–0.17 (0.50); $p = 0.74$
		Pirfenidone 1800 mg (*n* = 72)	**Placebo (*n* = 35)**	
Azuma *et al.* 2005[66]	Mean (SD) change lowest SpO_2 during 6MET, % at 6 months	0.6364 (3.5502)	–0.5484 (3.7933)	$p = 0.15$
	Mean (SD) change lowest SpO_2 during 6MET, % at 9 months	0.4697 (3.8838)	–0.9355 (3.3559)	$p = 0.07$

6MET, 6-minute exercise test; SE, standard error.

Study or subgroup	Treatment			Control			Weight	Mean difference IV, random, 95% CI	Mean difference IV, random, 95% CI
	Mean	SD	Total	Mean	SD	Total			
Pirfenidone vs. placebo at study end									
Azuma 2005[66]	0.4697	3.9	72	−0.9355	3.4	35	20.7%	1.41 (−0.04 to 2.85)	
Capacity 004[62]	−1.5	4.3	174	−2.3	4.3	174	34.7%	0.80 (−0.10 to 1.70)	
Capacity 006[62]	−1.9	34.4	171	−1.3	34.4	173	1.2%	−0.60 (−7.87 to 6.67)	
Taniguchi 2010[65]	−1.7	3.5	108	−1.53	0.35	100	43.4%	−0.17 (−0.83 to 0.49)	
Subtotal (95% CI)			**525**			**482**	**100.0%**	**0.49 (−0.32 to 1.30)**	

Heterogeneity: $\tau^2 = 0.28$; $\chi^2 = 5.41$, df = 3 ($p = 0.14$); $I^2 = 45\%$
Test for overall effect: $z = 1.18$ ($p = 0.24$)

FIGURE 8 Meta-analysis of pirfenidone trials: mean change in worst SpO_2 during 6MWT.

In the Taniguchi and colleagues[65] study, after 52 weeks no statistically significant differences between the high-dose pirfenidone group and placebo ($p = 0.74$) was observed. The change in the lowest SpO_2 on the 6MWT was the primary outcome in the Azuma and colleagues[66] trial. No statistically significant differences were observed when analysed at 6 months' ($p = 0.148$) or at 9 months' ($p = 0.722$) follow-up.

Combining the four trials in a random-effects meta-analysis found no significant difference in SpO_2 during the 6MWT (mean difference 0.49, 95% CI –0.32 to 1.30; $p = 0.24$). Moderate statistical heterogeneity was evident ($I^2 = 45\%$) between the trials (see *Figure 8*).

Azuma and colleagues[66] also report a categorical analysis (improved, stable, deteriorated) of the SpO_2 during the 6MWT at 9 months. A > 4% increase was classified as 'improved', a ≤ 4% decrease was classified as 'deteriorated' and any values in between these two was classified as 'stable'. Analysis of these data showed a statistically significant treatment effect in favour of pirfenidone ($p = 0.02$), although caution is required as this was unlikely to be a powered analysis, and the population was not the ITT population. Treatment with pirfenidone led to an improvement in 24% of cases, compared with 6% of cases in the placebo group. More participants deteriorated in the placebo group (18% pirfenidone, 33% placebo); however, the rates of participants classified as having a stable SpO_2 during the 6MWT were similar (58% vs. 61%, pirfenidone vs. placebo group, respectively).

Other secondary outcomes

Other outcomes reported in these studies include TLC, partial pressure of oxygen and HRCT outcome, the results of which can be found in *Appendix 5*.

Dyspnoea

Dyspnoea was measured by the University of California, San Diego, Shortness-of-Breath Questionnaire (UCSDSBQ) in the two Capacity trials, the results of which can be seen in *Table 32*.[62] The UCSDSBQ indicates severity of dyspnoea on a scale from 0 to 5 on 21 activities of daily living, along with three ratings on limitations caused by dyspnoea or fear of dyspnoea, for a total score ranging from 0 to 120, with a higher score indicating more dyspnoea. There were no statistically significant differences between the intervention and placebo groups on this outcome in either study ($p = 0.604$ Capacity 006; $p = 0.509$ Capacity 004). Dyspnoea was also reported by Azuma and colleagues;[66] however, the measure used was not described and no data were reported except the p-value which indicated no statistical significant difference between groups ($p = 0.64$).

TABLE 32 Dyspnoea outcomes: pirfenidone

		Interventions		
		Pirfenidone 2403 mg/day (*n* = 171)	Placebo (*n* = 173)	
Study	**Outcome**			**Difference between groups**
Noble *et al.* 2011[62] Capacity study 006	Mean change in UCSDSBQ score[a]	11.9	13.9	–2.0 (–7.6 to 3.6); $p = 0.60$
		Pirfenidone 2403 mg/day (*n* = 174)	**Placebo (*n* = 174)**	
Noble *et al.* 2011[62] Capacity study 004	Mean change in UCSDSBQ score[a]	12.1	15.2	–3.1 (–8.5, 2.3); $p = 0.51$
		Pirfenidone (*n* = 72)	**Placebo (*n* = 35)**	
Azuma *et al.* 2005[66]	Dyspnoea (not defined)	Data not reported	Data not reported	$p = 0.64$

a UCSDSBQ total score ranges from 0 to 120, with higher scores indicating greater shortness of breath.

Acute exacerbations

The rate of acute exacerbations was reported by two of the pirfenidone studies (*Table 33*). At 52 weeks Azuma and colleagues[66] demonstrated a statistically significant difference in the rate of acute exacerbations in those treated with pirfenidone and those treated with placebo, in favour of pirfenidone (0% vs. 14%, respectively; $p = 0.003$). Taniguchi and colleagues[65] also reported the incidence of acute exacerbation but found no difference between the high-dose pirfenidone and placebo groups at 52 weeks (see *Table 33*).

Quality of life

Only the study by Azuma and colleagues[66] reported having measured QoL; however, no data for this outcome were reported in the trial publication. The paper stated that there was no statistically significant difference between groups ($p = 0.8720$).

Adverse events

Serious adverse events were recorded in the Capacity RCTs[62] and can be seen in *Table 34*, which shows IPF-related serious adverse events occurring in at least two participants in any treatment group. The rate of any serious adverse event was observably similar between those treated with pirfenidone and those treated with placebo in the Capacity study 006[62] (pirfenidone 7.6% vs. 9.8% placebo), although no statistical testing of these rates were reported. All other serious adverse events reported in the study appear to be comparable between the treated and untreated study populations (see *Table 34*). A similar pattern was seen in the treatment arms of the Capacity 004 study (see *Table 34*).

TABLE 33 Acute exacerbations: pirfenidone

| | | Interventions | | |
| | | Pirfenidone 1800 mg (*n* = 72) | Placebo (*n* = 35) | Difference between groups |
Study	**Outcome**			
Azuma *et al.* 2005[66]	Acute exacerbation	0	5/35 (14%)	14%; $p = 0.0031$
		High-dose 1800 mg pirfenidone (*n* = 108)	**Placebo (*n* = 104)**	
Taniguchi *et al.* 2010[65]	Acute exacerbation rate	6 (5.6%)	5 (4.8%)	NS

NS, not significant.

TABLE 34 Serious adverse events for pirfenidone occurring in ≥ two people

SAEs	Pirfenidone, *n* (%) (*N* = 171)	Placebo, *n* (%) (*N* = 173)
Capacity 006[62]		
IPF	13 (7.6)	17 (9.8)
Pneumonia	7 (4.1)	7 (4.0)
Respiratory failure	4 (2.3)	6 (3.5)
Coronary artery disease	6 (3.5)	0
Acute respiratory failure	2 (1.2)	3 (1.7)
Bronchitis	0	5 (2.9)
Atrial fibrillation	2 (1.2)	1 (0.6)
Renal failure	2 (1.2)	2 (1.2)
Fall	2 (1.2)	1 (0.6)

TABLE 34 Serious adverse events for pirfenidone occurring in ≥ two people (*continued*)

SAEs	Pirfenidone, *n* (%) (*N* = 171)	Placebo, *n* (%) (*N* = 173)
Hypotension	2 (1.2)	1 (0.6)
Prostate cancer[a]	2 (1.6)	0
Colitis	2 (1.2)	0
Hip fracture	2 (1.2)	0
Hypertension	0	2 (1.2)
Hypoxia	0	2 (1.2)
Intervertebral disc protrusion	2 (1.6)	0
Liver function test abnormal	2 (1.2)	0
Nephrolithiasis	2 (1.2)	0
Sick sinus syndrome	2 (1.2)	0
Transitional cell carcinoma	0	2 (1.2)
	Pirfenidone 2403 mg/day, *n* (%) (*N* = 174)	Placebo, *n* (%) (*N* = 174)
Capacity 004[62]		
IPF	13 (7.5)	14 (8.0)
Pneumonia	4 (2.3)	6 (3.4)
Respiratory failure	2 (1.1)	2 (1.1)
Bronchitis	2 (1.1)	2 (1.1)
Lobar pneumonia	2 (1.1)	2 (1.1)
Myocardial infarction	0	4 (2.3)
Acute respiratory failure	2 (1.1)	3 (1.7)
Angina pectoris	2 (1.1)	1 (0.6)
Atrial fibrillation	1 (0.6)	1 (0.6)
Coronary artery disease	0	2 (1.1)
Pneumothorax	3 (1.7)	0
Pulmonary embolism	1 (0.6)	1 (0.6)
Syncope	3 (1.7)	1 (0.6)
Dyspnoea	0	3 (1.7)
Non-cardiac chest pain	2 (1.1)	2 (1.1)
Prostate cancer[a]	0	2 (1.6)
Aortic aneurysm	2 (1.1)	0
Chest pain	3 (1.7)	0
Hypoxia	1 (0.6)	2 (1.1)
Acute renal failure	1 (0.6)	0
Bladder cancer	2 (1.1)	0
Gastro-oesophageal reflux disease	2 (1.1)	0

SAE, serious adverse event.
a Percentage based on number of males in each treatment group.

All four included studies report details of any adverse events in their respective trials. In both Capacity trials,[62] adverse events occurring in 10% or more of participants receiving pirfenidone and an incidence of 1.5 times or greater in placebo were reported (*Table 35*). The authors state that events were generally mild or moderate, without clinical consequences. Rates of any adverse events were seen to be similar between treatment group and placebo in both studies. Nausea, fatigue, diarrhoea and rash were the most frequently reported adverse events in the pirfenidone-treated participants (incidence in the region of 25–38% in both Capacity studies). Rates of these adverse events were in the region of 13% to 21% in the placebo-treated participants (for full details see *Table 35*). No statistical analyses of the rates between the groups were reported. All other adverse events appeared to be similar across groups with the exception of dyspepsia, which was seen to be more widely recorded in the pirfenidone groups.

Adverse events recorded in $\geq 5\%$ participants in either the high-dose pirfenidone or the placebo group in the Taniguchi and colleagues trial[65] can be seen in *Table 36*. The rate of any adverse events was similar across the pirfenidone treatment arm and the placebo arm (high dose vs. placebo $p = 0.50$). However, there were some specific adverse events that were seen to be statistically significantly different between treatment and placebo groups. Photosensitivity was seen in 51.4% of high-dose pirfenidone participants, compared with 22.4% of placebo participants ($p < 0.01$). Anorexia, dizziness and raised glutamyl-transpeptidase were all higher in those treated with high-dose pirfenidone than in those treated with placebo (see *Table 36* for details). Two adverse events were more common in the placebo group than the high-dose pirfenidone group: nasopharyngitis and URTIs (*Table 36*). The report also provides details of statistical analyses of the rates between the two pirfenidone groups, which can be seen in *Appendix 5*. Specific adverse events leading to discontinuations for each treatment group were also reported but not data extracted. The numbers of dropouts for any adverse events were discussed earlier in this chapter (see *Pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *Pirfenidone*). No one particular adverse event appears to have been a more common reason for discontinuation.

Azuma and colleagues[66] reported adverse events occurring in $\geq 10\%$ of participants at 6 months. Rates of any adverse events were statistically significantly higher in the pirfenidone group than in the placebo group ($p = 0.04$) (*Table 37*). Photosensitivity, stomach discomfort, anorexia, nausea, fatigue and raised glutamyl-transpeptidase were all statistically significantly more common with pirfenidone than they were with placebo. All other adverse events were similar between the comparison groups. Reasons for treatment discontinuations were also reported by Azuma and colleagues[66] and can be seen in *Appendix 5*. Although adverse events were more frequently reported to be a reason for discontinuation, none of the specific adverse events or adverse events in general was statistically significantly more likely to lead to discontinuation.

Summary

On measures of FVC, pirfenidone appears to demonstrate an effect when compared with placebo treatment. One trial (Capacity 006[62]) appears to be an outlier with results being non-significant and the reason for this is unclear. Results on other outcomes such as DL_{CO}, 6MWT distance and the oxygen saturation during the 6MWT generally showed no effect of treatment with pirfenidone, with the exception of the 6MWT in the Capacity 006 trial.[62] Acute exacerbations were assessed in two studies, with results favouring pirfenidone in one but not the other study. QoL outcomes were not reported and, therefore, the effect of pirfenidone on QoL cannot be estimated. Adverse event rates were generally similar between those treated with pirfenidone and those treated with placebo, although some specific adverse events (gastrointestinal events and photosensitive skin rashes) appear more likely with pirfenidone. Two studies had low risk of bias, but two had an unclear risk of bias which should be considered when interpreting results.

TABLE 35 Adverse events: pirfenidone (Capacity trials)

AEs	Pirfenidone, *n* (%) (*N* = 171)	Placebo, *n* (%) (*N* = 173)
Capacity 006[62]		
Any AE	169 (99)	170 (98)
Nausea	65 (38)	28 (16)
Fatigue	56 (33)	35 (20)
Diarrhoea	56 (33)	37 (21)
Rash	58 (34)	22 (13)
Dizziness	30 (18)	18 (10)
Dyspepsia	36 (21)	10 (6)
Gastro-oesophageal reflux	10 (6)	12 (7)
Vomiting	23 (14)	8 (5)
Insomnia	12 (7)	11 (6)
Arthralgia	16 (9)	11 (6)
Anorexia	18 (11)	6 (4)
Abdominal distension	18 (11)	8 (5)
Photosensitivity reaction	17 (10)	4 (2)
Urinary tract infection	16 (9)	20 (12)
	Pirfenidone 2403 mg/day, *n* (%) (*N* = 174)	**Placebo, *n* (%) (*N* = 174)**
Capacity 004[62]		
Any AE	171 (98)	169 (97)
Nausea	60 (35)	32 (18)
Fatigue	48 (28)	36 (21)
Diarrhoea	43 (25)	30 (17)
Rash	53 (31)	18 (10)
Dizziness	33 (19)	17 (10)
Dyspepsia	30 (17)	16 (9)
Gastro-oesophageal reflux	26 (15)	14 (8)
Vomiting	24 (14)	7 (4)
Insomnia	22 (13)	12 (7)
Arthralgia	20 (12)	13 (8)
Anorexia	19 (11)	7 (4)
Abdominal distension	15 (9)	12 (7)
Photosensitivity reaction	25 (14)	2 (1)
Urinary tract infection	19 (11)	9 (5)
Stevens–Johnson syndrome	0	0
Toxic epidermal necrosis	0	0

AE, adverse event.

TABLE 36 Adverse events: pirfenidone (Taniguchi et al.[65])

| AEs ≥ 5% | Arm, n (%) | | Difference, p-value |
	High-dose pirfenidone (N = 109[a])	Placebo (N = 107[a])	
Any	109 (100)	106 (99.1)	p = 0.50
Photosensitivity	56 (51.4)	24 (22.4)	p < 0.01
Anorexia	18 (16.5)	3 (2.8)	p < 0.01
Abdominal discomfort	3 (2.8)	0	p = 0.25
Dizziness	8 (7.3)	1 (0.9)	p = 0.04
Nasopharyngitis	54 (49.5)	70 (65.4)	p = 0.02
URTI	1 (0.9)	9 (8.4)	p < 0.01
γ-GTP elevation	25 (22.9)	10 (9.3)	p < 0.01
WBC decrease	4 (3.7)	0	p = 0.12

AE, adverse event; γ-GTP, glutamyl-transpeptidase; WBC, white blood cell.
a Analysis on the full patient cohort including those reported to be ineligible and excluded post randomisation because no post-baseline data were available.

TABLE 37 Adverse events: pirfenidone (Azuma et al.[66])

| AEs occurring in ≥ 10% at 6 months | Arm, n (%) | | p-value |
	Pirfenidone (N = 72)	Placebo (N = 35)	
Any	72 (98.6)	32 (88.9)	p = 0.040
Photosensitivity	32 (43.8)	0	p = 0.000
Stomach discomfort	22 (30.1)	3 (8.3)	p = 0.0143
Anorexia	23 (31.5)	2 (5.6)	p = 0.0030
Nausea	16 (21.9)	2 (5.6)	p = 0.0314
Heartburn	12 (16.4)	1 (2.8)	p = 0.0566
Drowsiness	17 (23.3)	6 (16.7)	p = 0.4672
Fatigue	16 (21.9)	1 (2.8)	p = 0.0102
URTI	12 (16.4)	3 (8.3)	p = 0.3767
Fever	6 (8.2)	4 (11.1)	p = 0.7271
Elevation of GOT	4 (5.5)	6 (16.7)	p = 0.0785
Elevation of γ-GTP	20 (27.4)	3 (8.3)	p = 0.0249
Urinary occult blood positive	6 (8.2)	4 (11.1)	p = 0.7271
Elevation of CRP	15 (20.5)	10 (27.8)	p = 0.4694

AE, adverse event; CRP, C-reactive protein; GOT, glutamic oxaloacetic transaminase; γ-GTP, glutamyl-transpeptidase.

Thalidomide

The randomised crossover trial of thalidomide aimed to determine the efficacy of thalidomide in supressing cough, and as such the outcome measures focused on measurement of cough and QoL.[55] No lung function tests were reported.

Cough and quality of life

Cough measured by VAS and cough-related QoL measured by the CQLQ were statistically significantly improved after 12 weeks of thalidomide compared with placebo (*Table 38*). The authors state that the MCID of the CQLQ in IPF is unknown, but the MCID for the Leicester Cough Questionnaire, which has been shown to be similar to the CQLQ, is 1.3. A mean difference in score of 11.4 on the CQLQ is, therefore, likely to be a clinically important effect. Sensitivity analyses undertaken to explore the effect of missing data found that thalidomide use resulted in a statistically significant improvement in CQLQ scores in each of the analyses. A statistically significant improvement in the total score, symptom domain and impact domain, but not the activity domain, of the SGRQ was also found with thalidomide (see *Table 38*).

Adverse events

Significantly more people experienced at least one adverse event with thalidomide than with placebo (77% vs. 22%; $p = 0.001$) (*Table 39*). The most common adverse events with thalidomide were constipation (36%), dizziness (27%), viral upper respiratory tract infection (URTI) (23%) and malaise (14%). Adverse events requiring dose reduction occurred in three people [constipation in two (9%) and bradycardia in one (5%)].

TABLE 38 Cough and QoL outcomes: thalidomide

Study	Outcome	Interventions		Mean difference between groups (95% CI); *p*-value
		Thalidomide ($n = 23$)	Placebo ($n = 23$)	
Horton *et al.* 2012[55,75]	CQLQ, mean (SD) at baseline	60.5 (12.0)		
	CQLQ, mean (SD) at 12 weeks	47.2 (13.4)	58.7 (14.0)	−11.4 (−15.7 to −7.0); $p < 0.001$
	Cough VAS, mean (SD) at baseline	64.8 (21.4)		
	Cough VAS, mean (SD) at 12 weeks	32.2 (26.1)	61.9 (26.5)	−31.2 (−45.2 to −17.2); $p < 0.001$
	SGRQ total, mean (SD) at baseline	57.4 (18.8)		
	SGRQ total, mean (SD) at 12 weeks	43.9 (16.0)	56.9 (17.1)	−11.7 (−18.6 to −4.8); $p = 0.001$
	SGRQ symptom domain, mean (SD) at baseline	67.7 (19.7)		
	SGRQ symptom domain, mean (SD) at 12 weeks	50.3 (20.9)	62.0 (18.3)	−12.1 (−22.2 to −2.0); $p = 0.018$
	SGRQ impact domain, mean (SD) at baseline	48.1 (20.7)		
	SGRQ impact domain, mean (SD) at 12 weeks	34.3 (16.1)	49.0 (19.4)	−13.1 (−19.7 to −6.6); $p < 0.001$
	SGRQ activity domain, mean (SD) at baseline	64.3 (22.7)		
	SGRQ activity domain, mean (SD) at 12 weeks	60.9 (14.2)	65.8 (18.7)	−3.3 (−9.8 to 3.2); $p = 0.31$

TABLE 39 Adverse events: thalidomide

Adverse events, n (%)	Thalidomide (N = 22)	Placebo (N = 23)
Participants with ≥ 1 adverse event	17 (77)[a]	5 (22)[a]
Participants with a serious adverse event (influenza)	0	1 (4)
Adverse events requiring dose reduction		
Constipation	2 (9)	0
Bradycardia	1 (5)	0
Adverse events requiring drug discontinuation (progressive illness or inability to travel for visits)	2 (9)	1 (4)
Gastrointestinal adverse events		
Constipation	8 (36)	1 (4)
Change in taste	2 (9)	0
Dry mouth	2 (9)	0
Anorexia	1 (5)	NR
General adverse events		
Dizziness	6 (27)	0
Malaise	3 (14)	0
Oedema	2 (9)	0
Rash	2 (9)	0
Sleepiness	NR	1 (4)
Respiratory adverse event: worsening dyspnoea	2 (9)	2 (9)
Infectious adverse event: viral URTI	5 (23)	1 (4)
Cardiac adverse event: bradycardia	1 (5)	0

NR, not reported.
a $p = 0.001$.

Summary

One small randomised crossover trial with a low risk of bias provided data on the effectiveness of thalidomide for supressing cough in IPF. Cough, cough-related QoL and respiratory-related QoL were significantly improved with thalidomide compared with placebo. Adverse events were experienced with thalidomide. Caution is required in the interpretation of these data given the small sample size.

Assessment of effectiveness of pharmacological interventions for severe idiopathic pulmonary fibrosis

Sildenafil

Forced vital capacity

The results for per cent predicted FVC from the trial by the IPFCRN 2010 group (Zisman and colleagues)[72] can be seen in *Table 40*. After 12 weeks of treatment there was no statistically significant effect for sildenafil, compared with placebo (mean difference 0.32, 95% CI −1.12 to 1.76; $p = 0.66$). *Figure 2* presents the results of FVC graphically, where it can be seen that the CIs cross the line of no effect.

TABLE 40 Forced vital capacity: outcomes sildenafil

| Study | Outcome | Interventions | | Difference between groups |
		Sildenafil ($n = 89$)	Placebo ($n = 91$)	
Zisman *et al.* IPFCRN 2010[72]	FVC (% predicted), mean change (95% CI) at 12 weeks	–0.97 (–2.00 to 0.06)	–1.29 (–2.30 to –0.28)	0.32 (–1.12 to 1.76); $p = 0.66$

Diffusing capacity of the lung for carbon monoxide

Treatment with sildenafil led to a statistically significantly lower rate of decline of per cent predicted DL_{CO} from baseline than treatment with placebo: –0.33 versus –1.87%, respectively; $p = 0.04$ (*Table 41* and see *Figure 3*).

Six-minute walk test

The presence of an improvement of at least 20% in the 6MWT distance at 12 weeks was the primary outcome measure in the Zisman and colleagues IPFCRN 2010[72] trial of sildenafil. As seen in *Table 42*, the proportions with an improvement of at least 20% were not statistically significantly different between groups (10% vs. 7% in the sildenafil and placebo groups, respectively; $p = 0.39$).

Other secondary outcomes

Other outcomes reported in the Zisman and colleagues IPFCRN 2010 trial[72] include the partial pressure of oxygen, carbon dioxide and the oxygen saturation, the results of which can be found in *Appendix 5*.

Acute exacerbations

The incidence of acute exacerbation was not statistically significantly different between groups in the Zisman and colleagues IPFCRN 2010[72] trial, with 2% of the sildenafil-treated participants and 4% of the placebo-treated participants being recorded as having had an acute exacerbation (*Table 43*).

Dyspnoea

Two scales were used to measure dyspnoea in the Zisman and colleagues IPFCRN 2010[72] trial (*Table 44*): the UCSDSBQ and the Borg Dyspnoea Index. The UCSDSBQ was described above. The minimally important difference for this is reported by the study authors to be five points. The Borg Dyspnoea Index measures perceived breathlessness on a scale of 0 (none) to 10 (maximum) and the authors report that this has a

TABLE 41 Diffusing capacity of the lung for carbon monoxide outcomes: sildenafil

| Study | Outcome | Interventions | | Difference between groups |
		Sildenafil ($n = 89$)	Placebo ($n = 91$)	
Zisman *et al.* IPFCRN 2010[72]	DL_{CO} (% predicted), mean change	–0.33 (–1.36 to 0.71)	–1.87 (–2.91 to –0.83)	1.55 (0.08 to 3.01); $p = 0.04$

TABLE 42 Six-minute walk test outcomes: sildenafil

| Study | Outcome | Interventions | | Difference between groups |
		Sildenafil ($n = 89$)	Placebo ($n = 91$)	
Zisman *et al.* IPFCRN 2010[72]	6MWT at baseline	246.39 (103.40)	269.55 (129.83)	
	6MWT improvement in distance of ≥ 20%	9 (10%)	6 (7%)	3%;[a] $p = 0.39$

a Calculated by reviewer.

TABLE 43 Acute exacerbations: sildenafil

| | | Interventions | | Difference between groups |
Study	Outcome	Sildenafil (*n* = 89)	Placebo (*n* = 91)	
Zisman *et al.* IPFCRN 2010[72]	Acute exacerbation, *n* (%)	2 (2)	4 (4)	2%; *p* = 0.68

TABLE 44 Dyspnoea outcomes: sildenafil

| | | Interventions | | Difference between groups |
Study	Outcome	Sildenafil (*n* = 89)	Placebo (*n* = 91)	
Zisman *et al.* IPFCRN 2010[72]	UCSDSBQ (SD) at baseline[a]	50.71 (22.00)	43.28 (20.18)	
	UCSDSBQ mean change (95% CI)	0.22 (–3.10 to 3.54)	6.81 (3.53 to 10.08)	–6.58 (–11.25 to –1.92); *p* = 0.006
	Borg Dyspnoea Index after walk test (SD) at baseline[b]	3.82 (1.95)	3.33 (1.73)	
	Borg Dyspnoea Index after walk test, mean change (95% CI)	0.04 (–0.30 to 0.37)	0.37 (0.04 to 0.70)	–0.34 (–0.81 to 0.14); *p* = 0.16

a UCSDSBQ total score ranges from 0 to 120, with higher scores indicating greater shortness of breath.
b Measures perceived breathlessness on a scale of 0 (none) to 10 (maximum).

minimally important difference of 1. In this study the Borg Dyspnoea Index was applied after the participants had undertaken the 6MWT. The difference between groups was statistically significant when dyspnoea was measured using the UCSDSBQ (difference –6.58; *p* = 0.006) but was not statistically significantly different between groups on the Borg Dyspnoea Index after the walk test (difference –0.34; *p* = 0.16).

Quality of life

Quality of life of participants in the trial by Zisman and colleagues (IPFCRN)[72] was measured using the SGRQ, the SF-36 and the EQ-5D. The scoring for the SGRQ was previously described [see *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*, *BIBF 1120 (nintedanib)* and *Pirfenidone*] and fuller details can be found in *Chapter 4* (see *Systematic review of health-related quality of life studies*). The trial authors report that the MCID for the SGRQ is 5–8 points. The SF-36 measures aggregate scores on separate subscales for physical and mental health (see *Chapter 4*, *Systematic review of health-related quality of life studies* for more detail), with higher scores indicating better function. The study authors state that the SF-36 has a MCID of 2–4 points in IPF. The EQ-5D measures general QoL on a self-report questionnaire, with higher scores indicating better QoL (see *Chapter 4*, *Systematic review of health-related quality of life studies*) and on a VAS with a range of 0 to 100 (higher indicates better QoL). The study reports that the MCID is approximately 0.08 for the self-report questionnaire and 7 points for the VAS.

The changes from baseline scores of the SGRQ total score, symptoms score, activity score and impacts score were statistically significantly improved in those treated with sildenafil, compared with those treated with placebo (*Table 45*). On the SF-36 physical and mental health scores, and the EQ-5D self-report index and the VAS, there were no statistically significant differences between the two groups after 12 weeks of therapy. The study reported the baseline and end-point values on each of the SF-36 components, and full details of these can be seen in *Appendix 5*. Only the change from baseline score on the general health component was statistically significant between groups [sildenafil –1.04 (95% CI –2.52 to 0.44), placebo –3.89 (95% CI –5.37 to –2.42), difference 2.86 (95% CI 0.76 to 4.95); *p* = 0.008].

TABLE 45 Quality of life outcomes: sildenafil

Study	Outcome	Interventions		Difference between groups
		Sildenafil (*n* = 89)	Placebo (*n* = 91)	
Zisman *et al.* IPFCRN 2010[72]	SGRQ[a] mean total score (range) at baseline	54.55 (16.46)	51.72 (15.86)	
	SGRQ mean (95% CI) change in total score at 12 weeks	–1.64 (–3.91 to 0.64)	2.45 (0.17 to 4.72)	–4.08 (–7.30 to –0.86); *p* = 0.01
	SGRQ mean (SD) symptoms score at baseline	58.20 (17.75)	53.99 (18.90)	
	SGRQ mean (95% CI) change in symptoms score at 12 weeks	–3.58 (–7.02 to –0.13)	2.15 (–1.30 to 5.61)	–5.73 (–10.61 to –0.85); *p* = 0.02
	SGRQ mean (SD) activity score at baseline	71.20 (17.50)	68.02 (17.63)	
	SGRQ mean (95% CI) change in activity score at 12 weeks	–1.15 (–3.68 to 1.38)	2.49 (0.00 to 4.99)	–3.64 (–7.20 to –0.09); *p* = 0.04
	SGRQ mean (SD) impacts score at baseline	43.20 (19.26)	39.77 (18.81)	
	SGRQ mean (95% CI) change in impacts score at 12 weeks	–0.88 (–3.72 to 2.02)	2.82 (–0.03 to 5.67)	–3.70 (–7.76 to 0.37); *p* = 0.07
	SF-36[b] aggregate physical score[c] (SD) at baseline	33.17 (9.19)	34.84 (8.69)	
	SF-36 mean change (95% CI) aggregate physical score at 12 weeks	–0.51 (–1.86 to 0.83)	–0.35 (–1.68 to 0.99)	–0.17 (–2.06 to 1.73); *p* = 0.86
	SF-36 aggregate mental score (SD) at baseline	49.53 (9.76)	50.58 (9.52)	
	SF-36 mean change (95% CI) aggregate mental score at 12 weeks	1.30 (–0.59 to 3.18)	3.02 (1.15 to 4.89)	–1.72 (–4.38 to 0.93); *p* = 0.20
	EQ-5D[d] self-report score (SD) at baseline	0.71 (0.24)	0.74 (0.19)	
	EQ-5D mean change (95% CI) self-report questionnaire at 12 weeks	–0.01 (–0.06 to 0.03)	–0.03 (–0.08 to 0.01)	0.02 (–0.04 to 0.08); *p* = 0.54
	EQ-5D VAS (SD) at baseline	66.49 (17.45)	67.66 (16.98)	
	EQ-5D mean change (95% CI) VAS at 12 weeks	0.48 (–3.10 to 4.06)	–1.81 (–5.34 to 1.73)	2.28 (–2.75 to 7.32); *p* = 0.37

a   Scores range from 0 to 100, with lower scores indicating better QoL.
b   Score ranges from 0 to 100, with a higher score indicating better function.
c   SF-36 individual components can be seen in *Appendix 5*.
d   Scale of –0.59 to 1.00 (higher score indicates better QoL).

Adverse events

Serious adverse events were reported where at least one event was recorded during follow-up in the Zisman and colleagues IPFCRN 2010[72] trial. These are shown in *Table 46* within each of the major organ classes presented in the publication. *Appendix 5* has the complete details of each condition within each major class. The number of participants with at least one serious adverse event was similar between groups, with 15% and 16% for the sildenafil and placebo groups, respectively (see *Table 46*). In terms of number of events, the study publication reports that among all participants, 14 serious adverse events occurred in the sildenafil group and 23 in the placebo group. The rates of serious adverse events by each major organ class were also similar between the two groups.

A range of adverse events were reported for those randomised to treatment with sildenafil and those randomised to treatment with placebo and can be seen in *Table 47*, which reports these for each of the major organ classes presented in the report. As can be seen, there are no statistically significant differences in the proportion of adverse events recorded between the two groups. The rate of occurrence of any adverse events was 90% in the sildenafil-treated participants, compared with 87% in the placebo-treated participants. The majority of adverse events in both groups were either respiratory or infections.

Deaths from any cause were reported in the Zisman and colleagues IPFCRN 2010 trial,[72] with two (4%) participant deaths in the sildenafil-treated arm and four in the placebo-treated arm. There was no statistically significant difference between the proportions of deaths ($p = 0.43$).

Summary

The primary outcome of a 20% improvement on the 6MWT was not significant in the study comparing sildenafil with placebo. Lung function as measured by the FVC was also not statistically significant, but the DL_{CO} results favoured treatment with sildenafil and this may, therefore, represent a pulmonary vascular effect. Dyspnoea may be improved (depending on the measure used and test conditions) and acute exacerbations showed similar rates between groups. QoL was better in those treated with sildenafil when measured using the SGRQ, but not when using the SF-36 or the EQ-5D. Serious adverse events and adverse events showed similar rates between groups in this 12-week study. This study had an unclear risk of bias, which should be considered when interpreting the results.

TABLE 46 Serious adverse events for sildenafil and placebo, by major organ system

SAEs	Number (%) with at least one SAE; number observed (assumed by reviewer)		Difference between groups
	Sildenafil ($n = 89$)	Placebo ($n = 91$)	
Any	13 (15); 14	15 (16); 23	$p = 0.73$
Respiratory, thoracic or mediastinal disorder	7 (8); 7	9 (10); 11	$p = 0.63$
Infection or infestation	3 (3); 4	2 (2); 2	$p = 0.68$
Cardiac disorder	1 (1); 1	3 (3); 3	$p = 0.62$
Gastrointestinal disorder	2 (2); 2	1 (1); 1	$p = 0.62$
Hepatobiliary disorders	0	1 (1); 1	$p = 0.99$
Injury, poisoning and procedural complications	0	1 (1); 3	$p = 0.99$
Neoplasms benign, malignant and unspecified	0	1 (1); 1	$p = 0.99$
Nervous system disorders	0	1 (1); 1	$p = 0.99$
Death from any cause, *n* (%)	2 (2)	4 (4)	$p = 0.43$

SAE, serious adverse event.

TABLE 47 Adverse events for sildenafil and placebo, by major organ system

AEs	Number (%) with at least one AE; number of adverse events observed		Difference between groups
	Sildenafil ($n = 89$)	Placebo ($n = 91$)	
Any body system and event	80 (89.9); 442	79 (86.8); 453	$p = 0.52$
Respiratory, thoracic and mediastinal disorders	46 (51.7); 85	52 (57.1); 86	$p = 0.46$
Infections or infestations	42 (47.2); 58	39 (42.9); 52	$p = 0.56$
Nervous system disorders	35 (39.3); 56	35 (38.5); 58	$p = 0.91$
Gastrointestinal disorders	32 (36.0); 45	27 (29.7); 43	$p = 0.37$
General disorders and administration site conditions	27 (30.3); 45	26 (28.6); 44	$p = 0.80$
Musculoskeletal and connective tissue disorders	25 (28.1); 34	20 (22.0); 33	$p = 0.34$
Vascular disorders	15 (16.9); 19	13 (14.3); 16	$p = 0.64$
Skin and subcutaneous tissue disorders	12 (13.5); 21	14 (15.4); 15	$p = 0.72$
Cardiac disorders	11 (12.4); 15	13 (14.3); 15	$p = 0.70$
Investigations	10 (11.2); 11	13 (14.3); 18	$p = 0.54$
Psychiatric disorders	6 (6.7); 7	16 (17.6); 21	$p = 0.03$
Eye disorders	12 (13.5); 17	9 (9.9); 10	$p = 0.45$
Metabolism and nutritional disorders	5 (5.6); 5	11 (12.1); 16	$p = 0.13$
Blood and lymphatic system disorders	7 (7.9); 7	2 (2.2); 2	$p = 0.10$
Ear and labyrinth disorders	5 (5.6); 5	3 (3.3); 5	$p = 0.49$
Injury, poisoning and procedural complications	3 (3.4); 5	4 (4.4); 6	$p = 0.99$
Renal and urinary disorders	3 (3.4); 3	4 (4.4); 5	$p = 0.99$
Neoplasms benign, malignant and unspecified	2 (2.2); 2	3 (3.3); 3	$p = 0.99$
Hepatobiliary disorders	0	2 (2.2); 3	$p = 0.50$
Reproductive system and breast disorders	1 (1.1); 1	1 (1.1); 1	$p = 0.99$

AE, adverse event.

Assessment of effectiveness of non-pharmacological interventions for idiopathic pulmonary fibrosis

Disease management programmes

The trial by Lindell and colleagues[74] did not report lung function tests as outcomes. Two outcomes of relevance to the present systematic review were measures of dyspnoea and QoL. Other outcomes that were reported can be found in *Appendix 5* [these were anxiety as measured by the Beck Anxiety Inventory (BAI), depression as measured by the Beck Depression Inventory-II and stress as measured by the Perceived Stress Scale (PSS)]. This study had an uncertain risk of bias.

Dyspnoea

Participants' dyspnoea was measured in the Lindell and colleagues[74] trial using the UCSDSBQ. This measures shortness of breath on a scale of 0–120, with higher scores indicating more breathlessness. After the 6-week disease management programme or the usual care control, the scores on the UCSDSBQ were not statistically significantly different between the two groups ($p = 0.972$) (*Table 48*). No change scores of the UCSDSBQ were provided.

TABLE 48 Dyspnoea outcomes: disease management programme

| Study | Outcome | Interventions | | Difference between groups |
		Disease management ($n = 10$)	Usual care control ($n = 9$)	
Lindell et al. 2010[74]	UCSDSBQ adjusted mean (SD) score[a] at 6 weeks	49.51 (22.64)	49.88 (22.64)	$p = 0.972$

a UCSDSBQ total score ranges from 0 to 120, with higher scores indicating greater shortness of breath.

Quality of life

Quality of life of participants in the Lindell and colleagues[74] trial was measured using the SF-36 and results can be seen in *Table 49*. As discussed previously, the SF-36 measures QoL within two domains: physical and mental health. Each scale is scored from 0 to 100, with a higher score indicating better function (see also *Chapter 4*, *Systematic review of health-related quality of life studies*). After 6 weeks the SF-36 physical health score was lower (worse) in the participants assigned to the disease management programme than in those assigned to the usual care control. The difference between the groups was statistically significant ($p = 0.038$). No change scores on the SF-36 were provided. The scores on the SF-36 mental health scale were not statistically significantly different between the two interventions ($p = 0.772$).

Summary

One small pilot study with an unclear risk of bias reported limited evidence on the effects of a disease management programme in IPF. The disease management programme did not lead to any significant differences in dyspnoea. QoL appears to be adversely affected on measures of physical health but not on measures of mental health. The study was designed to obtain pilot data for a future study and is unlikely to be sufficiently powered.

Pulmonary rehabilitation

Forced vital capacity

After the 10-week pulmonary rehabilitation programme or usual care, the FVC was very similar between the two groups in the Nishiyama and colleagues[73] trial (*p*-value not stated; *Table 50*). The FVC, litres, at end point was not changed from baseline measurements in either group. This study had an uncertain risk of bias.

In the Jastrzebski and colleagues CCT[61] it is reported that there were no statistically significant changes in the results of lung function tests (assumed to include VC as this was an outcome) after 12 weeks, but no data were reported.

TABLE 49 Quality of life outcomes: disease management programme

| Study | Outcome | Interventions | | Difference between groups |
		Disease management ($n = 10$)	Usual care control ($n = 9$)	
Lindell et al. 2010[74]	SF-36 physical adjusted mean (SD) at 6 weeks[a]	31.06 (4.61)	36.04 (4.63)	$p = 0.038$
	SF-36 mental adjusted mean (SD) at 6 weeks	55.98 (2.71)	55.61 (2.71)	$p = 0.772$

a Score ranges from 0 to 100, with a higher score indicating better function.

TABLE 50 Forced vital capacity outcomes: pulmonary rehabilitation

| Study | Outcome | Interventions | | Difference between groups |
		Pulmonary rehabilitation (*n* = 13)	Control (*n* = 15)	
Nishiyama *et al.* 2008[73]	FVC, l, at 10 weeks	2.1 (0.4)	2.0 (0.8)	0.03 (–0.13 to 0.19); NS

NS, not significant.

Six-minute walk test

The distance walked on the 6MWT was statistically significantly lower in the pulmonary rehabilitation group than in the usual care group at 10 weeks ($p < 0.01$) in the Nishiyama and colleagues[73] trial (*Table 51*). This does not take into account the differences at baseline between the groups on this measure where the distance was also lower in the pulmonary rehabilitation group. When the mean change is estimated (by reviewers) it can be observed that there is a positive change in the distance walked in the pulmonary rehabilitation group and a negative change in the distance walked in the control group, but these data cannot be compared statistically to establish if there are any differences between the two groups.

The distance walked on the 6MWT at 12 weeks was 600.8 m in the intervention group and 544.5 m in the control group of the Jastrzebski and colleagues study[61] (see *Table 51*). The study reports only within-group comparisons, however, so it is unclear whether or not this is a statistically significant difference. The 6MWT distance at 6 weeks is also reported in this study and data can be found in *Appendix 5*.

Dyspnoea

Dyspnoea was measured using the BDI in the Nishiyama and colleagues[73] RCT of pulmonary rehabilitation versus control. The BDI comprises three categories: functional impairment, magnitude of task and magnitude of effort. The total BDI score ranges from 0 to 12, with higher scores denoting milder dyspnoea. Results can be found in *Table 52*. The difference between groups after the 10-week programme was small and not statistically significant (*p*-value not reported). The change in BDI was not presented but has been calculated here, where it can be seen that the BDI did not change in the pulmonary rehabilitation group and improved only slightly in the control group.

TABLE 51 Six-minute walk test outcomes: pulmonary rehabilitation

| Study | Outcome | Interventions | | Difference between groups |
		Inspiratory muscle training (*n* = 16)	Control (*n* = 14)	
Jastrzebski *et al.* 2008[61]	6MWT at baseline	487.4 (100.2)	485.6 (111.9)	
	6MWT distance (m) at 12 weeks	600.8 (93.7)	544.5 (121.5)	Not reported
		Pulmonary rehabilitation (*n* = 13)	Control (*n* = 15)	
Nishiyama *et al.* 2008[73]	6MWT at baseline	385 (116)	476 (128)	
	6MWT distance (m)	427 (84)	472 (130)	46.3 (8.3 to 84.4); $p < 0.01$

TABLE 52 Dyspnoea outcomes: pulmonary rehabilitation

| Study | Outcome | Interventions | | Difference between groups in change from baseline |
		Inspiratory muscle training (n = 16)	Control (n = 14)	
Jastrzebski et al. 2008[61]	Borg Dyspnoea Index[a] before 6MWT at baseline	1.7 (1.1)	1.9 (1.0)	
	Borg Dyspnoea Index before 6MWT at 12 weeks	1.2 (0.5)	1.6 (1.6)	Not reported
	Borg Dyspnoea Index after 6MWT at baseline	5.3 (2.2)	5.2 (2.3)	
	Borg Dyspnoea Index after 6MWT at 12 weeks	3.8 (2.3)	4.2 (2.1)	Not reported
	BDI[b] summary score at baseline	6.1 (2.3)	6.07 (2.3)	
	BDI summary score at 12 weeks	7.9 (2.7)	6.4 (2.5)	Not reported
		Pulmonary rehabilitation (n = 13)	Control (n = 15)	
Nishiyama et al. 2008[73]	BDI[b] at baseline	6.7 (1.4)	8.4 (1.5)	
	BDI score	6.7 (1.3)	8.0 (2.2)	0.4 (–0.6 to 1.4); NS

NS, not significant.
a   Borg dyspnoea ranges from a scale of 0 (none) to 10 (maximum) breathlessness.
b   BDI score ranges from 0 to 12, with higher score denoting milder dyspnoea in daily living.

In the Jastrzebski and colleagues study,[61] outcomes using both the Borg Dyspnoea Index and the BDI were reported. These included the Borg Dyspnoea Index before and after the 6MWT, the BDI functional impairment, magnitude of task, magnitude of effort, and the summary score. Results for the Borg Dyspnoea Index and summary score of the BDI at 12 weeks can be seen in *Table 52*, where it appears that these did not differ significantly between groups; however, no between-group analysis was presented in the study to confirm this. Results of the BDI functional impairment, magnitude of task and magnitude of effort and dyspnoea outcomes at 6 weeks can be found in *Appendix 5*.

Quality of life

Quality of life of participants in the RCT of pulmonary rehabilitation was assessed using the SGRQ.[73] As discussed previously, the SGRQ has a total score and three subscales, and each scale ranges from a score of 0 to 100 (lower scores indicating better QoL). *Table 53* summarises the results from the Nishiyama and colleagues[73] trial where the SGRQ outcomes were statistically compared using the end-point scores. This showed that the difference between the groups was statistically significant only on the total score ($p < 0.05$), suggesting that those in the control group had better QoL than those participating in the pulmonary rehabilitation programme. This does not take into account differences on these scales that were observed between the two groups at baseline where the pulmonary rehabilitation group had worse QoL. When the change from baseline scores are estimated, it would appear that QoL improved in the pulmonary rehabilitation group whereas it declined in the control group, although this is based on observation of the data only.

In the Jastrzebski and colleagues study,[61] QoL was assessed using the SF-36. This measure is described in more detail in *Chapter 4* (see *Systematic review of health-related quality of life studies*). Data from the 12-week assessment of the individual SF-36 components and the two summary scores can be seen in *Table 53*. No statistical analysis was presented of the comparison between treatment groups; however, it can be observed that scores were similar between groups.

TABLE 53 Quality of life outcomes: pulmonary rehabilitation

| Study | Outcome | Interventions | | Difference between groups |
		Inspiratory muscle training ($n = 16$)	Control ($n = 14$)	
Jastrzebski et al. 2008[61]	SF-36 PF at baseline	54.4 (23.6)	54.3 (17.4)	
	SF-36 PF at 12 weeks	68.1 (22.3)	62.5 (14.5)	Not reported
	SF-36 RP at baseline	42.8 (32.0)	44.6 (24.4)	
	SF-36 RP at 12 weeks	65.6 (31.5)	64.3 (30.6)	Not reported
	SF-36 BP at baseline	68.9 (27.2)	66.8 (22.2)	
	SF-36 BP at 12 weeks	75.7 (20.7)	69.6 (17.8)	Not reported
	SF-36 GH at baseline	37.8 (17.7)	37.4 (11.1)	
	SF-36 GH at 12 weeks	44.2 (22.4)	42.4 (13.6)	Not reported
	SF-36 VT at baseline	54.4 (18.2)	52.5 (13.3)	
	SF-36 VT at 12 weeks	60.0 (16.4)	57.1 (16.9)	Not reported
	SF-36 SF at baseline	58.9 (23.5)	58.0 (14.4)	
	SF-36 SF at 12 weeks	75.8 (30.8)	68.7 (18.2)	Not reported
	SF-36 RE at baseline	68.8 (39.4)	69.0 (44.3)	
	SF-36 RE at 12 weeks	70.8 (29.5)	78.6 (28.1)	Not reported
	SF-36 MH at baseline	64.2 (17.7)	65.1 (17.9)	
	SF-36 MH at 12 weeks	68.7 (16.8)	66.3 (20.2)	Not reported
	SF-36 PCS at baseline	38.4 (8.2)	36.1 (9.1)	
	SF-36 PCS at 12 weeks	44.8 (6.0)	42.3 (5.8)	Not reported
	SF-36 MCS at baseline	46.6 (9.9)	46.5 (10.9)	
	SF-36 MCS at 12 weeks	47.8 (11.22)	47.8 (11.6)	Not reported
		Pulmonary rehabilitation ($n = 13$)	Control ($n = 15$)	
Nishiyama et al. 2008[73]	SGRQ symptoms score (SD) at baseline	56.4 (22.3)	38.0 (25.8)	
	SGRQ symptoms score (SD) at 10 weeks	53.4 (25.8)	40.6 (21.2)	–5.7 (–18.7 to 7.2); NS
	SGRQ symptoms score change	3	–2.6	
	SGRQ activity score (SD) at baseline	64.7 (17.1)	50.4 (26.2)	
	SGRQ activity score (SD) at 10 weeks	62.5 (16.9)	54.0 (22.6)	–5.8 (–14.7 to 3.1); NS
	SGRQ activity score change	2.2	–3.6	
	SGRQ impacts score (SD) at baseline	39.7 (17.6)	29.9 (23.7)	
	SGRQ impacts score (SD) at 10 weeks	36.5 (17.5)	32.9 (23.5)	–6.2 (–12.8 to 0.3); NS
	SGRQ impacts score change	3.2	–3	
	SGRQ total score (SD) at baseline	50.2 (16.3)	37.8 (22.7)	
	SGRQ total score (SD) at 10 weeks	47.3 (17.4)	40.9 (20.7)	–6.1 (–11.7 to 0.5); $p < 0.05$
	SGRQ total score change	2.9	–3.1	

BP, bodily pain; GH, general health; MCS, mental cumulative score; MH, mental health; NS, not significant; PCS, physical cumulative score; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; VT, vitality.

Other secondary outcomes

Other outcomes reported in the Nishiyama and colleagues[73] trial include FEV_1, TLC, and the partial pressure of oxygen and carbon dioxide, the results of which can be found in *Appendix 5*. Other outcomes reported in the Jastrzebski and colleagues study[61] include inspiratory capacity and oxygen cost diagram, the results of which can be found in *Appendix 5*.

Summary

One study with an unclear risk of bias and one with a high risk of bias provided limited data on the effectiveness of a pulmonary rehabilitation programme or the addition of inspiratory muscle training, on IPF. Results are uncertain as to the effects of these types of intervention, and there were baseline differences between groups on many key outcomes (QoL, 6MWT), which should be considered when interpreting the results of one of the studies. On measures of lung function (FVC) and dyspnoea, there does not appear to be a beneficial effect of the programme.

Network meta-analysis

Network meta-analysis is an extension of conventional, pairwise, meta-analysis combing direct and indirect RCT evidence. This approach allows simultaneous comparison of multiple treatments from trials comparing different sets of treatments (provided there is a connected network). It is based on the assumption that, on a suitable scale, we can add and subtract the within-trial estimates of relative treatment effects. In this case, the difference in effect between treatments A and B is equal to the difference in effects between treatments A and C, and B and C: a difference in differences model $A - B = (A - C) - (B - C)$.[76,77] When direct randomised evidence is limited or unavailable, NMA is increasingly being used to inform clinical decision-making and its use has been embraced by reimbursement agencies including NICE in the UK and endorsed by health technology assessment bodies such as the International Society for Pharmacoeconomics and Outcomes Research.[78,79] NMA is typically performed in a Bayesian framework using vague priors to ensure posterior estimates are based on the observed data.

The IPF NMA focused on the FVC end point as clinical evidence has established a correlation with disease progression[80] and, thus, FVC drives the economic model (discussed in *Chapter 4*, *Independent economic evaluation*). Ten studies met the inclusion criteria and were included in the analysis (*Table 54*). The two non-pharmacological interventions were excluded from the NMA on clinical advice as they are not interchangeable treatments with the pharmacological interventions, but are rather given alongside them. In addition, thalidomide was excluded from the RCT as the focus of treatment is not to improve lung function. Inhaled NAC was considered separately from triple therapy as it has a different method of administration. Between-study heterogeneity has been examined in the systematic review section of this report.

The FVC end point was measured on two continuous scales: FVC per cent predicted; and absolute change from baseline (litres). Where both scales were presented in the same trial (Richeldi and colleagues[71]), per cent predicted was preferred as this measure was adjusted for age, height and sex. Trials were relatively homogeneous apart from Zisman and colleagues IPFCRN 2010,[72] which studied a more severe IPF population. This trial has nevertheless been included in the analysis as it only compares with placebo and, thus, does not impact on the other relative treatment effects; it may impact on the between-study SD which will be considered in a sensitivity analysis. It is also worth observing that the 1991 study by Raghu and colleagues[70] is in a very small population ($n = 19$). Furthermore, clinical opinion has identified that studies published prior to 2000 may have included NSIP patients along with IPF, with NSIP patients performing better.[1,8] This might have biased outcomes for Raghu and colleagues[70] if there were more NSIP subjects in the treatment arm; however, this is not reported. In any case, this study's effect on the analysis is dominated by the larger study that contributes azathioprine data.

The resulting evidence network is illustrated in *Figure 9*.

TABLE 54 Forced vital capacity data used in the analysis

Study	Treatment	FVC measure	Mean (SD)	n
Taniguchi et al. 2010[65]	Pirfenidone	Litres	–0.09 (0.20)	104
	Placebo	Litres	–0.16 (0.20)	103
Richeldi et al. 2011[71]	BIBF 1120	Per cent predicted	–1.04 (9.1)	85
	Placebo	Per cent predicted	–6.0 (9.4)	85
Raghu et al. 1991[70]	Azathioprine	Per cent predicted	6.5 (16.8)	10
	Placebo	Per cent predicted	1.7 (22.2)	9
Raghu et al. 2012[68]	NAC triple	Litres	–0.24 (0.40)	77
	Placebo	Litres	–0.23 (0.40)	78
Zisman et al. 2010[72]	Sildenafil	Per cent predicted	–0.97 (4.9)	89
	Placebo	Per cent predicted	–1.29 (4.8)	91
Homma et al. 2012[69]	Inhaled NAC	Litres	–0.09 (0.3)	38
	Placebo	Litres	–0.15 (0.2)	38
Demedts et al. 2005[67]	NAC triple	Litres	–0.06 (0.34)	71
	Azathioprine	Litres	–0.19 (0.41)	68
Capacity 006[62]	Pirfenidone	Per cent predicted	–9 (19.6)	171
	Placebo	Per cent predicted	–9.6 (19.1)	173
Capacity 004[62]	Pirfenidone	Per cent predicted	–8 (16.5)	174
	Placebo	Per cent predicted	–12.4 (18.5)	174
Azuma et al. 2005[66]	Pirfenidone	Litres	–0.03 (0.22)	72
	Placebo	Litres	–0.13 (0.19)	35
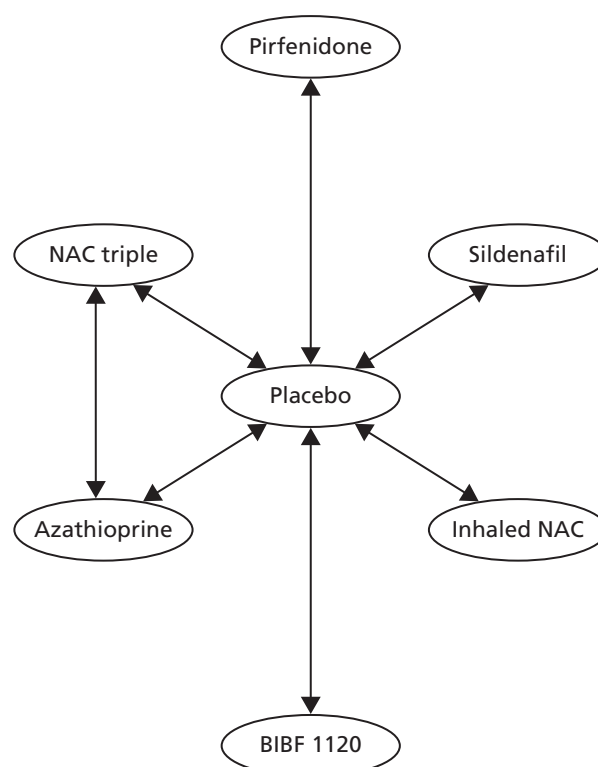
FIGURE 9 Evidence network for the FVC end point.

Were the network star-shaped,[81] with each treatment compared only with placebo, the NMA results would approximate the traditional pairwise meta-analysis in *Figures 2–8*. However, in this case there is one closed loop in the evidence network (placebo ↔ NAC triple therapy ↔ azathioprine) (see *Figure 9*). The presence of both direct and indirect evidence allows us to compare azathioprine with placebo directly and indirectly via NAC triple therapy and vice versa. Because of this closed loop, there is also potential for inconsistency between the direct and indirect evidence within the network.

The standardised mean difference approach was developed to combine treatment effects reported in distinct continuous measures which essentially measure the same thing and convert them to a common scale.[82] The assumption underlying this analysis is that mean change in FVC per cent predicted is measuring the same thing as absolute change in FVC, albeit we know that FVC per cent predicted is adjusted for age, height, sex and race. The mean difference between treatment arms is divided by the SD; thus, the effect measures are adjusted to be defined in terms of units of SD. A disadvantage is that the standardised mean difference measure is difficult to interpret and thereby populate a cost-effectiveness model.

Treatment effects are converted to standardised mean difference using Hedges' adjusted *g* method (*Table 55*). This was preferred over Cohen's method to overcome potential small sample bias;[82] one study consists of only 19 subjects split between two arms (Raghu and colleagues[70]). A sensitivity analysis was conducted using Cohen's method. The standardised mean difference calculation is performed with the treatment effect being subtracted from placebo; this is in order to produce output suitable for use in an economic model (where the treatment benefit is expressed as reduction in the rate of decline compared with placebo). The negative values in the table below, therefore, represent a treatment benefit.

The standardised mean difference scale measures were then meta-analysed as contrast data (mean differences) using the Decision Support Unit NMA code.[83] Fixed- and random-effects models were conducted in OpenBUGS (OpenBUGS Foundation, www.openbugs.net/w/Overview, with best model fit determined by deviance information criterion (DIC).[84] Models were run for a burn in of 20 k simulations, which were discarded, and a further 100 k simulations for estimation. Two chains were run, each with different starting values. Trace, Brooks Gelman Rubin and density plots were examined to ensure that models had converged.[85] Results [means, standard errors and 95% credible intervals (CrIs)] are reported in *Table 56*. The OpenBUGS model code is provided in *Appendix 6*. A validation of the code was conducted in SAS (SAS Institute Inc., Cary, NC, USA) and is reported in *Appendix 7*.

TABLE 55 Forced vital capacity outcome data converted to standardised mean difference scale (Hedges' *g* method)

Study	Treatment	Control	SMD	SE (SMD)
Taniguchi *et al.* 2010[65]	Pirfenidone	Placebo	−0.35	0.14
Richeldi *et al.* 2011[71]	BIBF 1120[a]	Placebo	−0.53	0.16
Raghu *et al.* 1991[70]	Azathioprine	Placebo	−0.23	0.46
Raghu *et al.* 2012[68]	NAC triple	Placebo	0.02	0.16
Zisman *et al.* 2010[72]	Sildenafil	Placebo	−0.07	0.15
Homma *et al.* 2012[69]	Inhaled NAC	Placebo	−0.23	0.23
Demedts *et al.* 2005[67]	NAC triple	Azathioprine	−0.34	0.17
Capacity 006[62]	Pirfenidone	Placebo	−0.03	0.11
Capacity 004[62]	Pirfenidone	Placebo	−0.25	0.11
Azuma *et al.* 2005[66]	Pirfenidone	Placebo	−0.47	0.21

SE, standard error; SMD, standardised mean difference.
a　BIBF 1120 compares the 300mg/day dose with placebo.

TABLE 56 Forced vital capacity NMA results (vs. placebo), standardised mean differences

Comparator	Fixed effects			Random effects		
	Mean	SE	95% CrI	Mean	SE	95% CrI
Azathioprine	0.25	0.21	–0.16 to 0.66	0.15	0.38	–0.63 to 0.81
BIBF 1120	–0.53	0.16	–0.84 to –0.23	–0.53	0.37	–1.24 to 0.17
NAC triple therapy	–0.03	0.15	–0.33 to 0.27	–0.08	0.32	–0.73 to 0.49
Inhaled NAC	–0.23	0.23	–0.68 to 0.22	–0.23	0.41	–1.00 to 0.53
Pirfenidone	–0.21	0.06	–0.34 to –0.09	–0.25	0.18	–0.62 to 0.08
Sildenafil	–0.07	0.15	–0.36 to 0.23	–0.07	0.37	–0.76 to 0.63
reSD	NA	NA		0.24	0.24	
DIC	2.02			1.99		

NA, not applicable; reSD, random-effects standard deviation; SE, standard error.

Parameters stabilised between 5000 and 10,000 simulations. There was no evidence to suggest that the random-effect model was a better fit than the fixed effect; a difference in DIC of 2–3 is needed to be indicative of improved fit.[84] There was also considerable uncertainty around the random-effects between-study SD. The random-effects model incorporates a measure of between-study variation (heterogeneity) into the calculations, thereby increasing the standard error, and can therefore be viewed as a more conservative estimate.[86] There were insufficient data to allow random-effects variances to vary between treatments. Neither BIBF 1120 nor pirfenidone reached statistical significance using the random-effects model; therefore, a scenario using the random-effects estimates is included as a sensitivity analysis.

Only the fixed-effect results for BIBF 1120 and pirfenidone were statistically significant, reducing the rate of decline in FVC compared with placebo. The results for BIBF 1120, inhaled NAC and sildenafil were very similar to the trial results. The pirfenidone estimate was a meta-analysis of the four individual trials, with more weight given to the larger Capacity trials.[62] The azathioprine estimate was driven by the indirect estimated via NAC triple therapy over the small placebo study. Similarly, the NAC triple therapy estimate assumed more weight via the direct route to placebo rather than the indirect evidence via the small azathioprine study.

However, we still have interpretation difficulties with the summary standardised mean difference scores. To make these interpretable, and thereby usable in our cost-effectiveness model, we converted the standardised mean difference scores to log-odds ratios using the following formulae:[87]

$$\log OR = \frac{\pi}{\sqrt{3}} SMD \tag{1}$$

and

$$se(\log OR) = \frac{\pi}{\sqrt{3}} se(SMD) \tag{2}$$

This transformation assumes that the data have a logistic distribution and that within-study arms have equal variances. The within-study variances appear equal in *Table 54*, apart from the small azathioprine study (Raghu and colleagues[70]) and the Demedts and colleagues study.[67] Nevertheless, F-tests ($F = 0.57$, $p = 0.42$; $F = 0.69$, $p = 0.12$, respectively) show it is likely that the two groups come from populations with equal variances. Without access to individual patient-level data, the distributional assumption is hard to

verify, hence a sensitivity analysis assumed a normal distribution. *Table 57* reports the standardised mean differences converted to log-odds ratios and odds ratios for use in the economic model. Odds ratios for slowing the decline in FVC are plotted in *Figure 10*.

Odds ratios of < 1 are indicative of reducing FVC decline relative to placebo. Only BIBF 1120 and pirfenidone produced a statistically significant decline in rate of FVC decline relative to placebo (odds ratios and 95% CrI < 1). A head-to-head comparison of BIBF 1120 versus pirfenidone showed a trend favouring BIBF, although this was not statistically significant (see *Table 57*).

TABLE 57 Network meta-analysis fixed-effects results, converted to log-odds ratios for slowing the decline in FVC

Comparator (vs. placebo)	Log-odds ratios		Odds ratios	
	Mean	95% CrI	Mean	95% CrI
Azathioprine	0.44	−0.30 to 1.19	1.56	0.74 to 3.29
BIBF 1120	−0.97	−1.52 to −0.41	0.38	0.22 to 0.66
NAC triple therapy	−0.06	−0.60 to 0.48	0.94	0.55 to 1.62
Inhaled NAC	−0.42	−1.24 to 0.40	0.66	0.29 to 1.49
Pirfenidone	−0.39	−0.62 to −0.16	0.68	0.54 to 0.85
Sildenafil	−0.12	−0.65 to 0.41	0.89	0.52 to 1.51
Head-to-head comparison				
BIBF 1120 vs. pirfenidone	−0.58	−1.18 to 0.03	0.56	0.31 to 1.03
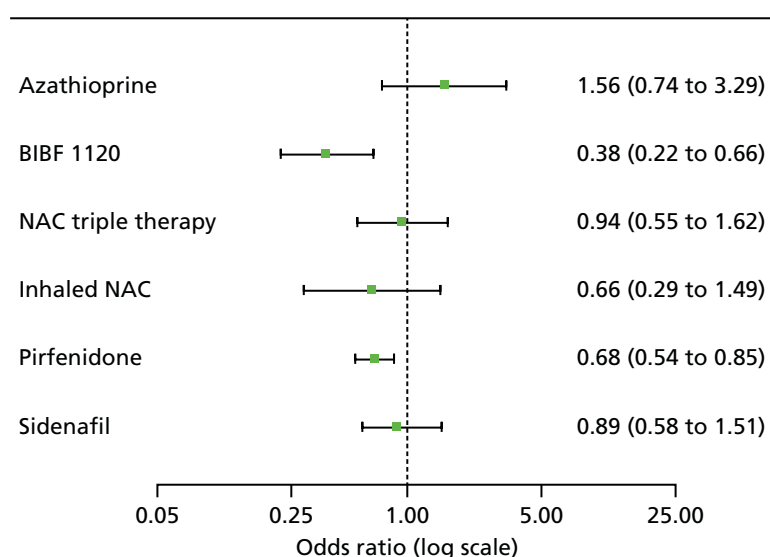
FIGURE 10 Forest plot of odds ratios slowing the decline in FVC (treatments vs. placebo).

Following best practice, we examined the level of agreement or consistency between the direct and indirect evidence.[88] We used the Bucher and colleagues[89] method to calculate the indirect estimates using the consistency equations and compared with the direct estimate, i.e.

$$\hat{\omega}_{BC} = d^D_{BC} - d^I_{BC} = d^D_{BC} - (d^D_{AC} - d^D_{AB}),$$ (3)

where $\hat{\omega}$ is the inconsistency parameter, and d^D and d^I are the direct and indirect evidence, respectively, comparing, for example, treatments B and C. The variance is then given by:

$$\text{var}\,\hat{\omega}_{BC} = \text{var}\,d^D_{BC} + \text{var}\,d^I_{BC} = \text{var}\,d^D_{BC} + \text{var}\,d^D_{AB} + \text{var}\,d^D_{AC}$$ (4)

These are then used to calculate a *z*-score which is then compared with the standard normal distribution:

$$z_{BC} = \frac{\hat{\omega}_{BC}}{\sqrt{\text{var}\,\hat{\omega}_{BC}}}$$ (5)

While more elaborate methods, such as Dias and colleagues'[90] node-splitting, plots of posterior mean deviance[91] or net heat plots[92] have been proposed for more complex networks with many loops, the Bucher and colleagues[89] approach is more efficient here as we have a simple network with only one loop and no multiarm trials. It is only necessary to calculate one measure of inconsistency within each loop; the others are intuitively the same.[88] The results of the calculations are presented in *Table 58*.

We can conclude that there is no evidence of inconsistency present in this treatment loop ($p = 0.25$). While the mean estimates from the azathioprine studies disagree, there is huge variability around the small Raghu and colleagues[70] study, which rules out any clear inconsistency.

In our sensitivity analysis, an alternative parameterisation was applied for the log-odds ratio, assuming that the data are normally distributed,[93] whereby a multiplier of 1.65 was applied to the mean and standard error standardised mean difference estimates. This resulted in slightly lower treatment effects and narrower CrIs but did not affect statistical significance (*Table 59*).

Fixed-effect results using Cohen's parameterisation were unchanged apart from azathioprine versus placebo, which gave an odds ratio of 1.57 (95% CrI 0.76 to 3.26).

In another sensitivity analysis using the random-effects results, arguably a more conservative scenario, none of the treatments reached statistical significance compared with placebo (*Table 60*).

Excluding the sildenafil study did not affect the log-odds ratios and had negligible impact on the between-study deviation, 0.23 (SD 0.23) compared with 0.24 (SD 0.24), when the study was included.

TABLE 58 Inconsistency estimates based on Bucher's equations

Loop	Pair	Direct, d^D Mean (SD)	Indirect, d^I Mean (SD)	Inconsistency, $\hat{\omega}$ Mean (SD)	*z*-score	*p*-value
Azathioprine, NAC triple, placebo	Azathioprine, placebo	−0.23 (0.46)	0.36 (0.23)	−0.59 (0.52)	−1.14	0.25

TABLE 59 Network meta-analysis fixed-effects results, converted to log-odds ratios for slowing the decline in FVC (assuming normal distribution)

Comparator (vs. placebo)	Log-odds ratios		Odds ratios	
	Mean	95% CrI	Mean	95% CrI
Azathioprine	0.40	−0.27 to 1.08	1.50	0.76 to 2.95
BIBF 1120	−0.88	−1.39 to −0.37	0.41	0.25 to 0.69
NAC triple therapy	−0.05	−0.55 to 0.44	0.95	0.58 to 1.55
Inhaled NAC	−0.38	−1.13 to 0.36	0.68	0.32 to 1.44
Pirfenidone	−0.35	−0.56 to −0.15	0.70	0.57 to 0.86
Sildenafil	−0.11	−0.59 to 0.37	0.90	0.55 to 1.45
Head-to-head comparison				
BIBF 1120 vs. pirfenidone	−0.53	−1.07 to 0.02	0.59	0.34 to 1.03

TABLE 60 Network meta-analysis random-effects results, converted to log-odds ratios for slowing the decline in FVC

Comparator (vs. placebo)	Log-odds ratios		Odds ratios	
	Mean	95% CrI	Mean	95% CrI
Azathioprine	0.27	−1.15 to 1.47	1.32	0.32 to 4.34
BIBF 1120	−0.97	−2.24 to 0.31	0.38	0.11 to 1.36
NAC triple therapy	−0.15	−1.33 to 0.90	0.86	0.27 to 2.45
Inhaled NAC	−0.43	−1.81 to 0.96	0.65	0.16 to 2.61
Pirfenidone	−0.45	−1.12 to 0.15	0.64	0.33 to 1.16
Sildenafil	−0.12	−1.38 to 1.14	0.89	0.25 to 3.13

Existing systematic reviews

Five existing systematic reviews have been identified with relevance to the research question in the present review, and these have been briefly summarised here.[94–98]

Spagnolo and colleagues,[94] in a Cochrane systematic review of non-steroid agents in the treatment of IPF (either as sole agents or in addition to corticosteroid treatment), identified 15 RCTs of relevance. Most of these were placebo-controlled trials and the interventions included were interferon gamma-1 beta, pirfenidone, cyclophosphamide, etanercept, imatinib, bosentan, colchicine, azathioprine, NAC and anticoagulant. Five of the interventions (eight RCTs) were not eligible for inclusion in the current review and two other RCTs of potentially eligible interventions did not meet the inclusion criteria of the present review. Included RCTs that were also included in the present review were three RCTs of pirfenidone[62,65] and one of azathioprine.[70] The date of the last searches of the Cochrane review was April 2010.

A separate Cochrane systematic review, by Richeldi and colleagues,[95] searched for evidence of the effects of corticosteroids in IPF. The date of the last searches was June 2008, at which time no trials were identified that met the inclusion criteria.

A systematic review of interferon gamma-1 beta for IPF was undertaken by Bajwa and colleagues in 2005.[96] This intervention was not eligible for the present systematic review. Four studies were included.

King and colleagues[97] presented the results of a systematic review and meta-analysis of pirfenidone in a conference abstract in 2011. Four RCTs were included, all of which were included in the present review. The abstract was sponsored by InterMune, the sponsor of pirfenidone in the UK, and was discussed in more detail in *Network meta-analysis*.

Ryerson and colleagues[98] focused their systematic review on the treatment and correlates of dyspnoea in IPF. Searches were undertaken until November 2010 and studies of any design were eligible if they included at least 10 adults with IPF. Twenty-nine studies met the inclusion criteria; eight of these were RCTs, all of which were identified in searches of the present systematic review.

Ongoing randomised controlled trials

Twenty-five ongoing studies of possible relevance have been identified. Eight of these were phase III RCTs and are summarised here. Seventeen were phase II studies and are summarised in *Table 61*.

TABLE 61 Ongoing/recently completed phase II studies of potential relevance

Treatment details	Estimated completion	Link
BIBF 1120	Completed	http://clinicaltrials.gov/show/NCT01136174
Pirfenidone	March 2013[a]	http://clinicaltrials.gov/show/NCT01504334
Carbon monoxide	July 2014	http://clinicaltrials.gov/show/NCT01214187
SAR156597	October 2013	http://clinicaltrials.gov/show/NCT01529853
Sirolimus	November 2015	http://clinicaltrials.gov/show/NCT01462006
Pulmonary rehabilitation	December 2013	http://clinicaltrials.gov/show/NCT01499745
Sildenafil, losartan, sildenafil and losartan	May 2013[a]	http://clinicaltrials.gov/show/NCT00981747
STX-100	December 2013	http://clinicaltrials.gov/show/NCT01371305
Lysophosphatidic acid receptor antagonist	January 2015	http://clinicaltrials.gov/show/NCT01766817
Tralokinumab	May 2015	http://clinicaltrials.gov/show/NCT01629667
Simtuzumab	July 2014	http://clinicaltrials.gov/show/NCT01769196
Cognitive–behavioural therapy	February 2014	http://clinicaltrials.gov/show/NCT01738711
GS-6624	Unknown	www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2012–001571–36
VRP700	Unknown	http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2012–005794–31-GB
Lecithinized superoxide dismutase nebuliser	Unknown	http://apps.who.int/trialsearch/Trial.aspx?TrialID=JPRN-UMIN000009572
Oxygen	Unknown	http://apps.who.int/trialsearch/Trial.aspx?TrialID=ACTRN12611001154998
Oxygen	Unknown	http://apps.who.int/trialsearch/Trial.aspx?TrialID=JPRN-UMIN000005098

a Date as per trial register.

A study protocol for a multicentre RCT comparing exercise training with a usual care control was published in February 2013.[63] This RCT aims to recruit 116 participants with ILD and will analyse data for those with IPF in planned subgroup analyses. The primary outcome for this RCT will be the 6MWT distance and follow-up is planned for 6 months. The study will be powered to detect a difference in the change in 6-minute walk distance (6MWD) in the IPF subgroup. The research is funded by the ATS Foundation, the Pulmonary Fibrosis Foundation, the Institute of Breathing and Sleep, the Eirlene Lucas Foundation and the National Health and Medical Research Council (MRC) of Australia. It is not clear from the publication when the study aims to be completed.

A placebo-controlled RCT of BIBF 1120 300 mg/day is currently ongoing. Patient recruitment is complete and the follow-up planned for 52 weeks. The study has an estimated enrolment of 515 participants and the effects of treatment will be tested on the primary outcome measure of annual rate of decline in FVC. The study is funded by Boehringer Ingelheim Pharmaceuticals (http://clinicaltrials.gov/show/NCT01335464, http://clinicaltrials.gov/show/NCT01619085).

A placebo-controlled RCT of pirfenidone is currently ongoing, although patient recruitment is complete. The study, known as the ASCEND trial, has a planned follow-up of 52 weeks and an estimated enrolment of 500 participants. The primary outcome measure is the change in the per cent predicted FVC. The study is funded by InterMune and the estimated final outcome assessment is February 2014 (http://clinicaltrials.gov/show/NCT01366209).

A RCT of sildenafil compared with placebo is currently recruiting participants with either COPD or IPF. It is unclear whether or not the IPF participant data will be reported separately. The study is sponsored by the Rabin Medical Centre, Israel, and plans to recruit 60 participants (http://clinicaltrials.gov/show/NCT01382368).

A placebo-controlled RCT of cotrimoxazole is under way in Spain. Participants will be followed for 24 weeks and efficacy will be evaluated by the decline in FVC per cent predicted. The study is sponsored by Fundación Pública Andaluza para la gestión de la Investigación en Sevilla and is expected to complete in September 2014 (http://clinicaltrials.gov/show/NCT01777737).

An ongoing RCT tested the effectiveness of pulmonary rehabilitation as measured by the 6MWT is expected to be completed in September 2013. The study has a usual care comparator group and is anticipating recruiting 50 participants. The study is funded by the Department of Veterans Affairs (http://clinicaltrials.gov/show/NCT01118221).

Another ongoing RCT is also assessing the effects of a pulmonary rehabilitation programme compared with usual care. The study is sponsored by Klinikum Berchtesgadener Land der Schön-Kliniken and aims to recruit 78 IPF participants. The primary outcome of the 6MWT will be assessed at 3 months' follow-up. The study aims to complete in July 2015 (http://clinicaltrials.gov/show/NCT01772667).

The effectiveness of treatment with fibrangi is being assessed in an ongoing RCT in the Islamic Republic of Iran. The study, funded by Pars Roos Biotechnology Co., is aiming to recruit 20 participants. The primary outcome measure is the DL_{CO} which will be assessed 3 months after cessation of treatment. The recruitment status is complete but there are no details as to when study completion is anticipated (http://apps.who.int/trialsearch/Trial.aspx?TrialID=IRCT138806152425N1).

In addition, the NAC versus placebo arms of the Raghu and colleagues[68] trial are expected to report in 2014.

Chapter 4 Economic analysis

Introduction

The aim of this section is to assess the cost-effectiveness of the different interventions for managing patients with IPF, including pharmacological and non-pharmacological interventions, compared with BSC in the UK. The economic evaluation incorporates:

- a systematic review of the literature on the cost-effectiveness of the different pharmacological and non-pharmacological interventions for managing patients with IPF
- a systematic review of studies assessing HRQoL in people with IPF, and
- an independent economic model developed specifically for this evaluation focusing on those interventions shown to be clinically effective in the systematic review (see *Chapter 3*), presenting the structure, parameterisation and results of the analysis.

Systematic review of existing cost-effectiveness evidence

A systematic review of the literature was undertaken to identify economic evaluations of interventions to manage IPF. The purpose was to assess the current evidence base for the cost-effectiveness of different interventions and to identify whether or not there was a need to conduct further economic modelling. If further modelling is required, the methods and parameters used in previous cost-effectiveness studies will be assessed to help inform the development of a new or updated evaluation.

The sources and search terms used the systematic literature search are described in *Appendix 1* and details of the study selection, data extraction and quality assessment processes are outlined in *Chapter 2* (see *Study selection and data extraction*).

Results of the systematic review

Searches identified a total of 225 references after deduplication; only two full references were retrieved (*Figure 11*). The majority of the 223 references excluded at the first screening for the review were due to an incorrect participant group or study type. Of the two studies retrieved as full papers, only one was judged to have met the inclusion criteria for the systematic review.[99] The other study was a review of an unpublished evaluation, providing insufficient information about the study itself.[100] The characteristics, methodological quality and results of the included economic evaluation are outlined in *Tables 62* and *63* and are discussed below. Data extractions and quality assessment can be seen in *Appendix 9*.

Critical appraisal of the studies

The economic evaluation by Hagaman and colleagues[99] included in the systematic review of cost-effectiveness focused on the evaluation of a test (i.e. TPMT testing) to identify patients with IPF who may incur severe adverse effects (i.e. leukopenia) that are associated with a component of a standard therapy for managing IPF (i.e. azathioprine which is administered with NAC and prednisolone).[99] It compared strategies that used or did not use the test prior to standard treatment and with an additional conservative therapy option. Although the use of the TPMT tests was not the focus of the inclusion criteria for this systematic review, the study also incorporated a comparison of standard treatments for IPF. This allowed an initial assessment of those treatments and, more importantly, an opportunity to examine the structure and data sources for the model with a view to its use in the model being developed for the current evaluation. The study by Hagaman and colleagues[99] evaluation was a cost–utility analysis within the US health-care system.[99] The study provided
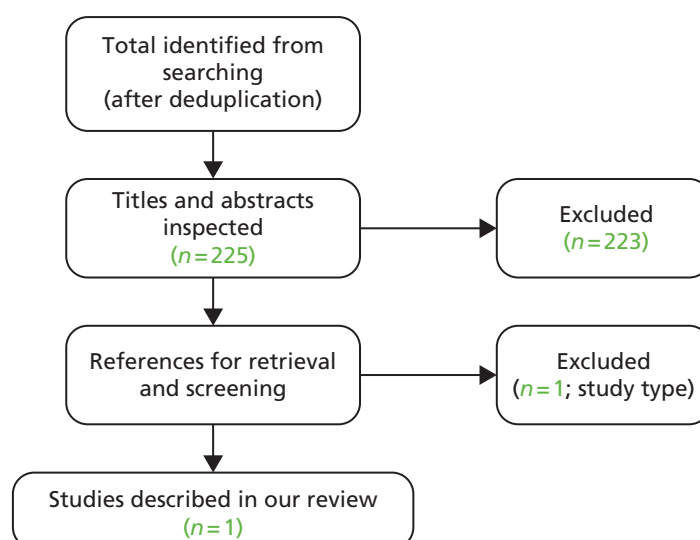
FIGURE 11 Flow chart of identification of studies for inclusion in the review of cost-effectiveness.

TABLE 62 Critical appraisal checklist of economic evaluation[60]

Item	Hagaman *et al.*[99]	Comments
1. Is there a clear statement of the decision problem?	Y[a]	
2. Is the comparator routinely used in UK NHS?	?	The intervention (screening) appears to be in more routine use than no screening
3. Is the patient group in the study similar to those of interest in UK NHS?	?	Limited information is provided about the patient group, with parameter values from different clinical and ethnic groups
4. Is the health-care system comparable with the UK?	?	US health-care system, though limited details provided
5. Is the setting comparable with the UK?	Y	Secondary care
6. Is the perspective of the model clearly stated?	N	No, although costs are clearly stated
7. Is the study type appropriate?	Y	
8. Is the modelling methodology appropriate?	?	Limited details provided. Only considers 1-year time horizon, with all events occurring at 0.5 years on average
9. Is the model structure described and does it reflect the disease process?	Y	
10. Are assumptions about model structure listed and justified?	N	
11. Are the data inputs for the model described and justified?	?	Not all inputs are described and justified (e.g. utility estimates, resource use)
12. Is the effectiveness of the intervention established based on a systematic review?	N	
13. Are health benefits measured in QALYs?	Y	
14. Are health benefits measured using a standardised and validated generic instrument?	?	Not stated
15. Are the resource costs described and justified?	?	Costs are stated, justification for resources not provided

TABLE 62 Critical appraisal checklist of economic evaluation[60] (*continued*)

Item	Hagaman *et al.*[99]	Comments
16. Have the costs and outcomes been discounted?	N	
17. Has uncertainty been assessed?	Y	Limited details provided
18. Has the model been validated?	?	Not stated

?, unclear; N, no; QALY, quality-adjusted life-year; Y, yes.
a States that the comparison will be between azathioprine, NAC and prednisolone with and without TPMT testing before initiating therapy and a conservative therapy strategy (i.e. no specific pharmacological treatment for IPF and using only supportive measures).

TABLE 63 Characteristics of economic evaluations

Characteristics	Study details
Author	Hagaman *et al.*[99]
Publication year	2010
Country	USA
Funding source	Not stated
Study type	Cost–utility analysis
Perspective	Not stated
Study population	IPF patients eligible for azathioprine treatment (no details stated)
Intervention(s)	TPMT screening prior to treatment with azathioprine, NAC and steroids
Intervention effect	Reduced incidence of leukopenia
Currency base	US$ 2007
Model type	Decision tree
Time horizon	1 year
Baseline cohort	Not stated
Base-case results	ICER of US$29,663 for comparison of azathioprine, NAC and prednisolone with and without prior TMPT testing

ICER, incremental cost-effectiveness ratio.

clear statements of the decision problem being assessed, used an appropriate model structure to reflect the disease process, included measures of benefit, resource use and cost and undertook an assessment of uncertainty through sensitivity analyses. It is less clear whether or not the study was relevant to the UK setting in terms of the comparators used, the patient group included or the nature of the service provided. It is also uncertain whether or not the modelling methodology was appropriate, the parameter values used were justified, the health benefits used were valid generic measures, and whether or not the modelling had been validated. No information was provided about the perspective, assumptions or discount rates used and it appears than none of the parameters were obtained through a systematic approach. While many of the apparent limitations in the evaluation may reflect the lack of detail in the reporting of the study, it causes uncertainty about the adequacy of the approach taken and the validity of the results.

Modelling approach

Hagaman and colleagues used a decision tree model to examine the cost-effectiveness of the addition of TPMT testing for people with IPF prior to initiating standard treatment.[99] The model evaluated three treatment arms for managing IPF, specifically azathioprine, NAC and steroids (prednisolone) 'triple therapy,' with and without prior TPMT testing, and a conservative therapy option (no specific pharmacological treatment for IPF, only supportive measures).

Patients undergoing TPMT testing prior to treatment were categorised into three different treatment groups depending on TPMT activity. Patients with normal TPMT activity received standard doses of azathioprine, NAC and prednisolone at the outset of treatment. Intermediate TPMT activity resulted in patients receiving a reduced dose of azathioprine and standard doses of NAC and prednisolone. Where TPMT activity was absent, patients received conservative therapy. Although thresholds for defining TPMT activity and dosages for the different regimens are not specified, references are provided for guidance on recommended levels. Patients who did not undergo TPMT testing received either standard doses of azathioprine, NAC and prednisolone or conservative therapy at the start of treatment.

Subsequent changes in treatment were incorporated within the model. Although Hagaman and colleagues state that this was based on clinical practice, no information was provided to justify the treatments subsequently given.[99] When patients suffered from severe complications following treatment with azathioprine, NAC and prednisolone, specifically leukopenia requiring hospitalisation, and survived, they were crossed over to receive conservative therapy. Where leukopenia following treatment with azathioprine, NAC and prednisolone was less severe and could be managed on an outpatient basis, the dose of azathioprine was reduced. Hagaman and colleagues state that other non-TPMT complications related to gastrointestinal symptoms may require discontinuation of treatment; however, they do not indicate the subsequent management strategies used.[99] Patients who survive leukopenia or who do not experience leukopenia may remain stable without progression, incur progression of IPF or die.

No systematic review was undertaken to identify evidence on the prevalence of TPMT activity, efficacy of treatment with azathioprine, NAC and prednisolone or conservative therapy, the occurrence of complications or for estimates on HRQoL. Instead, Hagaman and colleagues identified specific studies providing an estimate of the different parameters, which do not always focus solely on the relevant patient group.[99] Prevalence of TPMT activity originated from a population-based cohort study of the distribution of phenotypic TPMT activity in 14,545 Spanish patients with diseases amenable to treatment with azathioprine.[101] Gisbert and colleagues[101] found that TPMT activity was low in 0.5% of the cohort, intermediate in 11.9% and high in 87.6%. Treatment efficacy was assessed through disease progression, defined by a decline in FVC of at least 10%. A RCT (included in the clinical systematic review) comparing azathioprine, NAC and prednisolone with NAC and prednisolone in patients with IPF found that 37% of the former had progressed at 12 months compared with 51% in the latter (statistically significant difference – *p*-value not stated).[67] The occurrence of leukopenia by TPMT activity was taken from two studies of patients with inflammatory bowel disease who were treated with azathioprine.[102,103] A weighted average was calculated, showing that 5% of patients with normal TPMT activity, 21.4% with intermediate activity and 100% with no activity developed leukopenia following full-dose azathioprine. It was assumed from studies in alveolar haemorrhage and atopic dermatitis that a reduced dose would half the risk to 10%.[104,105] Estimates of HRQoL are presented for progression of IPF (base case 0.63, range for sensitivity analysis 0.5 to 0.9), leukopenia (0.95, range 0.92 to 0.98) and complicated leukopenia (0.76, range 0.66 to 0.86). Although there is no discussion about the HRQoL data, the studies from which they originate include patients with cancer and IPF.

The model adopted a 1-year time horizon, with all events (i.e. treatment-related adverse events, disease progression and death) assumed to occur on average at 0.5 years. It is unclear whether or not the time horizon was sufficient to capture all events, given that median survival ranged from 2 to 4 years from diagnosis. Limited information was provided concerning the basis for the structure of the model, the assumptions used or the parameterisation of the model.

Estimation of quality-adjusted life-years

Hagaman and colleagues estimated that patients receiving azathioprine, NAC and prednisolone with prior TPMT testing would gain 2.62 quality-adjusted life years (QALYs) during the 12-month time horizon, while those not undergoing TPMT testing would gain 2.61 QALYs. In contrast, patients receiving conservative therapy would gain 2.50 QALYs.[99]

Estimation of costs

Hagaman and colleagues included direct costs only in the evaluation, which were in 2007 US$.[99] Estimates were obtained from routinely published sources, including medication costs from www.drugstore.com and costs for procedures, office visits and hospitalisations from average Medicare reimbursement costs using current procedural terminology or diagnosis-related group codes. Costs of adverse events include both institutional and professional services. Detailed microcostings for baseline costs are provided; however, the basis for decisions concerning resource use and values for sensitivity analyses are not discussed. No source is provided for the cost of the TPMT assay. A summary of the cost estimates is provided in *Appendix 9*. Hagaman and colleagues estimated that the total cost per patient using azathioprine, NAC and prednisolone with prior TPMT testing was US$15,818 and without TPMT testing was US$15,802. The cost of treatment with conservative therapy was US$9691 per patient.[99]

Cost-effectiveness results

The base-case results presented by Hagaman and colleagues showed that azathioprine, NAC and prednisolone with prior TPMT testing was the most effective option and also the most costly.[99] The strategy without TPMT testing before treatment with azathioprine, NAC and prednisolone was excluded by extended dominance, having a higher incremental cost-effectiveness ratio (ICER) than other more expensive options. The base-case ICER for azathioprine, NAC and prednisolone with prior TPMT testing compared with conservative therapy was US$49,156 per QALY. Comparison of azathioprine, NAC and prednisolone with and without prior TPMT testing was undertaken, excluding extended dominance, which resulted in an ICER of US$29,663 per QALY.

Sensitivity analyses were undertaken on all parameters in the model, with the frequency of abnormal TPMT alleles, the probability of developing leukopenia and treatment efficacy reported as having considerable impact on the ICERs. An increase in the prevalence of abnormal TPMT alleles to ≥ 12% produced ICERs below US$50,000 per QALY when compared with both the no-TPMT testing strategy and conservative therapy. When frequencies of abnormal TPMT alleles are above 13.5%, the TPMT testing strategy dominates the no TPMT testing strategy. Increasing the probability of leukopenia following treatment with reduced-dose azathioprine in those with intermediate TPMT activity resulted in an increase in the cost relative to the non-TPMT testing strategy. As a result, the TPMT testing strategy was no longer cost-effective when the probability increased to be above 12%. Comparison of treatment efficacy between azathioprine, NAC and prednisolone following TPMT testing and conservative therapy showed that patients receiving the former had a consistently lower probability of progression than the latter (approximately 13% lower) at the US$50,000 willingness-to-pay (WTP) threshold (i.e. as rates of progression on conservative therapy increase, the strategy with TPMT testing remains cost-effective at increasingly high rates of progression following its use). It is stated that other parameters used in the model had less effect on the results. Unfortunately, details concerning any of the sensitivity analyses are limited and therefore it is difficult to interpret the robustness of the model in terms of structural or parameter uncertainty.

Summary

A systematic review of cost-effectiveness studies identified one economic evaluation of interventions to manage patients with IPF.[99] Its primary focus was on the benefits of a testing strategy prior to treatment with a standard intervention, with the intention of limiting the occurrence of serious adverse effects. Although the outcomes of the evaluation may not have been strictly relevant to the current report, the evaluation provides a basis for developing a de novo model (see *Independent economic evaluation*) and for providing parameters for the evaluation.

Systematic review of health-related quality of life studies

A systematic review was undertaken to assess the HRQoL of people with IPF. The aim of the review was to provide data to populate the economic model with health-state utility values appropriate to conduct a cost–utility analysis and to calculate the benefits of interventions in terms of QALYs. The sources and search terms used are described in *Appendix 1*. Studies reporting HRQoL in people with IPF were eligible for inclusion if they used either generic preference-based measures or the SGRQ, which is a disease-specific instrument used in IPF. Other recently developed disease-specific instruments such as the St George's Respiratory Questionnaire IPF (SGRQ-I),[106] the 'A Tool to Assess QOL in IPF' (ATAQ-IPF)[107] and the King's Brief Interstitial Lung Disease (K-BILD)[108] measures were not eligible for inclusion as there are currently no methods to map results of these to utility measures required for economic evaluation. Data from any study designs were eligible for inclusion in the systematic review. The inclusion and exclusion criteria for the review are shown in *Table 64*.

The search strategy identified 423 articles that were potentially relevant and one additional publication was identified in the searches for the systematic review of clinical effectiveness (reported in *Chapter 3*). The titles and abstracts were screened, with the full text of 52 studies retrieved for further inspection. After checking the retrieved papers, 23 studies met the inclusion criteria. A summary of the selection process is presented in *Figure 12*. A list of the excluded studies is shown in *Appendix 10*.

Nine studies[61,109–116] assessed HRQoL using the SF-36, seven used the SGRQ,[55,73,117–122] five studies used both the SF-36 and SGRQ[123–127] and two studies used the SF-36, the SGRQ and the EQ-5D[68,72] (*Table 65*).

TABLE 64 Inclusion/exclusion criteria for HRQoL of people with IPF

Category	Criteria
Patients	People with IPF. If data are lacking, participants with other ILDs will also be considered eligible for inclusion
Outcomes	Generic, preference-based (VAS/TTO/SG) measures such as EQ-5D, SF-36/6D and HUI were eligible. The disease-specific SGRQ was also included
Study design	Studies eligible for inclusion were either primary studies (prospective or retrospective observational studies) collecting QoL data, or trials where QoL data were reported as outcomes

HUI, Health Utilities Index; SF-36/6D, Short-Form 36 or 6 dimensions; SG, standard gamble; TTO, time trade-off.
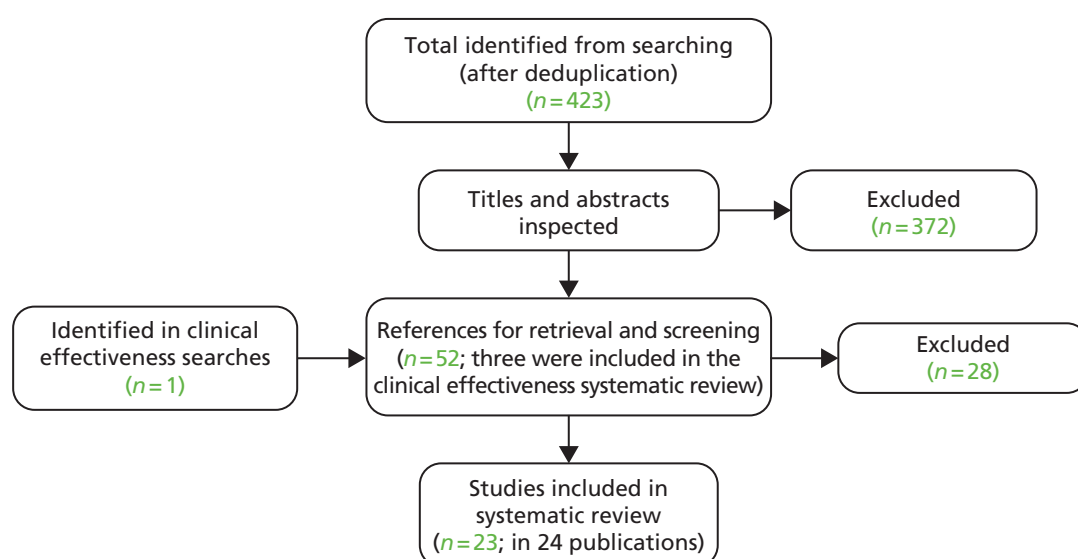
FIGURE 12 Flow chart of identified studies for HRQoL review in IPF.

TABLE 65 Characteristics of studies included in HRQoL systematic review

Study details	Study population	Details of HRQoL instrument(s)
Studies reporting EQ-5D (also SF-36 and SGRQ)		
Raghu _et al._ IPFCRN 2012[68] Country: USA Setting: outpatient clinic RCT Triple therapy compared with placebo (also included in _Chapter 3_)	Number: 155 (another 81 but data not shown) Age: ≈ 68 years M/F (%): 116/39 (75/25) Diagnosis: based on ATS/ERS criteria FVC per cent predicted: ranged between 69% and 72%	EQ-5D index EQ-5D VAS SF-36 (aggregate scores only) SGRQ
Zisman _et al._ IPFCRN 2010[72] Country: USA Setting: outpatient clinic RCT Sildenafil compared with placebo (also included in _Chapter 3_)	Number: 180 Age: ≈ 69 years M/F (%): 150/30 (83/17) Diagnosis: based on ATS/ERS criteria FVC per cent predicted: ranged between 55% and 59%	EQ-5D index EQ-5D VAS SF-36 SGRQ
Studies reporting SF-36 alone		
Mermigkis _et al._ 2013[116] Country: Greece Setting: outpatient clinic Single-arm cohort study (before and after) Continuous positive airways pressure therapy	Number: 12 Age: 67.1 years M/F (%): 10/2 (83/17) Diagnosis: ATS/ERS criteria FVC per cent predicted: 77.1%	SF-36 Overall score only presented
Kozu _et al._ 2011[113] Country: Japan Setting: outpatient clinic Single-arm cohort study (before and after) Pulmonary rehabilitation programme (8 weeks)	Number: 65 Age: 67.5 years M/F (%): 46/19 (70/30) Diagnosis: based on ATS/ERS criteria FVC per cent predicted: ranged between 51% and 83%	SF-36 Data presented by MRC dyspnoea grade
Kozu _et al._ 2011[114] Country: Japan Setting: outpatient clinic Cohort study (before and after) and additional control group Pulmonary rehabilitation programme (8 weeks)	Number: 45 Age: 67.5 years M/F (%): 37/8 (82/18) Diagnosis: based on ATS/ERS criteria FVC per cent predicted: 68.6%	SF-36

TABLE 65 Characteristics of studies included in HRQoL systematic review (*continued*)

Study details	Study population	Details of HRQoL instrument(s)
Ozalevli *et al.* 2010[110]	Number: 15	SF-36
Country: Turkey	Age: 62.8 years	
Setting: participants' home	M/F (%): 10/5 (67/33)	
Single-arm cohort study (before and after)	Diagnosis: based on ATS/ERS criteria	
Home-based pulmonary rehabilitation programme	FVC% predicted: 71.6%	
Jastrzebski *et al.* 2008[61]	Number: 30	SF-36
Country: Poland	Age: 56.35	
Setting: Outpatient clinic	M/F (%): 19/11 (63/37)	
CCT	Diagnosis: based on ATS/ERS criteria	
Inspiratory muscle training	FVC per cent predicted: 69.2%	
Tomioka *et al.* 2007[115]	Number: 46	SF-36 presented only as deviation values against national reference values
Country: Japan	Age: 69.9 years	
Setting: outpatient clinic	M/F (%): 32/14 (70/30)	
Cohort study (cross-sectional and longitudinal)	Diagnosis: based on ATS/ERS criteria	
No intervention	FVC per cent predicted: 71.0%	
Jastrzebski *et al.* 2005[112]	Number: 16	SF-36
Country: Poland	Age: 48.3 years	
Setting: outpatient clinic	M/F (%): 11/5 (69/31)	
Cohort study (cross-sectional and longitudinal) with additional control group	Diagnosis: based on ATS/ERS criteria and meeting criteria for lung transplant	
No intervention	FVC per cent predicted: 44.0%	
Baddini Martinez *et al.* 2002[111]	Number: 30	SF-36
Country: Brazil	Age: 58.6 years	
Setting: outpatient clinic	M/F (%): 18/12 (60/40)	
Cohort study	Diagnosis: by lung biopsy or HRCT, criteria used not reported	
No intervention	FVC per cent predicted: 61.9%	
Martinez *et al.* 2000[109]	Number: 34	SF-36
Country: Brazil	Age: 58.29 years	
Setting: outpatient clinic	M/F (%): 20/14 (59/41)	
Matched controlled study. Control group matched for sex and age	Diagnosis: by lung biopsy or HRCT, criteria used not reported	
No intervention	FVC per cent predicted: 62.41%	

TABLE 65 Characteristics of studies included in HRQoL systematic review (*continued*)

Study details	Study population	Details of HRQoL instrument(s)
Studies reporting SF-36 and SGRQ		
Verma *et al.* 2011[123]	Number: 137	SF-36
Country: Canada	Age: 59.4 years	SGRQ
Setting: outpatient clinic	M/F (%): 90/47 (66/34)	
Cohort study	Diagnosis: based on ATS/ERS criteria	
No intervention	FVC per cent predicted: 61.7%	
Raghu *et al.* 2010[125]	Number: 158	SF-36
Country: Europe (including UK), USA, Canada, Israel	Age: 65.2 years	SGRQ
Setting: outpatient clinic	M/F (%): 112/42 (71/29)	
RCT (post-hoc analysis)	Diagnosis: based on ATS/ERS criteria	
Bosentan compared with placebo	FVC per cent predicted: 67.7%[a]	
Swigiris *et al.* 2010[126]	Number: 158	SF-36
Country: Europe (including UK), USA, Canada, Israel	Age: 65.12 years	SGRQ
Setting: outpatient clinic	M/F (%): 73/27 (73/27)	
Retrospective analysis of RCT data	Diagnosis: based on ATS/ERS criteria	
Bosentan compared with placebo	FVC per cent predicted: 66.97%	
Lutogniewska *et al.* 2010[127]	Number: 30	SF-36
Country: Poland	Age: not reported	SGRQ
Setting: unknown	M/F (%): not reported	
Single-cohort study with subgroup data for IPF	Diagnosis: not reported	
No intervention	FVC per cent predicted: not reported	
Zimmerman *et al.* 2007[124]	Number: 20	SF-36
Country: Brazil	Age: 61.4 years	SGRQ
Setting: outpatient clinic	M/F (%): 12/8 (60/40)	
Cohort study	Diagnosis: based on ATS/ERS criteria	
No intervention	FVC per cent predicted: 70.4%	
Horton *et al.* 2012[55,75]	Number: 23	SGRQ
Country: USA	Age: 67.6 years	
Setting: university	M/F (%): 18/5 (78/22)	
RCT	Diagnosis by lung biopsy or HRCT, criteria used not reported	
Thalidomide	FVC per cent predicted: 70.4%	

TABLE 65 Characteristics of studies included in HRQoL systematic review (*continued*)

Study details	Study population	Details of HRQoL instrument(s)
Nishiyama *et al.* 2012[122]	Number: 87	SGRQ
Country: Japan	Age: 66.3 years	
Setting: outpatient clinic	M/F (%): 77/10 (89/11)	
Retrospective cohort study	Diagnosis by ATS/ERS criteria	
No intervention	FVC per cent predicted: 75.0%	
Nishiyama *et al.* 2008[73]	Number: 28	SGRQ
Country: Japan	Age: 66.3 years	
Setting: outpatient clinic	M/F (%): 21/7 (75/25)	
RCT	Diagnosis: based on ATS/ERS criteria	
Pulmonary rehabilitation compared with control (also included in *Chapter 3*)	FVC per cent predicted: 67.4%	
De Vries *et al.* 2000[117]	Number: 10	SGRQ
Country: the Netherlands	Age: 61.1 years	WHOQOL-100 (data not extracted)
Setting: university hospital setting	M/F (%): 4/6 (40/60)	
Focus groups/cross-sectional survey	Diagnosis: not stated	
No intervention	FVC per cent predicted: not reported	
Peng *et al.* 2008[118]	Number: 68	SGRQ
Country: China	Age: 64 years	
Setting: outpatient clinic	M/F (%): 54/14 (79/21)	
Cohort study (cross-sectional and longitudinal)	Diagnosis: by lung biopsy or HRCT	
	FVC per cent predicted: 66%	
No intervention		
Nishiyama *et al.* 2005[119]	Number: 41	SGRQ
Country: Japan	Age: 64 years	
Setting: outpatient clinic	M/F (%): 35/6 (85/15)	
Cohort study	Diagnosis: based on ATS/ERS criteria	
No intervention	FVC per cent predicted: 76.6%	
Tzanakis *et al.* 2005[120]	Number: 25	SGRQ
Country: Greece	Age: 66 years	
Setting: outpatient clinic	M/F (%): 21/4 (84/16)	
Matched controlled study	Diagnosis: based on ATS/ERS criteria	
No intervention	FVC per cent predicted: 68.8%	

F, female; M, male; WHOQOL, World Health Organization Quality of Life.
a Reported in linked publication;[128] see *Chapter 3*.

Two studies used the EQ-5D.[68,72] Both of these studies were also included in the systematic review of clinical effectiveness (see *Chapter 3*). The EQ-5D is a widely used instrument developed to measure health utility. It comprises five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) which can be rated with either three or five levels of severity (ranging from no problems to extreme problems) depending on the precise measure chosen.[129,130] Scores from the health states are converted through value sets (or preference weights) obtained from general population studies to generate a single utility value.[129,130] This utility ranges from 0.59 to 1 and represents health on a continuum from worst to best (negative values are valued as worse than dead). The value sets applied can be derived using either a VAS or time trade-off (TTO) technique. The former requires members of the general population to indicate where they believe an individual health state should be positioned on a scale ranging from worst imaginable health to best imaginable health. The TTO technique requires members of the general population to imagine that they live in a particular health state for 10 years and to then specify the amount of time they are willing to give up to live in full health. Applying these population value sets to the data allows generation of a utility value which has a societal perspective. This is, therefore, appropriate for use in health economic evaluations. The EQ-5D also has a VAS where individuals are required to rate their health on a vertical (thermometer) score ranging from worst imaginable health state (0) to best imaginable health state (100). This is a patient-based score which, therefore, is not representative of the general population, and as a result is not appropriate for use in health economic evaluation.[130]

Sixteen studies reported the SF-36;[61,68,72,109–116,123–127] one of these studies[61] was also included in the systematic review of clinical effectiveness (see *Chapter 3*). The SF-36 is a generic measure of health status. It is a 36-item survey and has eight scales covering functional health and well-being concepts.[131] These are physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health. The eight scales are aligned to their demonstrated validity to measure either physical health or mental health to aid the interpretation of the SF-36.[132] The first scale (physical functioning) has been shown to be the most valid measure of physical health, and the last scale (mental health) the most valid measure of mental health. The scales in between these are ordered according to their respective validity in measuring concepts of physical or mental health, with vitality and general health being moderately valid for both aspects of health.[132] Scores can also be combined to generate physical and mental health summary scores. Normative values for the UK have been published by a number of groups, including the *Health Survey for England*[133] and the ONS Omnibus Survey of Britain.[134] While there are limitations associated with any source of normative data, for example the demographic composition of the populations used will vary, these are essential to interpret results of the measure for the population of interest. For pragmatic reasons for the purposes of this systematic review, the normative data from the ONS has been used to aid the interpretation of the effects IPF has on HRQoL, as this has a reasonably wide UK base. Many of the included studies were not from UK populations (see below) and therefore observations on the impact of IPF on HRQoL, as measured by the SF-36, made in the current review should be treated as illustrative only.

Fourteen studies report data on the SGRQ.[55,68,72,73,117–127] The SGRQ is a disease-specific measure which aims to assess the impact of obstructive airways disease on overall health, daily life and perceived well-being.[135] It comprises 50 items within two parts, which cover three components of symptoms, activity and impact. A total score can also be calculated which summarises the impact of the disease on overall health status. Scores range from 0 to 100, where higher scores indicate greater negative impact on HRQoL. Few studies have investigated normative values for general populations. Population norms provided in the SGRQ manual[136] have been used to aid the interpretation of the data from the studies included in this systematic review (see *Results*, *St George's Hospital Respiratory Questionnaire* for more details).

Only one multicentre study (two publications from the same study[125,126]) included participants from centres within the UK. Two other multicentre studies were US-based only.[68,72] The remaining studies were single-centre studies from countries such as Japan (six studies[73,112–115,122]), Brazil (three studies[109,111,124]), Poland (three studies[61,112,127]), Turkey (one study[110]), Canada (one study[123]), the Netherlands (one study[117]),

China (one study[118]), Greece (two studies[116,120]) and the USA (one study[55]). Generalisability of these populations to those seen in current UK practice may, therefore, be limited. Most of the included studies were undertaken in hospital outpatient clinics. Sample sizes were generally small, ranging from 12 to 180 participants, with 15 studies including fewer than 50 participants. The average age of the participants ranged from 48 years to 69 years. Males were the dominant sex in all but one study,[117] with proportions of males ranging from approximately 58% to 85% in the remaining studies.

All studies were published after 2000. The diagnosis of IPF was according to the ATS/ERS criteria in 17 studies;[61,68,72,73,110,112–116,119,120,122–126] four studies stated that surgical lung biopsy or HRCT were used but did not state what criteria were used to assess the findings of these tests.[55,109,111,118] Two studies[117,127] did not report how the diagnosis was made. Per cent predicted FVC was reported in 21 of the included studies.[55,61,68,72,73,75,109–116,118–120,122–126] This ranged between approximately 62% and 77% in the majority of studies ($n = 18$) (see *Table 65*); one study had a much milder population (83%)[113] and two studies had more severe populations ($\approx 57\%$)[72] and (44%).[112] Therefore, there were a mixture of mild, moderate and severe cases of IPF if per cent predicted FVC is used as a proxy for disease severity.

Study designs varied. Six studies[55,68,72,73,125,126] were based on five RCTs of interventions where HRQoL was an outcome measure, one was a CCT,[61] and four studies were cohort 'before-and-after' studies that tested HRQoL at baseline and then again after an intervention;[110,113,114,116] one of these also had a control group of patients with COPD.[114] Five studies[111,119,122–124] were single-cohort studies with no intervention, three[112,115,118] were single-cohort studies that collected HRQoL data at two different time points (one also had an additional control group[112]), two studies[109,120] were cohort studies with matched controls, one was a cohort study with subgroup data[127] and one[117] was a combination of a focus group and survey. Seven of the 16 studies using the SF-36 did not report the overall physical or mental health summary scores,[109–111,113,114,124,125] one study reported only the overall summary scores (at baseline),[68] one study reported only an overall score encompassing both physical health and mental health component scores,[116] and one study did not present any useful data, reporting data as deviation values from national reference values.[115]

The limitations of the methodological designs of these studies and potential for bias should be considered in the interpretation of these data.

Results

European Quality of Life-5 Dimensions
Two RCTs report the EQ-5D at baseline and after an active intervention.[68,72] As comparisons between groups were not required for the purpose of this systematic review of HRQoL, the placebo group baseline data have been summarised (*Table 66*; for further study details see *Chapter 3*). The baseline FVC per cent predicted in one RCT was 72% in the placebo group,[68] indicating a mild to moderate population, and 58% in the second RCT, indicating a moderate to severe IPF population.[72] The EQ-5D index in the placebo group of the mild to moderate population was 0.8 (SD 0.2) and 0.74 (SD 0.19) in the moderate

TABLE 66 European Quality of Life-5 Dimensions scores from placebo group of included RCTs

Outcome	Raghu *et al.* IPFCRN 2012[68]	Zisman *et al.* IPFCRN 2010[72]
FVC per cent predicted at baseline	72.1 (14.1)	58.73 (14.12)
EQ-5D Index at baseline	0.8 (0.2)	0.74 (0.19)
EQ-5D VAS (SD) at baseline	78.1 (15.4)	67.66 (16.98)

to severe population, as seen in *Table 66*. Neither study reports what population tariff was applied to the patient-generated scores on the EQ-5D; however, given that the RCTs were both undertaken by the same research group in the USA, we have assumed that US preference weights were applied to generate the EQ-5D index score in both trials.

Both of these RCTs also presented data on the EQ-5D VAS; these were 78.1 (SD 15.4)[68] and 67.66 (SD 16.68)[72] in the placebo groups at baseline in the two studies, respectively.

Short Form questionnaire-36 items

Thirteen studies[61,72,109–114,123–127] provided data for the individual scales of the SF-36 and two[68,116] presented data only for the aggregate scales (*Table 67*). Four studies[68,72,112,125] were RCTs that provided baseline data for both treatment arms, but as comparisons between arms were not required for the purpose of this systematic review, the placebo group population data have been summarised (the treated group population data can be seen in *Appendices 5* and *11*). Twelve studies included participants with a range of FVC per cent predicted at baseline of approximately 58–72% (see *Table 67*), including subgroups within this range in one study.[61,68,72,109–111,113,114,123–126] In another study, the baseline characteristics of the participants with FVC per cent predicted indicated more severe IPF participants (44%)[112] and one other study[113] provided outcomes for four severity groups based on the MRC dyspnoea scale, and their corresponding mean FVC per cent predicted ranged from 83% to 51% as can be seen in *Table 67*. One study, in participants who were listed for lung transplantation, did not report the baseline FVC per cent predicted.[127]

Average (mean or median) scores on the different SF-36 scales varied across the studies, and this is likely in part to be a reflection of the different country of origin of the studies, the sample sizes which for many studies were small, and differences in participants including their ages, sex distribution and the severity of IPF. To aid interpretation we have crudely grouped the studies into those with an average baseline FVC per cent predicted 71–83% (to reflect a 'mild' IPF patient group); those with an average baseline FVC per cent predicted of 55–70% (to reflect a 'moderate' IPF patient group); and those with an average baseline FVC per cent predicted of < 55% (to reflect a 'severe' IPF patient group), in line with Nathan and colleagues.[22] However, interpretation of the ranges of results should be treated with caution. The one study that did not report FVC per cent predicted at baseline was categorised in the moderate group as participants were listed for lung transplantation.[127]

Four studies included populations or subgroups that had a mean FVC per cent predicted within the range of 71–83% ('mild' IPF),[68,110,113,116] two of which included data for the individual components of the scale.[110,113] The scores on the physical functioning scale of the SF-36 ranged between 55 and 56 and the scores on the mental health scale ranged between 50 and 62. Of the eight scales on the SF-36, these two have been shown to be the most valid measures of their respective constructs. Of the remaining scales relating to physical health, scores in these two populations ranged from 25 to 56 (role physical) and about 67 (bodily pain). Of the remaining scales relating to mental health, scores in these two populations ranged from 63 to 76 (social functioning) and from 29 to 67 (role emotional). The two scales that encompass elements of both physical health and mental health (general health and vitality) had ranges of scores of 51–57 and 52–55, respectively. The two studies show wide variance in some of the scales but overall it appears that IPF, in people who may crudely be characterised as having IPF of 'mild' severity, has a detrimental effect on HRQoL as measured by the SF-36 when compared observationally with age-standardised population norms for the UK[131] (*Table 68*).

Eleven studies included populations that may reflect a moderately severe IPF patient group (baseline per cent predicted FVC 55–70%).[61,72,109,111,113,114,123–127] In one study, there were two patient subgroups falling within this range, one with a baseline FVC per cent predicted of 60% and one with 67%, and these have been treated as two separate cohorts for the present review.[113] *Table 67* shows the scores on the eight domains of the SF-36 as measured in these studies; as can be seen, there are large variations in scores. The scores on the physical functioning and mental health scales ranged from 19 to 54 and from 42 to 76,

TABLE 67 Short Form questionnaire-36 items scores at baseline, mean (SD) unless stated

Study[a]	N	FVC per cent predicted	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
'Mild' IPF												
Mermigkis et al. 2013[116]	12	77.1%	–	–	–	–	–	–	–	–	63.2 (13.9)	–
Raghu et al. IPFCRN 2012[68] Placebo group	78	72.1%	–	–	–	–	–	–	–	–	55.7 (7.4)	40.6 (9.3)
Ozalevli et al. 2010[110]	15	71.6%	56.0 (5.7)	25.0 (1.7)	67.3 (2.6)	57.0 (4.6)	52.0 (4.9)	75.8 (2.7)	29.0 (1.3)	49.9 (6.7)	–	–
Kozu et al. 2011[113] One of four subgroups	16	83.0%	55.3 (7.2)	55.9 (15.9)	66.5 (25.1)	50.9 (11.0)	54.7 (11.7)	62.5 (18.8)	66.7 (15.2)	61.6 (14.3)	–	–
'Moderate' IPF												
Zisman et al. IPFCRN 2010[72] Placebo group	91	58.73%	31.18 (8.31)	36.38 (11.44)	49.72 (10.44)	37.66 (8.73)	45.30 (9.88)	43.06 (10.17)	44.0 (13.60)	50.95 (8.59)	50.58 (9.52)	34.84 (8.69)
Verma et al. 2011[123]	137	61.7%	25.6 ([b]21.9 to 29.4)	31.0 ([b]26.9 to 35.1)	68.0 ([b]63.4 to 72.7)	35.5 ([b]31.9 to 39.0)	39.3 ([b]35.8 to 42.9)	59.2 ([b]54.05 to 64.38)	74.5 ([b]69.6 to 79.3)	71.2 ([b]67.9 to 74.6)	29.4 ([b]27.9 to 30.9)	49.7 ([b]47.6 to 51.7)
Kozu et al. 2011[113] Two of four subgroups	17	67.0%	34.1 (18.4)	22.4 (17.3)	57.2 (29.0)	35.8 (21.5)	37.9 (21.5)	42.6 (27.6)	47.1 (28.2)	42.9 (20.8)	–	–
	17	60.0%	20.3 (7.0)	23.2 (13.4)	65.6 (29.1)	24.1 (16.8)	26.5 (18.0)	36.0 (15.2)	30.9 (21.4)	41.8 (17.2)	–	–
Raghu et al. 2010[125] Placebo group	158	67.7%	47.5[c] ([b]35.0 to 55.0)	25.0[c] ([b]25.0 to 50.0)	64.0[c] ([b]61.0 to 74.0)	46.0[c] ([b]40.0 to 55.0)	50.0[c] ([b]40.0 to 55.0)	81.3[c] ([b]62.5 to 87.5)	100.0[c] ([b]6.7 to 100.0)	76.0[c] ([b]68.0 to 80.0)	–	–

Study[a]	N	FVC per cent predicted	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
Swigiris et al. 2010[126]	158	66.97%	35.4 (10.3)	37.8 (11.6)	47.6 (10.7)	37.8 (9.4)	43.1 (9.2)	44.6 (12.3)	42.7 (14.2)	48.2 (10.1)	37 (10)	44.2 (10.84)
Lutogniewska et al. 2010[127]	30	NR	19.1 (17.0)	20.7 (23.2)	47.3 (25.9)	28.5 (10.7)	38.7 (19.8)	33.4 (26.0)	40.3 (36.9)	49.2 (22.8)	25.9 (7.8)	42.4 (14.2)
Kozu et al. 2011[114]	45	68.6%	35.7 (18.7)	33.9 (21.7)	62.4 (30.3)	34.7 (19.9)	38.3 (21.3)	48.3 (23.7)	36.5 (30.1)	50.0 (18.7)	–	–
Jastrzebski et al. 2008[61]	14	69.2%	54.3 (17.4)	44.6 (24.4)	66.8 (22.2)	37.4 (11.1)	52.5 (13.3)	58.0 (14.4)	69.0 (44.3)	65.1 (17.9)	36.1 (9.1)	46.5 (10.9)
Zimmermann et al. 2007[124]	20	70.4%	46 (18.3)	57.5 (39.8)	60.6 (31.9)	53.7 (24.1)	49.2 (24.3)	56.9 (32.2)	46.6 (39.5)	66.8 (17)	–	–
Baddini Martinez et al. 2002[111]	30	61.9%	40.83 ([d]4.41)	44.17 ([d]8.86)	78.57 ([d]4.40)	53.07 ([d]4.33)	50.50 ([d]5.10)	60.83 ([d]7.26)	64.44 ([d]8.74)	59.47 ([d]4.15)	–	–
Martinez et al. 2000[109]	34	62.41%	42.79 ([d]4.40)	44.12 ([d]8.11)	76.91 ([d]4.16)	53.50 ([d]3.90)	50.44 ([d]4.88)	60.29 ([d]6.69)	60.78 ([d]8.26)	57.33 ([d]4.07)	–	–
'Severe' IPF												
Kozu et al. 2011[113] One of four subgroups	15	51%	16.0 (9.1)	19.6 (10.3)	65.6 (28.4)	19.6 (15.3)	19.6 (15.3)	30.0 (14.8)	19.4 (15.3)	35.0 (12.0)	–	–
Jastrzebski et al. 2005[112]	16	44.0%	45	43	62	28	38	58	65	50	35	42

BP to bodily pain; GH to general health; MCS to mental cumulative score; MH to mental health; NR to not reported; PCS to physical cumulative score; PF to physical functioning; RE to role emotional; RP to role physical; SF to social functioning; VT to vitality.
a Two additional studies[115,137] used the SF-36 but did not present data in a useful format. Kozu and colleagues[113] provided outcomes for four severity groups based on the MRC dyspnoea scale. These subgroups have been displayed in the table under the 'mild' to 'moderate' and 'severe' categories.
b 95% CI.
c Median scores reported.
d Standard error.

TABLE 68 Short Form questionnaire-36 items dimension norms from British ONS survey, for those aged 55–84 years[131]

Dimensions by age and sex		PF, mean (SD)	RP, mean (SD)	BP, mean (SD)	GH, mean (SD)	VT, mean (SD)	SF, mean (SD)	RE, mean (SD)	MH, mean (SD)
Sex	Age group (years)								
Male	55–64	79.0 (25.6)	70.0 (41.9)	75.6 (28.3)	63.1 (26.9)	63.4 (25.0)	84.4 (26.1)	86.3 (31.5)	78.3 (19.5)
	65–74	76.2 (26.4)	74.4 (38.6)	78.8 (26.0)	64.8 (24.2)	65.0 (25.0)	86.0 (25.1)	89.5 (28.0)	81.4 (18.0)
	75–84	65.4 (27.5)	68.7 (41.1)	75.7 (26.8)	61.3 (26.7)	57.7 (26.5)	77.0 (28.0)	82.9 (35.3)	77.1 (19.4)
Female	55–64	77.2 (27.1)	75.2 (39.3)	73.2 (28.9)	67.4 (26.1)	60.5 (24.3)	85.0 (25.2)	77.7 (38.7)	74.0 (20.5)
	65–74	70.0 (26.8)	71.2 (40.2)	72.9 (27.8)	65.1 (24.1)	61.5 (23.1)	85.7 (26.1)	87.0 (31.1)	78.8 (18.5)
	75–84	52.9 (28.4)	57.9 (42.4)	66.7 (30.7)	58.7 (23.0)	51.9 (23.7)	75.2 (29.9)	82.7 (34.4)	74.6 (19.6)

BP, bodily pain; GH, general health; MH, mental health; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; VT, vitality.

respectively, suggesting that physical health is more affected than mental health when compared (by observation only) with the UK norms for people of these age ranges (see *Table 68*). Other physical health scores had similarly wide ranges, from 20 to 58 on the role physical and from 47 to 79 on the bodily pain scales. Other mental health scores of social functioning and role emotional ranged from 33 to 81 and from 31 to 100, respectively. The general health and vitality scales of the SF-36 ranged from 24 to 53 and 27 to 53, respectively. Scores vary widely between the studies; however, IPF of 'moderate' severity does appear to negatively affect HRQoL, as measured by the SF-36, when compared observationally with age-standardised population norms for similarly aged people in the UK.

Two studies included populations or subgroups that had a mean FVC per cent predicted of < 55%, reflecting severe IPF.[112,113] Data were from a subgroup of participants with baseline FVC per cent predicted of 51% from one study.[113] In the second study the mean FVC per cent predicted was 44%. In this latter study no measures of variance were provided for the mean scores reported. As can be seen in *Table 67*, scores on the physical functioning scale of the SF-36 ranged from 16 to 45 across the two cohorts, and scores on the mental health scale ranged between 35 and 50. On the other physical health scores, the ranges were 20–43 for role physical but higher (62–66) for bodily pain. On the other mental health domains of social functioning and role emotional, the ranges seen in these three cohorts were 30–58 and 19–65, respectively. The scores for general health and vitality ranged from 20 to 28 and from 19 to 38, respectively. From these patterns it can be seen that 'severe' IPF has a detrimental effect on most aspects of QoL as measured by the SF-36 (see *Table 68*).

Across all of the included studies there is wide variation in scores, which makes overall interpretation difficult. This is likely to be a reflection of the differences in study design, participants and sample sizes of the studies, as discussed above.

St George's Hospital Respiratory Questionnaire

Fourteen studies assessed HRQoL using the SGRQ.[55,68,72,73,117–120,122–127] IPF severities at baseline were similar, ranging from 59% to 77%. While the five RCTs[55,68,72,73,125] reported baseline data for two treatment arms, comparisons between groups were not required for the purpose of this systematic review and, as such, the placebo group data have been summarised here (the treated group population data can be seen in *Appendices 5* and *11*).

Scores on the different subscales of the SGRQ varied across the studies, and may reflect, in part, the differences between the studies in terms of the country, sample size and participant characteristics. In common with the review of the SF-36, we have crudely grouped the studies into those with mean baseline FVC per cent predicted 71–83% (to reflect a 'mild' IPF patient group) and those with a mean baseline FVC per cent predicted of 55–70% (to reflect a 'moderate' IPF patient group). There were no studies with a mean baseline FVC per cent predicted of < 55% (therefore reflecting a 'severe' IPF patient group). Interpretation of the ranges of results should be treated with caution.

Three studies[68,119,122] included a population that had a baseline FVC (per cent predicted) of > 71% ('mild' IPF) and results can be seen in *Table 69*. In these studies the mean score on the symptoms scale was between 40 and 46 and the score on the activity scale between 45 and 53. There appeared to be less effect of IPF on HRQoL as measured by the impacts scale, where the mean score ranged from 29 to 32. Overall, the total score on the SGRQ for this population was 36–39. This appears to be higher than the suggested population norms for the SGRQ as seen in *Table 70*, and suggests that IPF at a 'mild' severity has an impact on HRQoL as measured by the SGRQ.

Nine studies included populations that had baseline FVC per cent predicted between 55% and 70%, and may represent IPF of 'moderate' severity.[55,72,73,118,120,123–126] Two additional studies[117,127] did not report the baseline FVC per cent predicted but are also described here as the descriptions of the populations in the journal paper suggest that the individuals are unlikely to be mild IPF. Scores on the SGRQ ranged from 38 to 68 for symptoms, 36 to 82 for activity, 30 to 65 for impact and 38 to 64 for the total score. This suggests that the effect on HRQoL may be greater in populations with more 'moderate' IPF than in those with 'mild' IPF, and worse again than the suggested population norms for this measure (see *Table 70*).

Summary of health-related quality of life review

- The systematic review found 23 relevant studies.
- Two studies were identified that used the preferred measure of HRQoL, the EQ-5D, 16 used the SF-36 and 14 the SGRQ.
- Types of studies were mixed; some RCTs were included of which the present review focused on the placebo groups at baseline, other designs were before-and-after studies, single-cohort studies, cohort studies with matched controls, and cohort studies with a longitudinal analysis of data. Although not formally critically appraised, these studies have potential for bias and results should, therefore, be treated with caution.
- Results were mixed but appear to show that IPF has an adverse effect on HRQoL compared with population norms, and this is likely be diminish as IPF become more severe.
- Results of EQ-5D from the placebo arms of two included studies from an international research group for IPF suggest that in a 'mild' population the utility is 0.8 and in a more 'severe' population the utility diminishes to 0.74.
- Few studies were undertaken with UK participants and therefore generalisability of these data to a UK population is likely to be low.

TABLE 69 St George's Hospital Respiratory Questionnaire scores at baseline, mean (SD) unless stated

Study	N	FVC per cent predicted	Symptoms	Activity	Impacts	Total
'Mild' IPF						
Nishiyama et al. 2012[122]	87	75.0	45.0 (23.3)	48.0 (24.7)	31.6 (20.7)	39.0 (20.2)
Raghu et al. IPFCRN 2012[68]	78	72.1	45.6 (21.8)	52.7 (21.0)	28.8 (17.3)	39.4 (17.4)
Nishiyama et al. 2005[119] Placebo group	41	76.6	40.1 (24.6)	44.5 (26.7)	28.9 (19.8)	35.7 (20.6)
'Moderate' IPF						
Verma et al. 2011[123]	137	61.7	59.8 (95% CI 56.2 to 63.4)	81.6 (95% CI 78.7 to 84.4)	54.1 (95% CI 50.6 to 57.6)	63.4 (95% CI 60.4 to 66.3)
Zisman et al. IPFCRN 2010[72]	91	58.73	53.99 (18.9)	68.02 (17.63)	39.77 (18.81)	51.72 (15.86)
Raghu et al. 2010[125] Placebo group	158	67.7	54.8[a] (95% CI 47.9 to 59.1)	59.5[a] (95% CI 56.0 to 65.6)	30.5[a] (95% CI 25.7 to 38.4)	44.3[a] (95% CI 40.2 to 50.0)
Swigiris et al. 2010[126]	158	66.97	50.1 (21.9)	60.6 (22.8)	33.7 (20.6)	44.8 (19.5)
Horton et al. 2012[55]	23	70.4	67.7 (19.7)	64.3 (22.7)	48.1 (20.7)	57.4 (18.8)
Lutogniewska et al. 2010[127]	30	NR	58.7 (20.3)	70.4 (20.1)	65.4 (24.8)	64.0 (19.0)
Zimmermann et al. 2007[124]	20	70.4	46.4 (20.3)	62.4 (19)	43.6 (20.9)	48.4 (17.9)
Nishiyama et al. 2008[73]	28	67.4	38.0 (25.8)	50.4 (26.2)	29.9 (23.7)	37.8 (22.7)
Peng et al. 2008[118]	68	66	65 (16)	56 (15)	49 (19)	54 (15)
Tzanakis et al. 2005[120]	25	68.8	55.9 (25.3)	36.2 (21.4)	29.6 (21)	37.7 (18.9)
De Vries et al. 2000[117]	10	NR	56.0 (20.2)	38.7 (22.7)	46.1 (20.8)	44.5 (17.8)

NR, not reported.
a Median.

TABLE 70 Means for SGRQ scores in normal subjects with no history of respiratory disease (modified from the SGRQ manual[136])

Characteristics	Symptoms	Activity	Impacts	Total
N = 74	12 (95% CI 9 to 15)	9 (95% CI 7 to 12)	2 (95% CI 1 to 3)	6 (95% CI 5 to 7)
Age: 46 (range 17–80) years				
FEV$_1$ per cent predicted: 95 (91–99)				

Independent economic evaluation

Overview of economic evaluation

A comparison was made of the costs and benefits of pharmacological interventions for the treatment of IPF within the framework of a decision-analytic economic model. The cost-effectiveness of six interventions against BSC was assessed: azathioprine and prednisolone, BIBF 1120, NAC triple therapy, inhaled NAC, pirfenidone, and sildenafil. Non-pharmacological interventions were not included, for reasons discussed in *Chapter 3* (see *Network meta-analysis*). Thalidomide was not included as it is not currently used as a disease-modifying intervention. In the context of the model the terms placebo and BSC are considered to be equivalent, although BSC is preferred because it is more descriptive of treatment actually received.

The patient population included in the economic model reflects to some extent the patient populations of the key trials included in the systematic review of clinical effectiveness (see *Chapter 3*) as treatment effects are taken from these trials.[62,65–72] Transition probabilities for IPF progression are taken from the Capacity trial[62] where average age at baseline in the pooled population was 66 years, and 73% were men. This is broadly comparable with the UK IPF population, which between 1990 and 2004 had a mean age at presentation of 71 years, and was 62% male.[11] All-cause mortality in the model is defined using these UK data. Mean overall FVC per cent predicted at baseline in the Capacity trial was 74.7% (see *Table 5*).

The analysis takes the perspective of the NHS and Personal Social Services in the UK. The lifelong costs and benefits associated with each treatment are estimated by the economic model. The base price year for costs is 2012. Future costs and benefits are discounted at 3.5% per annum as recommended by the UK Treasury.[138] The interventions act to slow the rate of decline in FVC per cent predicted. The estimates of intervention effectiveness used by the model for reduced rate of decline in FVC per cent predicted are obtained from the NMA reported in *Chapter 3*.

Methods for economic analysis

Model structure

A Markov state-transition model was developed to simulate the progression of IPF in a cohort of patients and to estimate the cost-effectiveness of the pharmacological treatments under consideration. The model was constructed using the TreeAge Pro 2013 software (TreeAge Software Inc., Williamstown, MA, USA). The model structure was informed by a review of the available literature and expert opinion on the clinical progression of the disease. The decision to use a Markov cohort structure in preference to a patient-level model was pragmatic and based on available data. While a patient-level model is more flexible in reproducing patient experience, it requires more data and/or modelling assumptions. These data were not available and it was felt that such a model would, therefore, be less robust than a Markov model.

The model uses four distinct health states: unprogressed IPF, progressed IPF, lung-transplant and dead (*Figure 13*). Non-dead health states are associated with a HRQoL utility and a cost estimate. Progression is defined by an absolute decline in FVC per cent predicted of ≥ 10% from a baseline (recently diagnosed) value. Acute exacerbations are not modelled as separate health states but are associated with a cost and utility decrement.

All patients start the model in the unprogressed state and may then either stay in the unprogressed state; suffer an acute exacerbation and move to the progressed state; move to the progressed state without acute exacerbation; or die. From the progressed state a patient may either stay in the progressed state; suffer an acute exacerbation and return to the progressed state; undergo a lung transplant; or die. From the lung transplant state patients may either remain in the lung transplant state or move to the dead state (see *Figure 13*).
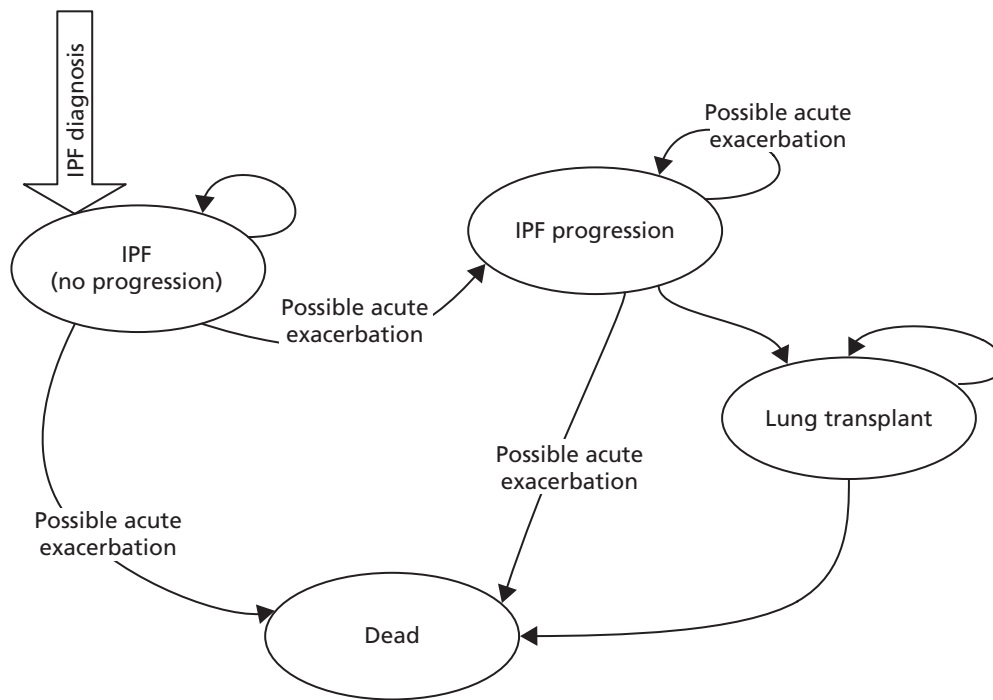
FIGURE 13 Influence diagram for the Southampton Health Technology Assessments Centre (SHTAC) IPF cost-effectiveness model.

Model cycle length is 1 month and a lifetime horizon of 30 years is adopted in the base case, which is sufficiently long to capture all clinically and economically important events. A half-cycle correction was applied.

The baseline disease progression parameters used in the model were obtained from the IPF clinical literature (Section 3). These inform the monthly probability of an absolute decline of 10% or more in FVC per cent predicted; the monthly probabilities of death from IPF (unprogressed and progressed states); and the monthly probabilities of an acute exacerbation (unprogressed and progressed states). Other sources were used to inform the probability of a lung transplant; survival after lung transplant; and all-cause mortality by age.

Parametric curves were fitted to Kaplan–Meier PFS data for treatment with BSC in order to provide the probability of disease progression in this treatment arm. The outputs of the NMA of FVC described in *Chapter 3* were used to modify this probability for each treatment and so provide a treatment-specific probability of progression. Overall survival curves were fitted to the progressed state and lung transplant Kaplan–Meier survival data in order to provide the probability of death in each cycle for these health states.

The costs included in the model are those for drug treatment; long-term oxygen use; oxygen monitoring; acute exacerbation of IPF; lung transplant; and adverse events associated with treatment. Full details of the costs used in the model are given in *Data sources*.

The model includes the following assumptions:

- All patients enter the model in the unprogressed state.
- Patients who experience a ≥ 10% absolute decline in FVC per cent predicted are considered to be in the progressed health state.
- Acute exacerbations are associated with a drop in FVC per cent predicted of ≥ 10% in an unprogressed patient.

- The utility decrement associated with acute exacerbation lasts for 1 month.
- No lung transplants occur in patients who are in the unprogressed disease state. It is only possible to transit to the lung transplant state from the progressed disease state.
- Pharmacological treatment has a constant effect on relative rate of FVC per cent predicted decline compared with BSC in the unprogressed state, irrespective of the time for which the treatment is taken or the FVC achieved.
- Treatments are assumed to act solely on the rate of FVC decline and do not have an additive effect on the incidence of acute exacerbation.
- Home oxygen is not used in the unprogressed state but all progressed IPF patients use home oxygen on a long-term basis.
- Treatment-emergent adverse events are not modelled with separate health states. They are instead reflected in additional costs associated with the relevant treatment.

It is further assumed that a per cent predicted FVC of ≥ 70% in an IPF patient population at baseline indicates unprogressed IPF. The FVC per cent predicted at baseline of recently diagnosed IPF in the key trials is generally > 70% (see *Chapter 3*) and this is, consequently, an appropriate, if somewhat arbitrary, threshold for distinguishing between progressed and unprogressed IPF when information on prior rate of predicted FVC per cent decline is not available. In common with the key trials, we do not impose an upper limit on FVC per cent predicted when defining unprogressed IPF.

An absolute decline of ≥ 10% in FVC per cent predicted is an established marker of disease progression and mortality in patients with IPF.[67,139,140] FVC or VC decline has been used as a marker of disease progression in many RCTs among IPF populations[62,141,142] and guidelines recommend that an absolute decrease in FVC of ≥ 10% can be used as a surrogate marker for mortality.[140] Recent work has indicated that a relative decline in FVC may be similarly prognostic of mortality.[139] However, given the available evidence, this economic evaluation uses absolute decline in FVC per cent predicted of ≥ 10% as its marker of disease progression.

In each cycle, the total costs and QALYs are calculated by multiplying the individual costs and HRQoL of each model state by the proportion of the model cohort in that state for each of the treatments. The total discounted lifetime costs and QALYs are calculated by aggregating the costs and QALYs for all cycles. The ICER of each of the treatments versus BSC is calculated as

$$\text{ICER} = \frac{\text{Cost of treatment} - \text{cost of BSC}}{\text{QALYs of treatment} - \text{QALYs of BSC}} \qquad (6)$$

The associated net monetary benefit (NMB) of a specific treatment compared with BSC may be calculated as

$$\text{NMB} = \text{incremental QALYs} \times \text{WTP} - \text{incremental costs}, \qquad (7)$$

where the incremental QALYs and incremental costs are simply the denominator and numerator, respectively, of *Equation 6* and WTP is the maximum amount a decision-maker is prepared to pay per QALY gained.[143] As long as the NMB is more than zero, a treatment is cost-effective, and larger NMBs represent greater cost-effectiveness than smaller NMBs. The NMB framework facilitates comparison of multiple treatment options simultaneously and it is used in later analysis to identify the most cost-effective treatment at various assumed costs for BIBF 1120.

Model validation

The model was validated by checking the model structure, calculations and data inputs for correctness. The structure was reviewed by clinical experts to establish that it was appropriate for the disease and its treatment. Internal consistency was examined by varying input values and subsequent verification that any change to the input values produced changes in the model outputs of the expected direction and magnitude. To establish its external consistency, the model results were compared with outcomes reported in trial and other publications.

Evaluation of uncertainty

The evaluation of the cost-effectiveness of pharmacological treatments for IPF is based on uncertain information which includes uncertainty about the clinical effects of treatment, HRQoL while in the various health states, and resource use. Such uncertainty is examined using deterministic and probabilistic sensitivity analyses (PSAs).

One-way deterministic sensitivity analyses were conducted to test the robustness of the cost-effectiveness results to variations in parameter input values when altered one at a time (see *Results of independent economic analysis*).

Joint variation and potential correlation in multiple parameters was addressed using PSA (see *Results of independent economic analysis*). In the PSA, probability distributions were assigned to the parameter point estimates used in the base-case analysis. The model was then run for 1000 iterations with parameter values sampled at random from these distributions. The uncertainty surrounding the cost-effectiveness of the treatments is represented on a cost-effectiveness acceptability curve (CEAC), which plots the probability that an intervention will be cost-effective at a particular WTP threshold.

Scenario analysis was used to investigate the effect of uncertainty in model assumptions and structure.

Value of information

Decision-makers are interested in the value of further research to gain more precise information on uncertain parameters in an economic model, and need to identify the research priorities where additional information would provide the most benefit. Value of information (VOI) analysis attempts to answer these questions by analysing the hypothetical case for which perfect information could be obtained through further research.[144] The expected value of perfect information (EVPI) shows the value of reducing the uncertainty around the decision of whether or not to adopt the intervention. If the EVPI is greater than the cost of obtaining new information, then further research will be valuable. The EVPI is estimated for different cost-effectiveness thresholds and this gives an indication of an upper cost bound for further research at these thresholds.

Expected value of perfect information analysis alone does not indicate what further information to obtain. In order to determine what further research would have the most effect in reducing uncertainty, a partial EVPI is conducted. In this analysis, the uncertainty around particular input parameters in the model is investigated.

Value of information analysis was conducted to investigate the expected pay-off of further research into selected treatments for IPF and is described in *Results of independent economic analysis*.

Data sources

Unprogressed state: probability of progression

The baseline risk of disease progression in the economic model is taken from the Capacity trials (see *Chapter 3*).[62] These were large RCTs with 347 patients in the pooled placebo population who were followed up for 72 weeks (16.5 months). Mean per cent predicted FVC at baseline in the pooled placebo population was approximately 75%, and so the population meets the model definition of unprogressed

IPF. PFS was defined as a composite of time to either confirmed ≥ 10% decline in per cent predicted FVC; ≥ 15% decline in per cent predicted DL_{CO}; or death. Kaplan–Meier curves for this outcome are given in the trial publication.[62] Numbers experiencing a confirmed ≥ 10% decline in per cent predicted FVC were also reported as a secondary endpoint. The high numbers enrolled in the trial and the detailed reporting of outcomes which broadly coincide with the model definition of IPF progression make this an appropriate and robust source of data to inform disease progression in the economic model.

The placebo population in the Capacity trial is considered to correspond to the BSC treatment arm in the economic model. PFS probabilities in the pooled placebo population were extracted from a Kaplan–Meier plot in the trial publication using the digitising software Engauge (Engauge Digitizer, San Francisco, CA, USA: http://digitizer.sourceforge.net) and the method of survival curve reconstruction described in Guyot and colleagues.[145] Parametric survival curves were then fitted to the observed data using Stata software (StataCorp LP, College Station, TX, USA) in order to extrapolate PFS beyond the 72 weeks in the trial. In line with the recommendation of Latimer,[146] all of the 'standard' parametric models were considered (exponential, Weibull, Gompertz, log-logistic and log-normal).

Akaike information criterion (AIC) values obtained for each distribution are given in *Table 71* which shows that the log-normal and Weibull distributions provide the best fit to the data based on this criterion. The Gompertz and exponential distributions fit the data less well. The log-normal and Weibull fits are compared graphically in *Figure 14*. The figure demonstrates that while the initial behaviour of the log-normal and Weibull fits is similar, the log-normal predicts better PFS than the Weibull after approximately 2 years. PFS predicted by the log-normal fit at 10 years is approximately 6% and declines only slowly after this. The Weibull fit is, therefore, considered to be more clinically plausible and is preferred in the model base case. The log-normal fit is examined in scenario analysis discussed in *Results of independent economic analysis*.

Progression-free survival, as reported in the Capacity trials, is for model purposes confounded with decline in DL_{CO}. The numbers experiencing only a categorical change in FVC ≥ 10% in the pooled population from both Capacity trials is reported in Noble and colleagues[62] but no Kaplan–Meier curve is supplied for this outcome. The TTP outcome also does not incorporate deaths, which are included in PFS and are required by the model. PFS data are consequently adopted for use in the model.

TABLE 71 Values of AIC obtained for parametric survival models fitted to reconstructed PFS data from Capacity trial[62]

Model	AIC
Log-normal	640.4
Weibull	641.6
Log-logistic	641.7
Gompertz	648.9
Exponential	669.5

Lower values of AIC indicate a better fit to the data.
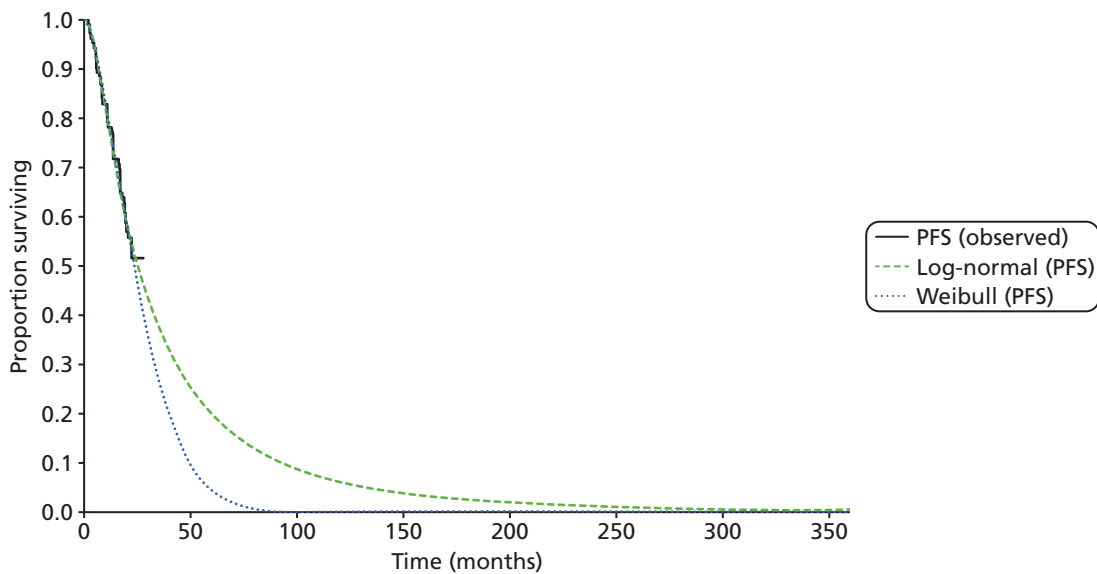
FIGURE 14 Kaplan–Meier plot of PFS in the pooled placebo population of the Capacity trials[62] compared with fitted log-normal and Weibull PFS curves.

Unprogressed state: probability of death from idiopathic pulmonary fibrosis

Given that the model uses a PFS curve to inform disease progression in the base case, all deaths in the unprogressed state are incorporated in the monthly probability of progression discussed above (there are no deaths in those remaining in the unprogressed state). The conditional probability that a patient has died, given that they are moving out of the PFS state, is, consequently, required. The Capacity trial reports 34 deaths from any cause in the pooled placebo population.[62] The total number of patients moving out of the PFS state is not reported in the trial publication but from the survival curve reconstruction described above we estimate that there were 138 such events. Assuming that all reported deaths in this trial occurred to patients in the unprogressed state gives a probability of death, given progression of 34/138, of 0.25 (see *Table 74*). Variations to this proportion are examined in sensitivity analysis.

Progressed state: overall survival

The monthly probability of death in the progressed state is obtained from the overall survival Kaplan–Meier curve given in Richeldi and colleagues[139] for IPF patients exhibiting a ≥ 10% absolute decline in FVC per cent predicted at 12 months after diagnosis. This decline is consistent with the definition used by the economic model of progression. The study is more recent and follows a bigger sample (*n* = 26) than other published studies which differentiate survival by FVC decline[147,148] and, consequently, it was adopted to inform the economic model.

Overall survival probabilities were extracted from the Kaplan–Meier curve using the digitising software Engauge and the method of survival curve reconstruction described in Guyot and colleagues.[145] Parametric survival curves were then fitted to the observed data using Stata software to allow for more straightforward examination of uncertainty in transition probabilities in the economic model, and in order to extrapolate overall survival beyond the 5 years of observed data. All of the 'standard' parametric models were considered (exponential, Weibull, Gompertz, log-logistic and log-normal).[146]

The exponential distribution provided the best fit to the data based on the AIC (*Table 72*). This fit represents the data reasonably well and is the distribution adopted in the economic model base case (*Figure 15*). The alternative functional forms of Weibull and Gompertz provide similar fits to the observed Kaplan–Meier curve as the exponential and so were not examined in sensitivity analysis (see *Figure 15*).

The exponential fit to the overall survival data gives the monthly probability of death from IPF in the progressed disease state as 0.0304 (see *Table 74*).

TABLE 72 Values of AIC obtained for parametric survival models fitted to reconstructed overall survival data from Richeldi *et al.*[139]

Model	AIC
Exponential	83.1
Gompertz	84.8
Weibull	85.0
Log-normal	86.8
Log-logistic	87.2

Lower values of AIC indicate a better fit to the data.
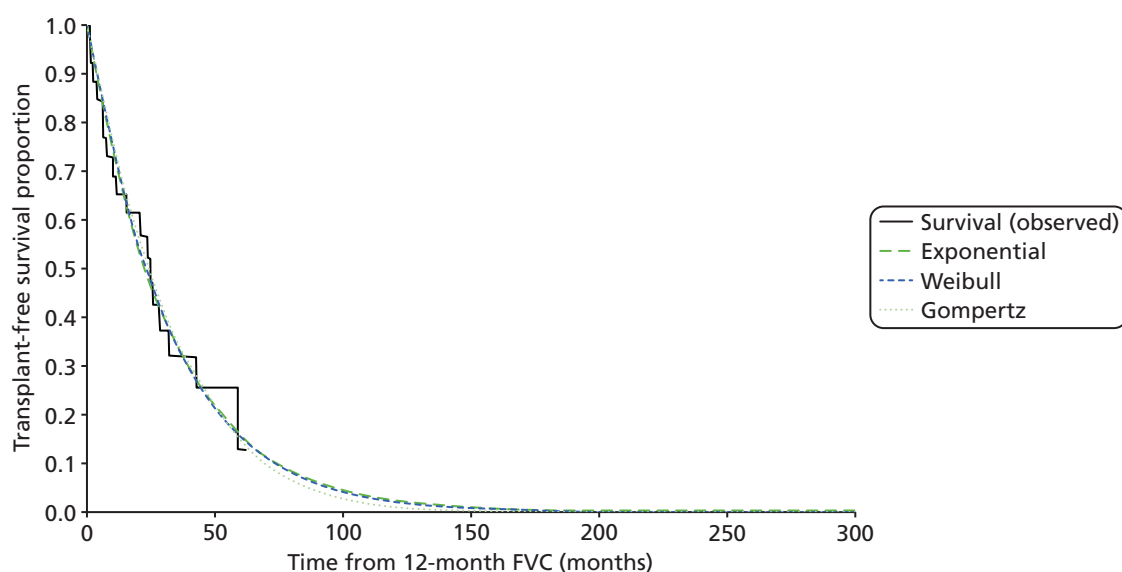
FIGURE 15 Overall survival curves fitted to data from Richeldi *et al.*[139] for IPF patients with ≥ 10% absolute decline in FVC per cent predicted at 12 months.

The data in *Figure 15* represent patients who have experienced a ≥ 10% absolute decline in FVC per cent predicted and in this respect conform to the economic model definition of IPF progression. However, they also pertain to a particular rate of decline as they were obtained for patients with ≥ 10% decline at 1 year after diagnosis. The data may consequently overstate the probability of death for those in the progressed state who have experienced a ≥ 10% decline in FVC per cent but did so over a longer time period than 1 year. Alternative, less steep overall survival curves for those in the progressed state are examined in deterministic sensitivity analysis described in *Results of independent economic analysis*.

Acute exacerbation

The baseline probability of an acute exacerbation while in the unprogressed state is taken from a study by Song and colleagues.[17] This was a retrospective review of 461 patients in South Korea who were diagnosed with IPF according to the ATS/ERS criteria between 1990 and 2009. Two hundred and sixty-nine of the cases were biopsy-proven. Acute exacerbation was defined by the criteria of Collard and colleagues.[6] FVC per cent predicted was > 70% at baseline and so the patient group meets the model definition of unprogressed IPF. Three-year incidence of first acute exacerbation was found to be 20.7% and corresponding 1-year incidence was 14.2%. This compares with the incidence of 10–14% reported in *Chapter 1* (see *Description of underlying health problem*). The 3-year incidence of 20.7% is converted to a monthly probability of 0.006 for use in the economic model (see *Table 74*).[149] The 1-month probability of death from acute exacerbation is also taken from Song and colleagues and is 0.5 (see *Table 74*).

Given the model assumption that acute exacerbations entail a drop in FVC per cent predicted of ≥ 10% in an unprogressed patient, all patients undergoing acute exacerbation move out of the unprogressed state and are included in the monthly probability of progression discussed above. In order to reflect the costs and utility decrement associated with acute exacerbation, the model calculates the probabilities that a progressing patient has either survived or died from an acute exacerbation using Bayes' theorem:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)},$$
(7)

where, for the probability that a progressing patient has a survived an acute exacerbation, $p(B)$ is the 1-month probability of moving to the progressed state; $p(A)$ is the 1-month probability of an acute exacerbation in the unprogressed state; $p(B|A)$ is the 1-month probability of surviving an acute exacerbation; and $p(A|B)$ is the probability that a progressing patient has survived an acute exacerbation (see *Table 74* for all input probabilities). The probability that a progressing patient dies from an acute exacerbation is calculated using the same method.

A transition cost and utility decrement are applied to those undergoing an acute exacerbation before they transit to their destination state, i.e. progressed or dead (see *Figure 13*). The utility decrement is assumed to last for 1 month only.

The probability of acute exacerbation while in the progressed state is drawn from the placebo arm of the sildenafil trial.[72] FVC per cent predicted was < 60% at baseline in this trial, thus meeting the model definition of progressed disease. Four of 91 patients experienced acute exacerbation on the placebo arm in the 12-week follow-up period,[72] giving a monthly probability of acute exacerbation of 0.016, more than twice the monthly probability of acute exacerbation for the unprogressed health state (see *Table 74*).

The overall survival curve used to inform transitions from the progressed state already incorporates deaths due to acute exacerbation. The proportion of these deaths due to acute exacerbation is calculated using Bayes' theorem with the method described above. The proportion of acute exacerbations in those remaining in the progressed state is calculated in the same way. A transition cost and 1-month utility decrement are then applied to those undergoing an acute exacerbation before they transit to their destination state, i.e. progressed or dead (see *Figure 13*).

Lung transplant frequency in progressed idiopathic pulmonary fibrosis patients

Specific criteria have been proposed to guide referral of IPF patients for lung transplant (see *Chapter 1*, *Current service provision*). These include a reduction of 10% in FVC over a 6-month follow-up period. The NHS Choices website currently states that a lung transplant is used to treat people with advanced lung disease who are failing to respond to other treatment.[150] Notwithstanding the fact that clinicians may make a quick referral for a deteriorating IPF patient even if they do not meet the established criteria (see *Chapter 1*, *Current service provision*), it is assumed in the economic model that only patients with progressed IPF are eligible for a lung transplant.

The UK Cardiothoracic Transplant Audit notes 166 lung transplants in the UK in 2010–11.[151] Using the IPF incidence and prevalence in the UK given by Navaratnam and colleagues,[13] and outputs from the economic model, leads to an estimate for the size of the progressed IPF population in the UK of approximately 4505. Given that 17% of lung transplants are for IPF patients (see *Chapter 1*, *Current service provision*), the overall annual probability of lung transplant for a progressed IPF patient in the UK is 0.0062, or 0.6%, equivalent to 28 transplants per year (see *Table 74*). This probability does not reflect comorbidities or other factors which may make progressed patients ineligible for transplant and reduce the effective population size to < 4505, and so will understate the probability of transplant achieved in practice for a patient on the waiting list. It is, however, a necessary assumption in the economic model where patients in the progressed state are not differentiated by comorbidity or other status.

The figure of 28 lung transplants per year is in line with the range of 23–33 given in *Chapter 1* (see *Current service provision*). Variations to the annual probability of lung transplant are examined in a sensitivity analysis.

Survival after lung transplant

Figure 17 in the 28th Annual Report of the Registry of the ISHLT gives a Kaplan–Meier survival curve for IPF patients after lung transplant between 1990 and 2008.[151] The standard parametric survival curves were fitted to a digitised copy of this curve in order to inform survival after lung transplant in the economic model.[146] The digitising software Engauge was used, followed by the method of survival curve reconstruction described in Guyot and colleagues.[145]

The Weibull and log-normal distributions were found to provide the best fit to the data by the AIC (*Table 73*) and are compared with the observed Kaplan–Meier curve in *Figure 16*. *Figure 16* shows that 5-year survival after lung transplant is 40–50% in all cases, in line with the survival estimate given in *Chapter 1* (see *Current service provision*).

TABLE 73 Values of AIC obtained for parametric survival models fitted to reconstructed lung transplant overall survival data[152]

Model	AIC
Weibull	22,257
Log-normal	22,262
Gompertz	22,391
Exponential	22,506
Log-logistic	22,527

Lower values of AIC indicate a better fit to the data.
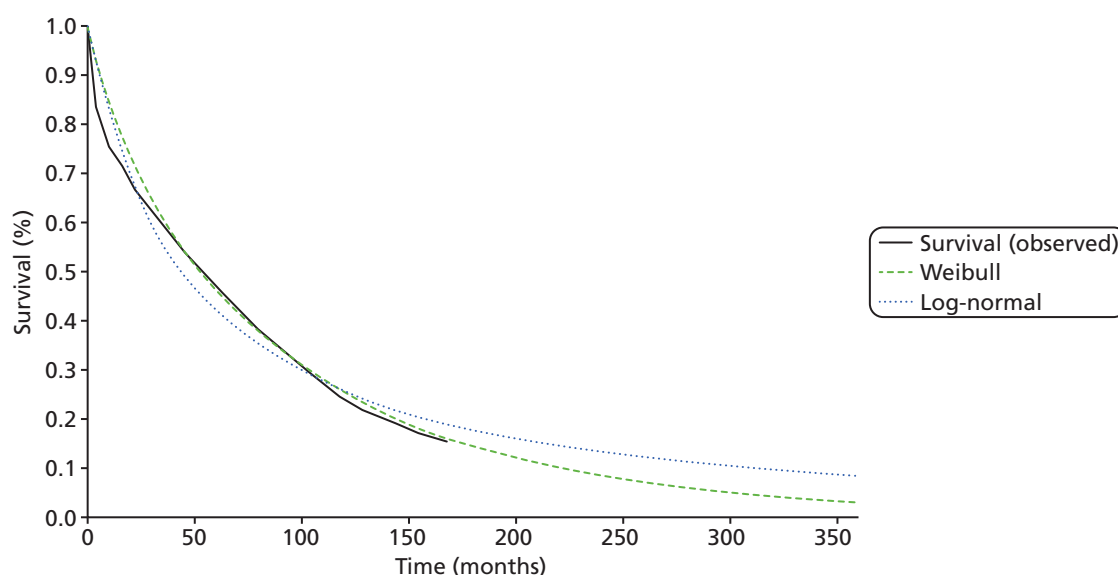
FIGURE 16 Observed Kaplan–Meier survival by IPF diagnosis for adult lung transplants performed between January 1990 and June 2008,[152] compared with fitted Weibull and log-normal survival curves.

Lung transplant is associated with high hazard of death soon after transplant, which reduces as time progresses.[151] Both the Weibull and log-normal curves overestimate lung transplant survival until around 30 months after transplant (see *Figure 16*). Because of this poor initial fit to the data, it was decided to use the observed Kaplan–Meier probability of mortality in the first 30 months after transplant. After this time the model uses the fitted Weibull survival curve to estimate post-lung transplant mortality as it provides a slightly better fit to the observed data than the log-normal, and also provides a more clinically plausible extrapolation of survival, i.e. with a progressive decline in survival after around 170 months, rather than a levelling-off as seen in the log-normal curve (see *Figure 16*).

The lung transplant state is modelled as a tunnel state because the hazard of death after lung transplant varies over time and it is not the starting state – new patients enter the state throughout the model time horizon.

A summary of baseline transition probabilities used in the economic model is given in *Table 74*.

TABLE 74 Summary of baseline probabilities

Variable	Source point estimate (time period)	Point estimate per 1-month model cycle	Source
Unprogressed state – probability of progression or death	NA – KM data	Not constant – Weibull curve	Noble *et al.*[62]
Unprogressed state – probability of death from IPF given progression	0.25	0.25	Noble *et al.*[62] and survival curve reconstruction
Unprogressed state – probability of acute exacerbation	0.207 (3 years)	0.006	Song *et al.*[17]
Progressed state – probability of acute exacerbation	0.044 (12 weeks)	0.016	IPFCRN Zisman *et al.*[72]
1 month probability of death from acute exacerbation	0.5	0.5	Song *et al.*[17]
Progressed state – probability of death from IPF	NA – KM data	0.0304	Richeldi *et al.*[139]
Progressed state – probability of lung transplant	0.0062 (1 year)	0.0005	RCS[151] Navaratnam *et al.*[13] 17% (NHS Blood and Transplant, 25 February 2013, personal communication)
Lung transplant state – probability of death	NA – KM data	Not constant – Weibull curve	Christie *et al.*[152]

KM, Kaplan–Meier; NA, not applicable; RCS, Royal College of Surgeons.

All-cause mortality associated with age

The general underlying risk of mortality was modelled using a cohort life table generated from the 2009–11 male and female interim life tables for England and Wales.[153] The age-related mortality for each year in the model was determined from these data using the demographic characteristics of IPF patients in the UK reported in Gribbin and colleagues.[11] Specifically, in the base case, patients enter the model at an age of 71 years, and a 62%/38% male/female split is assumed.

Treatment effects

Absolute rate of decline in forced vital capacity

The effects of the various pharmacological treatments on slowing down (or speeding up) the rate of decline in FVC per cent predicted are derived from the NMA described in *Chapter 3* (see *Network meta-analysis*). It is assumed in the model that the odds ratio obtained from the standardised mean difference in FVC seen in the key trials can be applied to an absolute decline in FVC per cent predicted of 10%, even when a 10% absolute decline in FVC per cent predicted was not seen in a trial itself. This is equivalent to assuming that a pharmacological treatment has a constant effect on relative rate of FVC per cent decline compared with BSC in the unprogressed state, irrespective of the time for which the treatment is taken or the FVC achieved.

The odds ratios for the treatment effects given in *Table 59* are used in the economic model. The odds ratios shown in this table assume that the standardised mean difference data are normally distributed,[93] and this results in slightly smaller treatment effects and narrower CrIs than the alternative parameterisation (see *Table 57*). The treatment effects are applied to the baseline probability of disease progression given by the curve fitted to PFS data for BSC (see *Figure 14*). The baseline probability is converted to odds before it is multiplied by the odds ratio, as follows:

$$odds_i = \frac{p}{1-p} * OR_i,$$ (8)

where $odds_i$ is the odds of progression for treatment i, p is the baseline monthly probability of progression in the BSC arm, and OR_i is the treatment effect of the ith treatment.

The odds of progression for treatment i are converted back to a monthly probability using the following formula:

$$p_i = \frac{odds_i}{1 + odds_i},$$ (9)

where p_i is the monthly probability of progression for treatment i.

In the base case, for all treatments except sildenafil, treatment effects are applied only in the unprogressed disease state. For sildenafil, the effect of treatment is applied to the progressed disease state only. This conforms to the study populations of the relevant trials. The application of treatment effects in both unprogressed and progressed disease states is examined in scenario analysis described in *Results of independent economic analysis*.

Treatments are assumed to act solely on the rate of FVC decline and do not have a separate additive effect on the incidence of acute exacerbation. Any benefits in reduced acute exacerbation incidence associated with a pharmacological treatment are considered to be reflected in the change in rate of FVC decline associated with the treatment.

Health-related quality of life

The economic model applies different utility values to each of the non-dead health states. These are not differentiated by treatment received. Any impact of treatment on HRQoL is assumed to occur because of delay to disease progression, and this delay is already accounted for in the model. The utility values used for the health states are discussed below.

Unprogressed and progressed idiopathic pulmonary fibrosis

The systematic review of clinical effectiveness and HRQoL (see *Systematic review of health-related quality of life studies*) identified two studies which used the EQ-5D to establish health state values in IPF patients.[68,72] The sildenafil trial[72] included a population of severe IPF patients with a baseline mean FVC per cent predicted of 59% for the placebo arm. This group meets the model definition for progressed IPF. The placebo population of the triple-therapy NAC trial[68] had a baseline mean FVC per cent predicted of 72% on the placebo arm, and so meets the model definition of unprogressed IPF.

Both studies were conducted under the auspices of the IPFCRN, with all study centres in North America. Neither study protocol mentions the tariff used in calculation of EQ-5D health state values but it may be assumed to be the US tariff. Huang and colleagues[129] note that for assessing treatment benefit in a single population, the use of either the UK or the US EQ-5D weights as a measure of HRQoL will not change inferences.

Given that both the sildenafil and the NAC studies were conducted by the same clinical network, it is likely that their estimates of EQ-5D are consistent and, consequently, it is reasonable to contrast them in the economic model. The model adopts the self-report questionnaire EQ-5D value of 0.8 ($\pm$ 0.2) for the placebo population at baseline, given in Raghu and colleagues,[68] as the estimate of utility in the unprogressed IPF health state. The self-report questionnaire EQ-5D value of 0.74 ($\pm$ 0.19) for the placebo population at baseline, given in Zisman and colleagues,[72] is used as the estimate of utility in the progressed IPF health state (*Table 75*).

Other studies in the systematic review of HRQoL (see *Systematic review of health-related quality of life studies*) use the SF-36 to measure HRQoL. Mapping the SF-36 scores reported in these studies to EQ-5D using the algorithm of Ara and Brazier[156] (their Equation 1) was investigated. However, for the single study which reported both SF-36 and EQ-5D,[72] the mapped EQ-5D value of 0.46 was very low compared with the directly elicited EQ-5D value of 0.74. Longworth and Rowen[157] recommend that mapping of utilities should usually be viewed as a second best solution to direct collection of data. For this reason, a decision was made to use directly collected EQ-5D in the economic model.

TABLE 75 European Quality of Life-5 Dimensions utility values by model health state

Model health state	EQ-5D (SD)	Source
Unprogressed IPF	0.80 (0.20)	IPFCRN Raghu *et al.*[68]
Progressed IPF	0.74 (0.19)	IPFCRN Zisman *et al.*[72]
Lung transplant		
0–6 months after transplant	0.71 (0.38)	Anyanwu *et al.*[154] (weighted average assuming two-thirds single and one-third double lung)
7–18 months after transplant	0.72 (0.31)	Anyanwu *et al.*[154] (weighted average assuming two-thirds single and one-third double lung)
19–36 months after transplant	0.70 (0.33)	Anyanwu *et al.*[154] (weighted average assuming two-thirds single and one-third double lung)
> 36 months after transplant	0.68 (0.38)	Anyanwu *et al.*[154] (weighted average assuming two-thirds single and one-third double lung)
Acute exacerbation decrement	0.20 (not available)	Lloyd *et al.*[155]

Lung transplant

The utility associated with the lung transplant state is taken from a UK-based study by Anyanwu and colleagues,[154] which assessed the HRQoL in lung transplantation using the EQ-5D. This study finds that utility after transplant tends to increase through time in cases of a bilateral transplant (within the follow-up period), but decreases through time for a single lung transplant. Although, since April 2006 the number of bilateral sequential lung grafts has increased to 72.3% of total lung transplant activity,[151] IPF is a disease of older adults and, traditionally, single lung transplant has been preferred over double lung transplant because it is not such a major operation (see *Chapter 1*, *Current service provision*). A double lung transplant may be performed in younger IPF patients. Given this mix of transplant types, we use a weighted average of the single and double transplant utilities reported in Anyanwu and colleagues[154] in the economic model. We assume that two-thirds of IPF patients receive a single lung transplant, and that one-third receive a double lung transplant. The weighted average utilities are given in *Table 75* by time after transplant.

The weighted utilities after lung transplant initially increase from 0.71 to 0.72, then decrease to 0.68 after 36 months. These utilities are, in all cases, lower than the utility value assumed for progressed IPF of 0.74 (see *Table 75*). These differences may arise because of the relative age of the lung transplant utility data which were collected in 1998,[154] in contrast to the more recent progressed state utility value for which data were collected between 2007 and 2009.[72] It is worth noting that the EQ-5D value for the intervention population at baseline, reported in Zisman and colleagues,[72] is 0.71, lower than the value of 0.74 obtained for the placebo population which is used in the economic model. We examine variations to these base-case utilities in deterministic and PSAs described in *Results of independent economic analysis*.

Acute exacerbations

No studies were found which directly examined EQ-5D in acute exacerbations of IPF. HRQoL for patients experiencing acute exacerbation is obtained from a study by Lloyd and colleagues,[155] which concerned hospitalised asthma patients experiencing an acute exacerbation. The sample size with hospitalisations in this study was only five. However, the mean EQ-5D decrement for these five, of 0.2, is in agreement with the mean decrement of 0.19 found by Menn and colleagues[158] for acute exacerbations in COPD using a German EQ-5D tariff. In the absence of better data, this decrement has been used in the economic model (see *Table 75*). Variations to the decrement are examined in sensitivity analysis (see *Results of independent economic analysis*).

Summary

A summary of the health state utility values used in the economic model is given in *Table 75*. To avoid double counting, the pharmacological treatments for IPF are assumed to not have any effect on HRQoL beyond that achieved by delaying the progress of the disease.

Costs

Five types of cost are considered in the economic model:

1. Costs associated with a particular pharmacological treatment. These include both the costs of the drug and the monitoring costs associated with the treatment.
2. Hospital admission costs arising from acute exacerbations of IPF.
3. Ongoing non-pharmacological treatment costs for management of the condition. These include oxygen costs and the costs of long-term oxygen monitoring.
4. Costs associated with lung transplant.
5. Cost of adverse events attributed to the pharmacological intervention.

Treatment costs

Drug unit costs were taken, where available, from the NHS electronic drugs tariff for May 2013.[159] Otherwise, costs were taken from *British National Formulary* (BNF) 65[160] or other appropriate online sources (*Table 76*). Dose information was obtained from either BNF 65 or the NICE clinical guideline for IPF,[31] except for BIBF 1120 for which doses were drawn from the trial publication.[71]

TABLE 76 Drug unit costs, doses and monitoring costs used in the economic model

Drug	Standard dose	Unit cost of intervention	Drug cost per month	Monitoring costs	Monitoring cost total per month
Azathioprine	2 mg/kg 125 mg/day assumed[31]	Cost per 25 mg 28-tab pack = £4.71 Cost per 50 mg 56-tab pack = £4.20 Cost per day = £0.32[159]	£9.69	13 liver function tests, seven full blood counts and one TPMT assay assumed per year at total annual cost of £279.96[31]	£23.33
BIBF 1120	150 mg two times daily[71]	Currently unlicensed. Use assumed cost of pirfenidone (month 2 onwards) + 50%	£3273.71	Not known – none applied	–
NAC	600 mg three times daily[31]	Only available as an unlicensed generic. Cheapest cost found for 600 mg 250-tab pack = £18.47 (supplied by Now® Foods, price on www.amazon.co.uk, 17 May 2013) Cost per day = £0.22	£6.75	No additional monitoring required	–
Inhaled NAC	352.4 mg two times daily, diluted with saline to a total volume of 4 ml[69]	Acetylcysteine cost per 200 mg/ml 10 ml ampoule = £2.25 (as Parvolex®, Phoenix) Sodium chloride nebuliser solution cost per 20 × 2.5 ml = £11.50 Cost per day = £1.94 assuming no wastage[160] Nebuliser cost = £79.78 (one-off)[a] Face mask cost = £0.59 (one per month assumed)[a]	£59.14	No additional monitoring required	
Pirfenidone	267 mg three times daily for first week; 534 mg three times daily for second week; 801 mg three times daily ongoing	Cost per 267 mg 63-tab pack = £501.92 Cost per 267 mg 252-tab pack = £2,007.70 Cost per 267 mg 270-tab pack = £2,151.10 Cost per week For week 1 = £167.31 For week 2 = £334.62 For weeks 3+ = £501.92[160]	Month 1 = £1680.55 Month 2+ = £2182.47	Eight liver function tests per 48 weeks[23] Assumed number of tests per year = 9 Cost of liver function test including nurse time = £13.12[31] Cost of tests per year = £118.08	£9.84

TABLE 76 Drug unit costs, doses and monitoring costs used in the economic model (*continued*)

Drug	Standard dose	Unit cost of intervention	Drug cost per month	Monitoring costs	Monitoring cost total per month
Prednisolone	40 mg daily for first 4 weeks	Cost per 5mg 28-tab pack = £3.43	For month 1 = £29.23	Assessment for corticosteroid complications. Annual cost of £220.11[31]	£18.34
	30 mg daily for weeks 5–8	Cost per week	For month 2 = £21.18		
		For weeks 1–4 = £6.86			
	20 mg daily for weeks 9–12	For weeks 4–8 = £5.15	For month 3 = £13.28		
	10 mg daily thereafter[31]	For weeks 8–12 = £3.43	For month 4+ = £8.11		
		For weeks 12+ = £1.86			
Sildenafil	By mouth, 20 mg three times daily	Cost per 20 mg 90-tab pack = £373.50[160]	£378.95	No additional monitoring required	
		Cost per day = £12.45			

a University Hospital Southampton NHS Foundation Trust Procurement Team, 17 July 2013, personal communication.

The costs of monitoring associated with treatment were also included in the economic model and are given in *Table 76*, together with full details of assumed doses, unit costs and data sources.

BIBF 1120 was, at the time of writing (June 2013), unlicensed and a price is not available from standard sources. The manufacturer (Boehringer Ingelheim) was contacted with a request for price information. Their response indicated that the price is currently commercially sensitive and therefore unavailable. Consequently, the model uses an assumed cost for BIBF 1120. This is fixed at the monthly cost of pirfenidone plus 50% in the base case but is subject to full sensitivity analysis (see *Results of independent economic analysis*).

Acute exacerbations

The cost of an acute exacerbation was estimated as the activity-weighted average cost of the DZ25A and DZ25B currency codes given in NHS Reference Costs 2011–12.[161] These codes relate to fibrosis or pneumoconiosis either with or without complications and comorbidities. Elective inpatient data were excluded from calculations of the weighted average of DZ25A and DZ25B as they do not represent sudden unplanned deterioration in IPF. The calculated unit cost and lower and upper quartiles used in sensitivity analysis are given in *Table 77*. These general costs were used for all treatments, as no treatment-specific acute exacerbation costs were available.

TABLE 77 Unit costs assumed by economic model

HRG/code	Description	Unit cost (£)	Lower quartile (£)	Upper quartile (£)	SD (£)
DZ25A DZ25B	Fibrosis or pneumoconiosis, with and without CC (excluding elective inpatient data)	1361.04	1002.20	1578.20	–
DZ38Z	Oxygen assessment and monitoring	173.94	121.40	209.37	–
–	Home oxygen – 12 months' supply	824.30	–	–	142.80
DZ01Z	Lung transplant (excluding outpatient procedures)	35,468.61	28,334.00	36,356.36	–

CC, complications and comorbidities; HRG, Healthcare Resource Group.

Oxygen and oxygen monitoring costs

The economic model assumes that home oxygen is not used in the unprogressed state, but that all progressed IPF patients use home oxygen on a long-term basis. This is associated with a monitoring cost and the cost of the oxygen itself.

The cost of oxygen assessment and monitoring is taken from currency code DZ38Z given in NHS Reference Costs 2011–12.[161] A weighted average of items under currency code DZ38Z was calculated and is given in *Table 77*, together with the lower and upper quartiles used in sensitivity analysis. A frequency of two oxygen-monitoring assessments per year is assumed.

The cost of oxygen was taken from the home oxygen service costing tool made available by the UK Department of Health.[162] This tool estimates the likely cost to a primary care trust (PCT) of providing home oxygen to patients by year over the period 2012–16, and the additional cost of running a Home Oxygen Assessment and Review Service to ensure that home oxygen is appropriately prescribed to those people who clinically need it.[163] Using this tool, the baseline cost of oxygen in 2012 was extracted for each of the 151 PCTs it covers. An expected oxygen cost for one patient was obtained using the average proportional incidence of oxygen subcharge categories within the PCT given by the costing tool. In this way, the average annual cost to a PCT of home oxygen was calculated as £824.30 per patient, with a SD of £142.80 (see *Table 77*).

Lung transplant

The cost of a lung transplant was calculated as a weighted average of items under currency code DZ01Z in NHS Reference Costs 2011–12, excluding outpatient procedures.[161] Outpatient procedures have an average cost of £140 and do not relate to transplant per se, although they may be a consequence of transplant. A weighted average unit cost for lung transplant of £35,468.61 was obtained (see *Table 77*).

Adverse events

The incidence of adverse events for each pharmacological treatment was taken from the studies included in the systematic review of clinical-effectiveness (see *Chapter 3*). Although adverse event data are consistently reported across studies as a percentage of patients, the types of adverse event reported differ between the studies. Only serious adverse events of Common Terminology Criteria (CTC) grades 3 and 4 which occur in > 5% of patients in any treatment arm are included in the model, as these are considered to incur additional NHS costs. Adverse events are, moreover, included only if the adverse event incidence differs significantly between treatment arms, in line with the modelling guidelines of Philips and colleagues.[60] Adverse event costs are generally small in comparison with other costs in the model as serious events are relatively rare.

The adverse events which required cost estimates were gastrointestinal and respiratory system related. A mean cost for treatment of each adverse event was obtained from NHS Reference Costs[161] using the currency codes shown in *Table 78*. When incorporating costs for adverse events into the model, we used the expected cost per patient shown in *Table 78* and assumed that patients would have only one episode of any adverse event during their treatment. This cost is applied in the first cycle of the model.

Model validation

The Weibull and log-normal PFS predictions from the model are compared with the observed Kaplan–Meier data for the BSC arm[62] in *Figure 17*. The predictions look very reasonable when compared with the observed data and the original fitted curves shown in *Figure 14*. The model thus appears to be adequate in its simulation of PFS in the BSC arm.

TABLE 78 Mean cost estimates for AEs included in the economic model

Treatment	AE	Currency code	Cost (£) per event	AE incidence (% patients)	Expected cost per patient (£) per event
BIBF 1120[71]	Serious GI	PA25A[a]	3387	4.7	159
	Severe GI	PA26A[b]	1559	5.9	92
Triple NAC[68]	Respiratory system	DZ25A[c]	1150	16	184

AE, adverse event; GI, gastrointestinal.
a Major gastrointestinal disorders with complications and comorbidities.
b Other gastrointestinal disorders with complications and comorbidities.
c Fibrosis or pneumoconiosis, without complications and comorbidities.
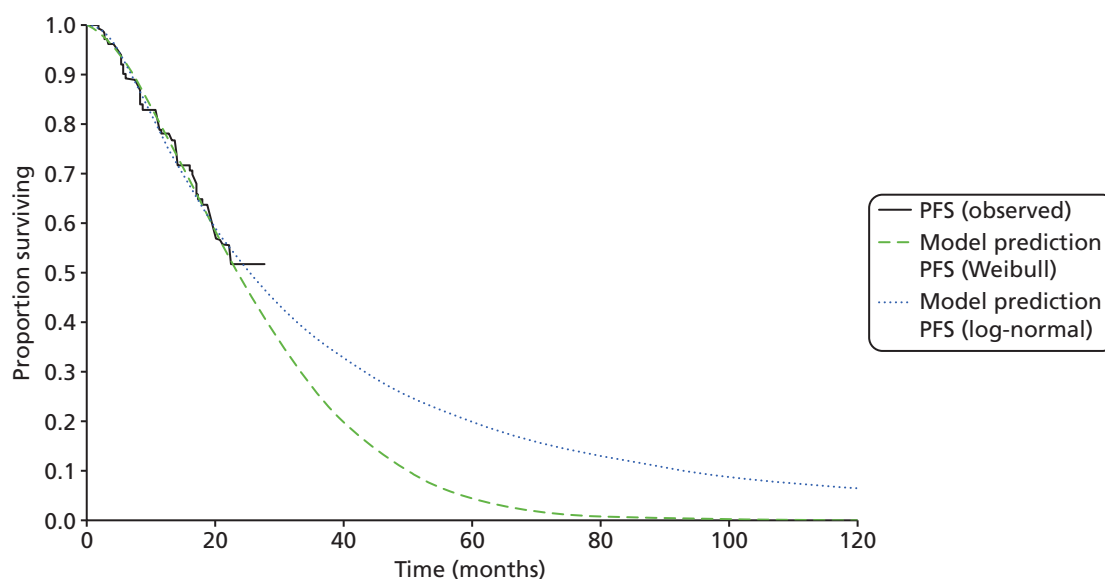
FIGURE 17 Progression-free survival predicted by model compared with PFS observed in the Capacity trial (pooled placebo arm).

It was verified that the treatment effects were being correctly applied in the unprogressed state by producing PFS curves for both pirfenidone and BIBF 1120 and comparing these with predicted PFS in the BSC arm (*Figure 18*). Pirfenidone is associated with longer median PFS than BSC, and BIBF 1120 is associated with longer median PFS than either pirfenidone or BSC. This is consistent with the treatment effects estimated by the NMA (see *Chapter 3, Network meta-analysis*).

Overall survival curves for each treatment arm were examined and used to obtain median overall survival predictions which are shown in *Table 79* along with median PFS by treatment arm. The model predicts median overall survival of 43 months (3.6 years) in the BSC arm which is consistent with the median survival of 2–5 years given in Noble and colleagues[62] and the median survival of 3.03 years estimated by Navaratnam and colleagues.[13]
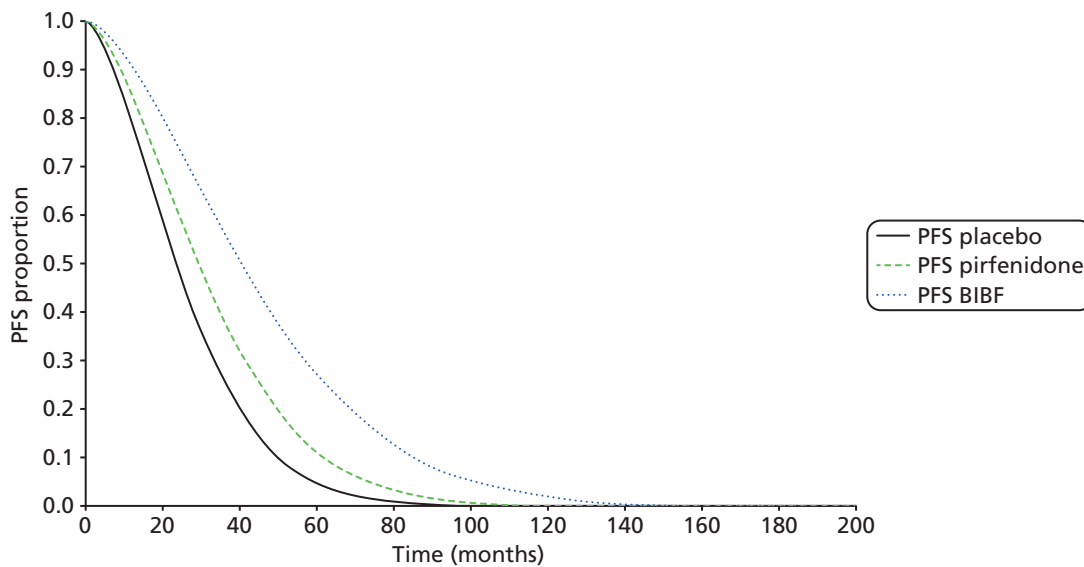
FIGURE 18 Progression-free survival predicted by model, by treatment arm.

TABLE 79 Median PFS and overall survival predicted by model, by treatment arm

Treatment arm	Median PFS (months)	Median OS (months)	Difference (OS-PFS)
BSC	23.5	43	19.5
Pirfenidone	29	50	21
BIBF 1120	40	61.5	21.5
OS, overall survival.			

Table 79 also indicates that the difference between median PFS and median overall survival in the BSC arm is 19.5 months, or 1.6 years. Median survival after a > 10% decline in absolute FVC per cent predicted is given in the source Richeldi and colleagues data[139] as approximately 24.5 months (see *Figure 15*). These conditional data may not be recreated exactly from the economic model as the model does not distinguish overall survival for the progressed state from overall survival overall. After accounting for this dissimilarity, a reduction of 5 months in the difference of medians appears of satisfactory magnitude and varies in the expected direction given the mortality associated with a ≥ 10% decline in FVC per cent predicted. In summary, the model appears to be performing adequately when compared with external data sources.

Results of independent economic analysis

This section reports the cost-effectiveness of the pharmacological treatments compared with BSC in a cohort of IPF patients.

Base-case results for all treatments are given in *Table 80* and are shown on the cost-effectiveness plane in *Figure 19*. The cost-effectiveness frontier, which indicates the maximum health gain which may be attained for any given level of spending on the available treatment options, is also shown in *Figure 19*. The only three treatments on the cost-effectiveness frontier are BSC, inhaled NAC and BIBF 1120. *Table 80* gives the total discounted costs for all treatments together with the discounted QALYs expected for patients in the cohort, by treatment arm. The associated ICERs comparing each treatment with both BSC and the next best option (the next cheapest option on the cost-effectiveness frontier) are also shown.

Table 80 and *Figure 19* indicate that azathioprine and prednisolone is dominated by BSC as it is more costly but gives rise to fewer total QALYs than BSC. NAC triple therapy is associated with an ICER of £41,811 per QALY when compared with BSC but does not lie on the cost-effectiveness frontier because of extended dominance by a blend of BSC and inhaled NAC. Inhaled NAC is associated with an ICER of £5037 per QALY when compared with the next best option (BSC). Sildenafil and pirfenidone are dominated by inhaled NAC and are not cost-effective at a WTP of £30,000 per QALY when compared with BSC. BIBF 1120 is the most costly treatment (assumed) but is also associated with the greatest number of QALYs. It has an ICER of £132,659 per QALY when compared with BSC and an ICER of £209,251 when compared with the next best option (inhaled NAC).

TABLE 80 Summary of base-case cost-effectiveness results for all treatments

Treatment	Total costs (£)	Total QALYs	ICER vs. BSC (£/QALY)	ICER vs. next best option (£/QALY)
BSC	3084	2.98	–	–
Azathioprine and prednisolone	4313	2.66	Dominated	Dominated
NAC triple therapy	5021	3.03	41,811	Extended dominance
Inhaled NAC	5029	3.37	5037	5037
Sildenafil	12,008	3.11	68,116	Dominated
Pirfenidone	70,118	3.34	190,146	Dominated
BIBF	139,613	4.01	132,658	209,246

BIBF uses an assumed cost.
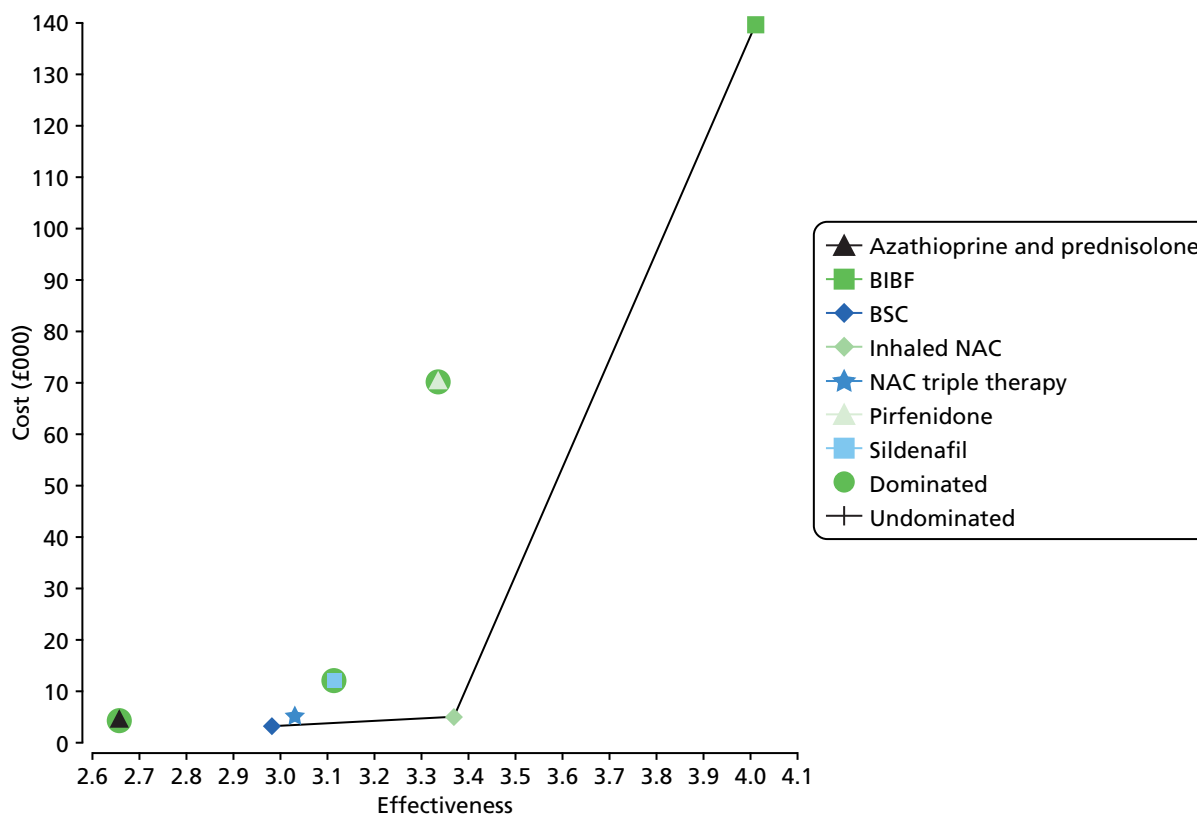
FIGURE 19 Cost-effectiveness plane and frontier for all treatments.

Although *Table 80* and *Figure 19* indicate that inhaled NAC is the only cost-effective treatment compared with BSC at a WTP threshold of £30,000 per QALY, the treatment effect of inhaled NAC does not achieve statistical significance at the 95% level in the NMA (see *Table 59*), and so it is unlikely that this treatment would be adopted in clinical practice. The clinical effectiveness review also notes that this study had an unclear risk of bias (see *Table 6*). The treatments which achieve statistical significance in the NMA are pirfenidone and BIBF 1120. Results for only these treatments are shown on the cost-effectiveness plane in *Figure 20*.

It is clear from *Table 80* that neither BIBF nor pirfenidone is a cost-effective treatment option for IPF at a WTP of £30,000 per QALY. Compared with BSC, pirfenidone has an ICER of £190,146 per QALY gained, and BIBF 1120 has an ICER of £132,658 per QALY gained. Pirfenidone is subject to extended dominance by a blend of BSC and BIBF 1120 and does not lie on the cost-effectiveness frontier (see *Figure 20*).

In summary, although these base-case findings indicate that inhaled NAC is a cost-effective treatment option for IPF based on a point estimate of its efficacy, this estimate does not achieve statistical significance. On this basis, inhaled NAC is unlikely to be adopted as a treatment option in clinical practice. No other pharmacological intervention is cost-effective at a WTP threshold of £30,000 per QALY in the model base case.
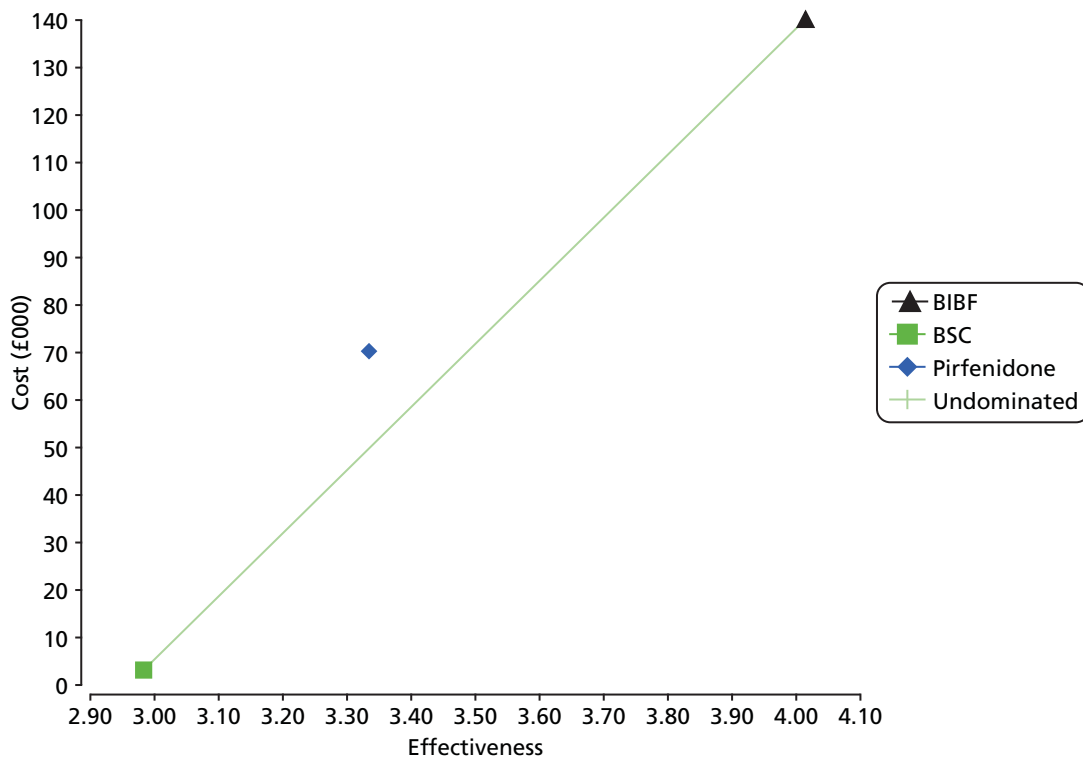
FIGURE 20 Cost-effectiveness plane and frontier for BSC, pirfenidone and BIBF 1120.

Sensitivity analysis

Deterministic and PSAs were conducted in order to investigate the effect of uncertainty in model parameter values on the cost-effectiveness results. Deterministic sensitivity analysis was used to highlight the most influential parameters while the effect of uncertainty and interaction in multiple parameters was examined using PSA. Scenario analysis was used to investigate the effect of uncertainty in model assumptions and structure. An additional one-way sensitivity analysis was performed on the price of BIBF 1120 in order to identify a price at which it becomes a cost-effective treatment option compared with BSC at a WTP of £20,000 and £30,000.

Each parameter was assumed to follow a probability distribution and these are given, with the distribution parameters, in *Table 81*. For beta distributions, the distribution parameters were fitted using either the method of moments or information on the sample size and number of events when available. Distribution parameters were fitted to the gamma distributions using the method of moments. In cases where a standard error or SD was not supplied in the source literature, the standard error was calculated using an arbitrary ± 20% from the base-case value. Correlation between the parameters of the Weibull distribution used to inform disease progression was incorporated by sampling from a bivariate normal distribution with covariance matrix as specified in *Table 81*. Possible correlation between treatment effects obtained from the NMA was addressed by using values sampled from the converged Markov chain (see *Table 81*).

The model parameters were varied in deterministic sensitivity analysis between the 2.5th and 97.5th percentiles of the assumed parameter distribution of the mean value, and these are given in *Table 81*. *Table 82* gives upper and lower bounds for parameters examined in deterministic sensitivity analysis only.

TABLE 81 Parameters, distributions and associated upper and lower values used in probabilistic and deterministic sensitivity analysis

Name	Distribution	Distribution parameters	Mean/base case	2.5th percentile	97.5th percentile
Costs					
Acute exacerbation	Gamma	$\alpha = 10.16$	£1361	£657	£2317
		$\beta = 133.95$			
Lung transplant	Gamma	$\alpha = 35.57$	£35,469	£24,784	£48,038
		$\beta = 997.09$			
O_2 monitoring	Gamma	$\alpha = 7.11$	£174	£71	£323
		$\beta = 24.45$			
O_2	Gamma	$\alpha = 5031.30$	£824	£802	£847
		$\beta = 0.16$			
Probabilities					
Acute exacerbation unprogressed state	Beta	$\alpha = 75$	0.207	0.167	0.250
		$\beta = 287$			
Acute exacerbation progressed state	Beta	$\alpha = 4$	0.044	0.012	0.094
		$\beta = 87$			
Death from acute exacerbation within 1 month	Beta	$\alpha = 38$	0.5	0.394	0.619
		$\beta = 37$			

continued

TABLE 81 Parameters, distributions and associated upper and lower values used in probabilistic and deterministic sensitivity analysis (*continued*)

Name	Distribution	Distribution parameters	Mean/base case	2.5th percentile	97.5th percentile
Death in unprogressed state	Beta	$\alpha = 34$	0.246	0.178	0.321
		$\beta = 104$			
Lung transplant	Beta	$\alpha = 28$	0.006	0.004	0.009
		$\beta = 4477$			
OS and PFS curve parameters					
OS progressed state	Normal[a]	SE = 0.213	–3.549	–3.966	–3.131
PFS	Bivariate normal[a]				
Lambda		Covariance matrix	–5.412	–5.668	–5.194
Shape		$\begin{pmatrix} 0.1404 & \\ -0.0287 & 0.0062 \end{pmatrix}$	0.469	0.459	0.480
Utilities					
Unprogressed state	Beta	$\alpha = 248.8$	0.80	0.754	0.842
		$\beta = 62.2$			
Progressed state	Beta	$\alpha = 358.2$	0.74	0.700	0.778
		$\beta = 125.8$			
After lung transplant					
0–6 months	Beta	$\alpha = 20.0$	0.71	0.533	0.859
		$\beta = 8.2$			
7–18 months	Beta	$\alpha = 29.4$	0.72	0.571	0.842
		$\beta = 11.6$			
19–36 months	Beta	$\alpha = 33.8$	0.70	0.568	0.822
		$\beta = 14.2$			
> 36 months	Beta	$\alpha = 41.1$	0.68	0.558	0.791
		$\beta = 19.3$			
Decrement associated with acute exacerbation[b]	Beta	$\alpha = 19.8$	0.20	0.128	0.284
		$\beta = 79.2$			
Treatment effects (FVC decline)					
Azathioprine and prednisolone	WinBUGS[c] output from NMA used to preserve any correlation in treatment effects (sample of 1000 iterations used, taken after burn-in of 50,000 iterations with thinning interval of 50)[a]		0.40	–0.27	1.08
BIBF 1120			–0.88	–1.39	–0.37
Inhaled NAC			–0.38	–1.13	0.36
NAC			–0.05	–0.55	0.44
Pirfenidone			–0.35	–0.56	–0.15
Sildenafil			–0.11	–0.59	0.37

OS, overall survival; SE, standard error.
a Parameters on log scale.
b Distribution calculated after arbitrary ± 20% variation applied to mean to obtain standard error.
c WinBUGS, MRC Biostatistics Unit, Cambridge, UK.

TABLE 82 Parameters examined in deterministic sensitivity analysis only with upper and lower values

Parameters	Base case	Lower value	Upper value
Age of cohort entering model	71	65	75
Discount rate for costs	3.5	0	6
Discount rate for health	3.5	0	6

Deterministic sensitivity analysis

Tables 83 and *84* show the results of the deterministic sensitivity analyses for pirfenidone and BIBF 1120 compared with BSC for the most influential parameters. Tornado diagrams depicting the range in ICER given in these tables are shown in *Figures 21* and *22*.

TABLE 83 Deterministic sensitivity analyses for pirfenidone vs. BSC

Variable	Low value	High value	Minimum ICER (£/QALY)	Maximum ICER (£/QALY)	Difference
Treatment effect on FVC decline – pirfenidone (log-odds ratio)	−0.56	−0.15	126,431	417,752	291,321
Discount rate utilities	0	0.06	158,838	214,124	55,286
Utility of unprogressed state	0.754	0.842	180,062	202,572	22,510
Discount rate costs	0	0.06	181,928	203,155	21,227
Progressed state OS parameter	−3.966	−3.131	186,925	196,960	10,035
Shape parameter PFS Weibull curve	0.459	0.480	188,814	191,767	2953
Lambda parameter PFS Weibull curve	−5.668	−5.194	189,200	191,566	2366
Probability of death given progression	0.178	0.321	188,951	191,245	2294
Utility in the progressed state	0.700	0.778	189,499	190,765	1266
Utility decrement associated with acute exacerbation	0.128	0.284	190,059	190,248	189
Probability of LT	0.00413	0.0087	190,065	190,243	178
Probability of acute exacerbation (unprogressed state)	0.167	0.25	190,073	190,231	159
Probability of acute exacerbation (progressed state)	0.012	0.094	190,058	190,201	143
Generic cost of acute exacerbation	657	2317	190,093	190,219	127
LT utility months 36 +	0.558	0.791	190,092	190,196	104
Age of cohort entering model	65	75	190,112	190,185	73
Cost of oxygen monitoring	71	323	190,118	190,166	48
Probability of death from acute exacerbation	0.394	0.619	190,132	190,160	29
LT utility months 19 to 36	0.568	0.822	190,135	190,157	22
LT utility months 7 to 18	0.571	0.842	190,136	190,155	19
LT utility months 0 to 6	0.533	0.859	190,140	190,152	12
Cost of LT	24,784	48,038	190,140	190,152	12
Cost of oxygen	802	847	190,144	190,148	4

LT, lung transplant; OS, overall survival.

TABLE 84 Deterministic sensitivity analyses for BIBF 1120 vs. BSC

Variable	Low value	High value	Minimum ICER (£/QALY)	Maximum ICER (£/QALY)	Difference
Treatment effect on FVC decline – BIBF (log-odds ratio)	−1.39	−0.37	96,399	272,859	176,460
Discount rate utilities	0	0.06	108,168	151,739	43,571
Discount rate costs	0	0.06	125,025	145,149	20,124
Utility of unprogressed state	0.754	0.842	125,622	141,327	15,705
Progressed state OS parameter	−3.966	−3.131	130,419	137,413	6994
Shape parameter PFS Weibull curve	0.459	0.480	131,884	133,606	1722
Probability of death given progression	0.178	0.321	131,832	133,412	1580
Lambda parameter PFS Weibull curve	−5.668	−5.194	132,087	133,554	1467
Utility in the progressed state	0.700	0.778	132,206	133,090	883
Utility decrement associated with acute exacerbation	0.128	0.284	132,599	132,726	127
Probability of LT	0.0041	0.0087	132,601	132,725	124
Generic cost of acute exacerbation	657	2,317	132,606	132,728	122
Probability of acute exacerbation (unprogressed state)	0.167	0.250	132,601	132,715	114
Probability of acute exacerbation (progressed state)	0.012	0.094	132,588	132,701	113
LT utility months 36 +	0.558	0.791	132,619	132,693	74
Age of cohort entering model	65	75	132,633	132,684	52
Cost of oxygen monitoring	71	323	132,630	132,677	47
Probability of death from acute exacerbation	0.394	0.619	132,635	132,672	37
LT utility months 19 to 36	0.568	0.822	132,650	132,665	15
LT utility months 7 to 18	0.571	0.842	132,651	132,664	14
Cost of LT	24,784	48,038	132,651	132,664	12
LT utility months 0 to 6	0.533	0.859	132,653	132,662	9
Cost of oxygen	802	847	132,656	132,660	4

LT, lung transplant; OS, overall survival.
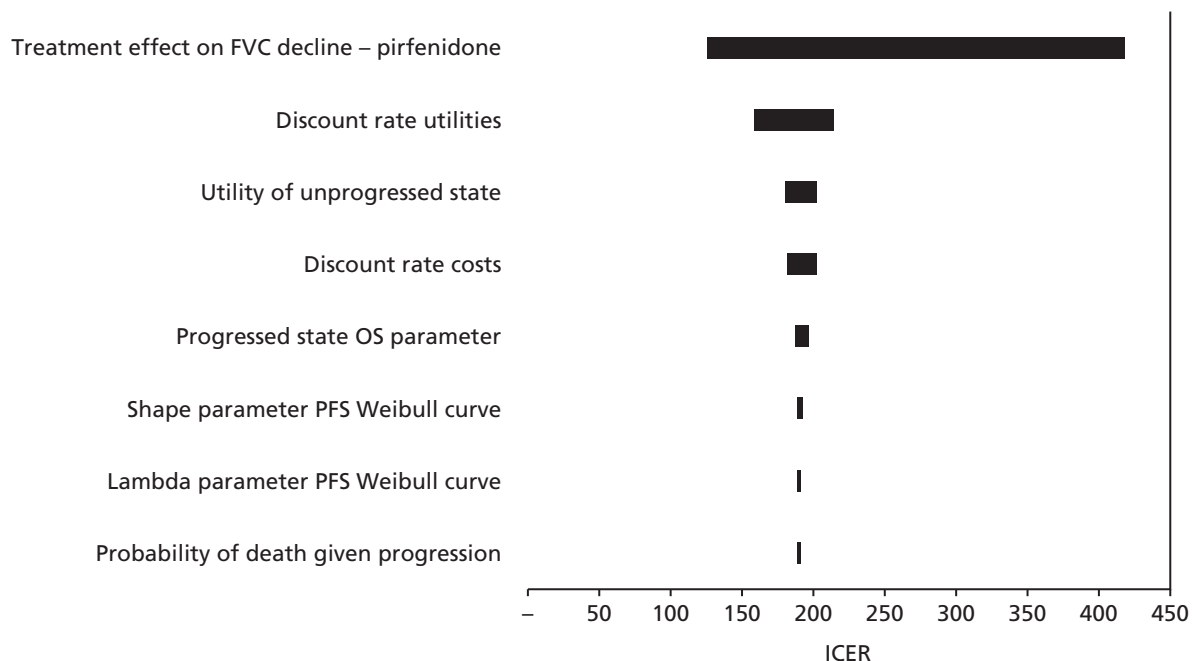
FIGURE 21 Tornado diagram showing results of deterministic sensitivity analysis for pirfenidone vs. BSC. Bars indicate spread in ICER between parameter bounds (£000). OS, overall survival.
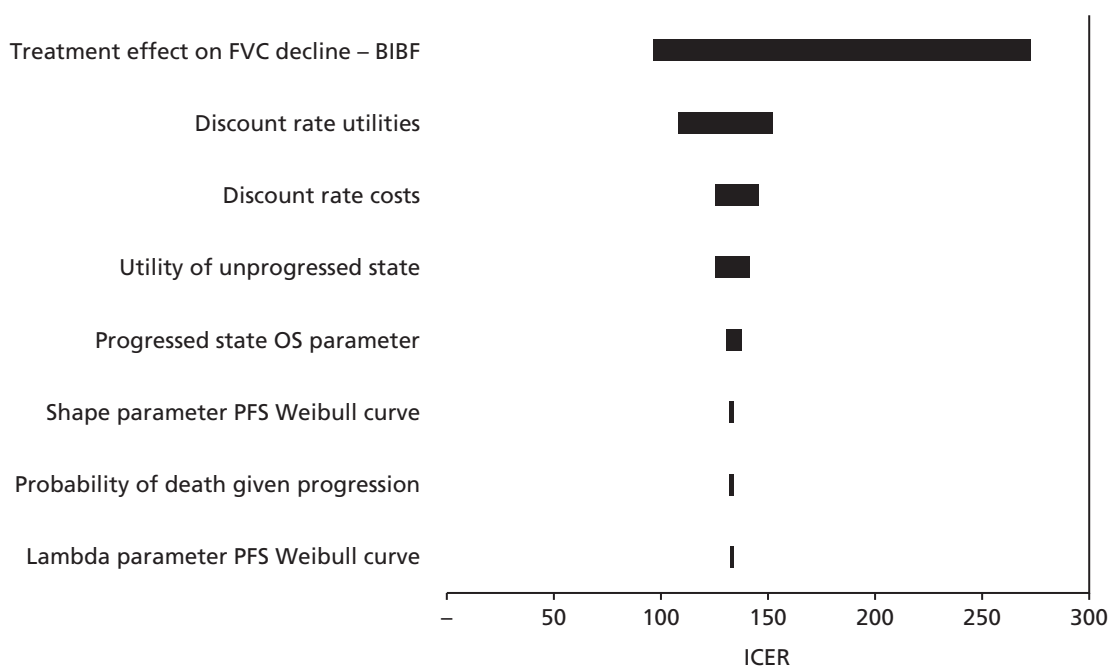
FIGURE 22 Tornado diagram showing results of deterministic sensitivity analysis for BIBF vs. BSC. Bars indicate spread in ICER between parameter bounds (£000). OS, overall survival.

The results indicate that the ICER is, above all, very sensitive to the assumed treatment effects of pirfenidone and BIBF 1120, as there is a very wide difference in the ICERs produced by the low and high values of these parameters. However, even in the best case, the ICERs for both treatments are above £90,000 per QALY when compared with BSC, at £126,431 per QALY for pirfenidone and £96,399 per QALY for BIBF 1120 (see *Tables 83* and *84*, respectively). In the worst case, the ICERs for both treatments are above £250,000 per QALY gained.

A degree of sensitivity is also shown to the discount rates for utilities and costs; to assumed utility in the unprogressed state; and to the parameter used for the overall survival curve in the progressed state. However, the differences in ICER here are all less than one-quarter of the differences associated with the treatment effect variation.

Figure 23 shows the NMB associated with BSC, pirfenidone and BIBF 1120 at a WTP of £30,000 per QALY gained. In this analysis, the cost of BIBF 1120 was varied between £0 and £5000 per month, while the cost of the other two treatment options was held constant.

The monthly cost of BIBF 1120 assumed in the model base case is £3274. *Figure 23* demonstrates that, given a WTP of £30,000 per QALY, BIBF 1120 must cost less than £736 per month to be considered as the cost-effective treatment option compared with BSC and pirfenidone. At BIBF 1120 costs higher than this, BSC treatment is preferred as it has a higher NMB. Using a lower WTP, of £20,000, the corresponding monthly cost threshold for BIBF 1120 is £489.
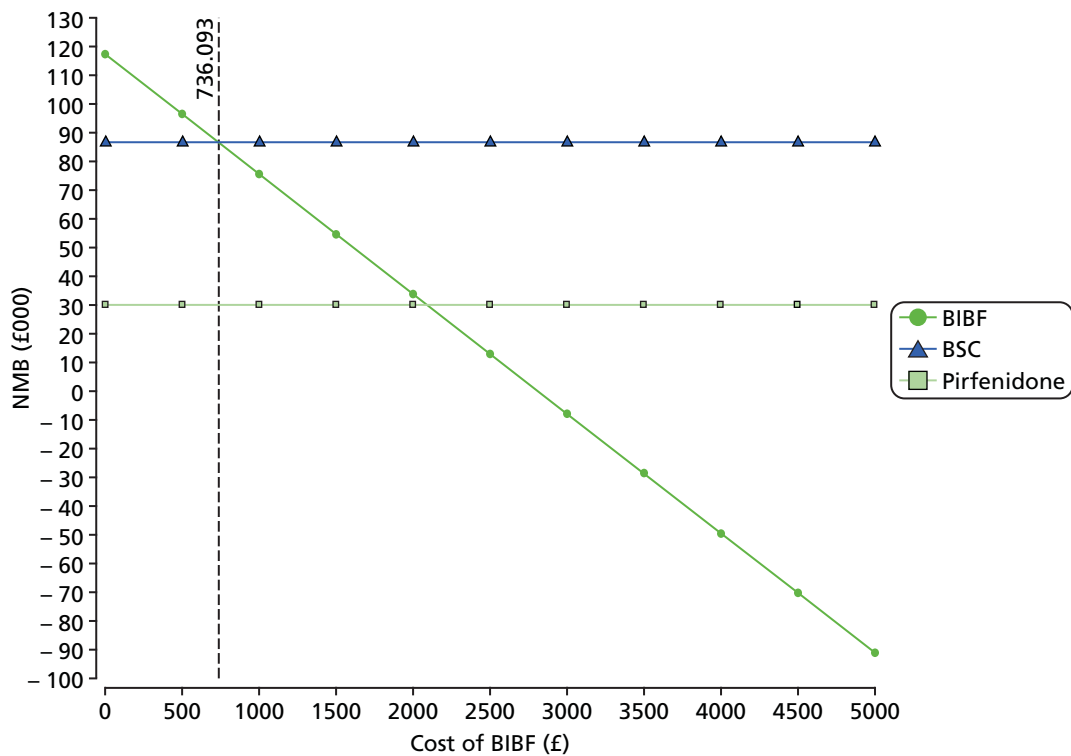
FIGURE 23 One-way sensitivity analysis of monthly cost of BIBF 1120, showing effect on BIBF 1120 NMB compared with BSC and pirfenidone.

Probabilistic sensitivity analysis

Parameters entered in PSA and their assumed distributions are given in *Table 81*. One thousand simulations were run.

The mean PSA results are presented in *Table 85* and are similar to the results for the base case given in *Table 80*. Scatterplots for cost and health outcomes are shown in *Figures 24* and *25* for pirfenidone and BIBF 1120, respectively. The CEAC for all treatments is given in *Figure 26* and indicates that at the £20,000 WTP threshold, inhaled NAC has the highest probability (65%) of being cost-effective. Inhaled NAC also has the highest probability of being cost-effective (64%) at a WTP of £30,000 per QALY. At £90,000 and £100,000 WTP thresholds, BIBF 1120, the other undominated treatment in the base-case analysis, has the probability of being cost-effective of 0.2% and 2%, respectively (see *Figure 26*). BIBF 1120 has a higher probability of being cost-effective than BSC at WTP thresholds in excess of £110,000 per QALY.

TABLE 85 Baseline PSA cost-effectiveness results vs. BSC

Treatment	Mean total costs (£)	Mean total QALYs	ICER vs. BSC (£/QALY)
BSC	3126	3.00	–
Azathioprine and prednisolone	4391	2.72	Dominated
NAC triple therapy	5086	3.07	25,147
Inhaled NAC	5074	3.40	4866
Sildenafil	12,405	3.15	58,555
Pirfenidone	70,266	3.36	184,099
BIBF	140,749	4.06	129,878
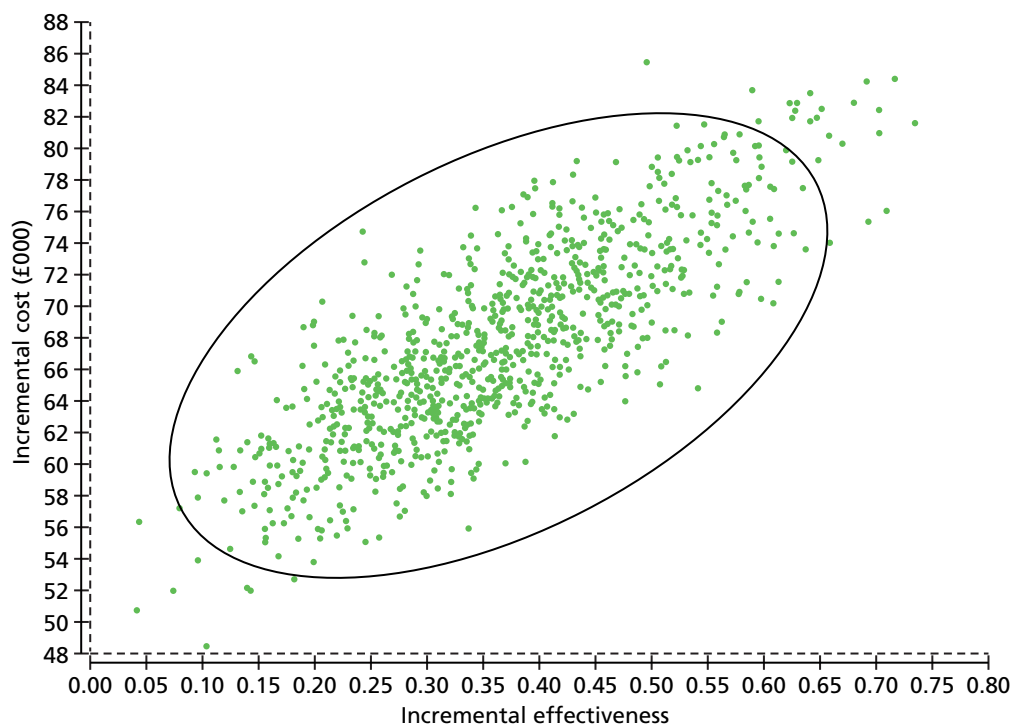
FIGURE 24 Scatterplot of the costs and health benefits from PSA: pirfenidone vs. BSC.
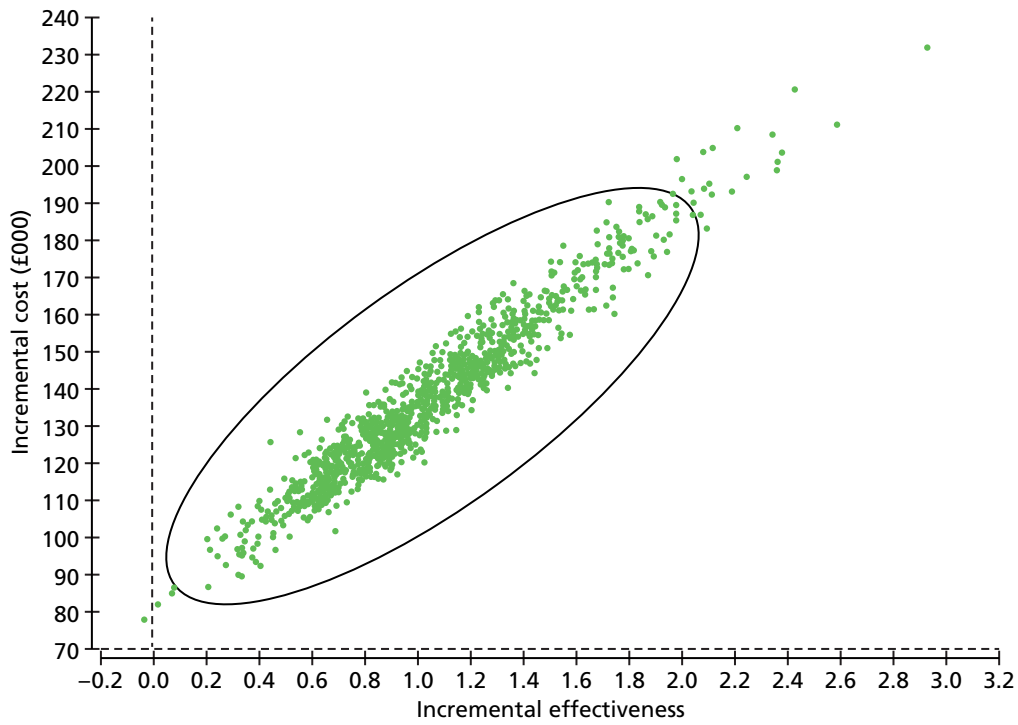
FIGURE 25 Scatterplot of the costs and health benefits from PSA: BIBF vs. BSC.
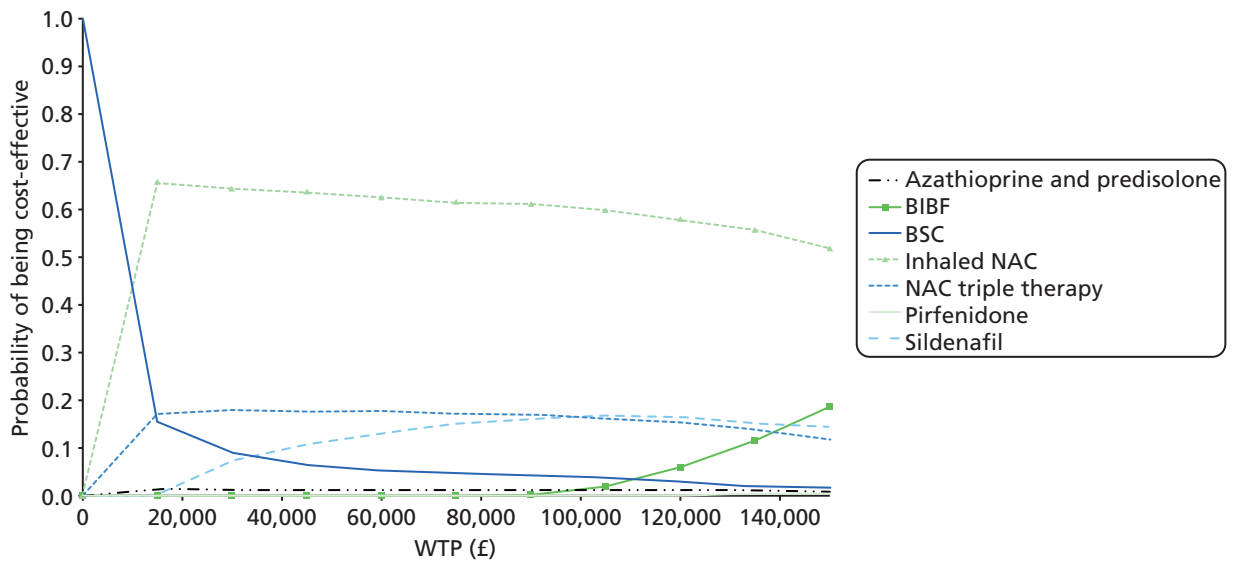
FIGURE 26 Cost-effectiveness acceptability curve from the PSA.

Scenario analysis

In addition to the sensitivity analyses, three alternative scenarios were examined to investigate the uncertainty surrounding the structural assumptions made by the model.

Exponential fit to progression-free survival data from Noble and colleagues[62]

The model base case uses a Weibull fit to PFS data. The use of a log-normal fit to these data was examined and produced the results given in *Table 86*.

Median PFS is greater with a log-normal fit than with a Weibull (see *Figure 14*) and, consequently, *Table 86* shows more QALYs accruing to each treatment arm than in the base case (see *Table 80*). Total costs are also greater than the base case, but overall ICERs compared with BSC are reduced to £121,519 per QALY gained for pirfenidone and £95,786 per QALY gained for BIBF 1120. Pirfenidone is still subject to extended dominance by a blend of BSC and BIBF 1120. Neither pharmacological treatment is cost-effective at a WTP threshold of £30,000 per QALY.

Treatment continuation after progression

The model base case assumes that pharmacological treatment is discontinued on progression of IPF, for all treatments except sildenafil, and that sildenafil is only used after progression. This assumption is in line with the trials which inform the evidence for these treatment effects. However, it is possible that treatments would be effective for both unprogressed and progressed cases to some degree. *Table 87* shows the results obtained when the treatment effect is also applied after disease progression (for all treatments except sildenafil) and before disease progression (sildenafil only). The estimates of the effects of the treatments are assumed to be the same and they act in a similar fashion, i.e. by delaying progression or death from IPF.

Total QALYs and total costs for the pharmacological treatments are higher than the base case (see *Table 80*) but ICERs compared with BSC are lower. The pirfenidone ICER of £159,825 per QALY gained compares with £190,146 achieved in the base case, while the BIBF 1120 ICER of £115,892 per QALY gained compares with £132,658 achieved in the base case. Pirfenidone continues to be subject to extended dominance by a blend of BSC and BIBF 1120.

TABLE 86 Cost-effectiveness results using log-normal fit to PFS data

Treatment	Total costs (£)	Total QALYs	ICER vs. BSC (£/QALY)	ICER vs. next best option (£/QALY)
BSC	3069	3.69	–	–
Pirfenidone	115,671	4.62	121,519	Extended dominance
BIBF	273,966	6.52	95,786	95,786

TABLE 87 Cost-effectiveness results assuming treatment continuation after progression

Treatment	Total costs (£)	Total QALYs	ICER vs. BSC (£/QALY)	ICER vs. next best option (£/QALY)
BSC	3084	2.98	–	–
Pirfenidone	131,378	3.79	159,825	Extended dominated
BIBF	275,674	5.33	115,892	115,892

Log-normal fit to progression-free survival data and treatment continuation after progression

This scenario combines the previous two scenarios and assumes both a log-normal fit to PFS data and treatment continuation after IPF progression. Results are given in *Table 88*. Total costs and total QALYs in the treatment arms are higher than in either of the separate scenarios, but total costs in the BSC arm do not increase at the same rate. Consequently, the ICERs achieved are worse than the ICERs shown in *Table 86* for the log-normal-only scenario.

Expected value of perfect information for inhaled *N*-acetylcysteine and BIBF 1120

The base-case results for the IPF model indicate that inhaled NAC is the only non-dominated treatment within the £30,000 per QALY WTP threshold. However, the treatment effect of inhaled NAC does not achieve statistical significance in the NMA and, as such, is unlikely to be adopted in clinical practice. The potential benefit of reducing the uncertainty around the inhaled NAC treatment effect was examined in EVPI and EVPPI analyses. The EVPI of BIBF 1120 compared with BSC was also investigated, as base-case results show that it is similarly undominated by other treatments.

Expected value of perfect information and EVPPI were calculated for current and future IPF populations, assuming an annual IPF incidence in the UK of 5000,[13] a discount rate of 3.5% and a conservative estimate of 5 years for the lifetime of the decision problem. The PSA was run for 1000 iterations; in EVPPI analysis this was repeated for 50 different values of the inhaled NAC treatment effect.

Figure 27 displays the population EVPI at increasing thresholds of WTP for inhaled NAC compared with BSC and for BIBF 1120 compared with BSC. At a threshold for cost-effectiveness of £20,000 per QALY, the population EVPI in the case of inhaled NAC versus BSC is £19.4M, and in the case of BIBF 1120 versus BSC is £0 (see *Figure 27*). This indicates that there is some value in further research into inhaled NAC, but that there is no value in further research into BIBF 1120 at the £20,000 WTP threshold. The EVPI of £0 for BIBF 1120 arises because given the assumed probability distributions and base-case values, BIBF 1120 is never cost-effective at the £20,000 WTP threshold. However, the base-case price of BIBF 1120 is assumed, and fixed in PSA, and variations to its price would affect this finding.

The EVPPI for the treatment effect of inhaled NAC at a WTP of £20,000 per QALY is £15.8M. This makes up a large proportion of the total EVPI for inhaled NAC versus BSC at £20,000 per QALY (£19.4M, see *Figure 27*) and demonstrates that the treatment effect of inhaled NAC has a large bearing on the uncertainty in this comparison. This is consistent with the results of the deterministic sensitivity analyses, which indicate that the model is most sensitive to assumed treatment effects. These findings suggest that any future research into the treatments considered should, as a priority, focus on reducing the uncertainty in estimates of the treatment effect of inhaled NAC versus BSC.

TABLE 88 Cost-effectiveness results assuming treatment continuation after progression and log-normal fit to PFS data

Treatment	Total costs (£)	Total QALYs	ICER vs. BSC (£/QALY)	ICER vs. next best option (£/QALY)
BSC	3069	3.69	–	–
Pirfenidone	172,179	5.03	126,215	Extended dominated
BIBF	383,892	7.58	98,049	98,049
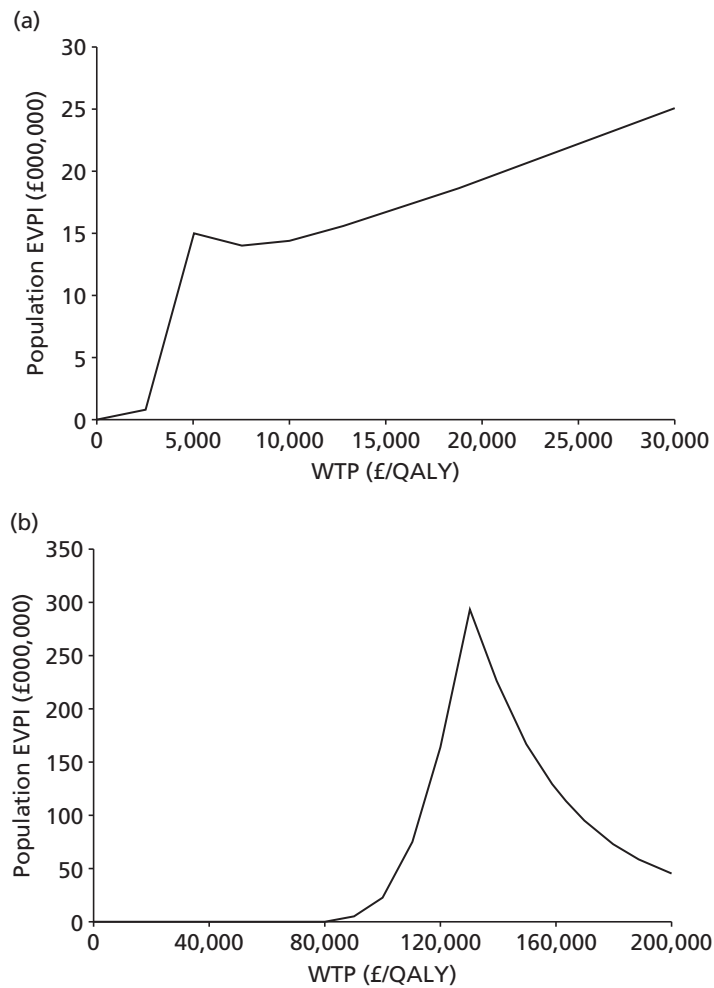
FIGURE 27 Expected value of perfect information curves for inhaled NAC and BIBF 1120 compared with BSC (base-case model). (a) Inhaled NAC vs. BSC and (b) BIBF 1120 vs. BSC.

Summary of independent model of cost-effectiveness

- An independent economic model was developed. The model incorporates three survival curves which are used to inform the probabilities of transition from three health states: unprogressed IPF, progressed IPF and lung transplant. Treatment effects are applied in the model base case to the survival curve which informs transition from the unprogressed disease state. Utility values are applied to the health states to estimate total QALYs. Costs are included for treatments, treatment monitoring, acute exacerbations, lung transplant and adverse events.

- The model base-case results show increased survival for five of the treatments compared with BSC, at increased cost. Only one treatment, inhaled NAC, is cost-effective at a WTP threshold of £30,000, but its treatment effect does not achieve statistical significance in NMA or the primary study. The clinical effectiveness review found the primary study to have an unclear risk of bias. Of the remaining treatments, only pirfenidone and BIBF 1120 achieve a statistically significant treatment effect. However, pirfenidone is subject to extended dominance by a blend of BSC and BIBF 1120.

- The effect of variation to the parameter values used in the economic model was evaluated in sensitivity analyses. Results were found to be generally robust to such variation but were particularly sensitive to changes in value of the treatment effect parameters.

- PSA was used to estimate the probabilities that the treatments are cost-effective at the £20,000 and £30,000 WTP thresholds. Inhaled NAC was the most cost-effective option at both thresholds, with a probability of around 65% in both cases.

- The monthly cost of BIBF 1120 is not yet available from conventional sources (e.g. BNF) and was not provided by the manufacturer. The monthly cost assumed in the model base case is £3274. One-way sensitivity analysis indicates that BIBF 1120 must cost less than £736 per month to be considered the cost-effective treatment option at a WTP threshold of £30,000 per QALY, compared with BSC.

- The treatment effect of inhaled NAC compared with BSC is associated with an EVPPI of £15.8M at a WTP threshold of £20,000. If estimates of this treatment effect can achieve the level of statistical significance demanded by clinical practice, and are consistent with earlier estimates,[69] then inhaled NAC would represent a clinically acceptable and cost-effective treatment option for IPF at a WTP threshold of £20,000. Further research to more precisely estimate its effect on rate of FVC per cent decline is desirable.

Chapter 5 Discussion

Statement of principal findings

Clinical effectiveness

For patients with mild to moderate IPF, 10 studies evaluating five pharmacological interventions were included. In a small RCT with an unclear risk of bias, treatment with azathioprine and prednisolone led to an improvement in survival, compared with placebo and prednisolone (when this was age adjusted), but not in lung function. It is possible that this trial included participants who would be diagnosed with NSIP according to current criteria, which may explain the treatment effect. Follow-up was 12 months. BIBF 1120 300 mg/day was more favourable on some measures of lung function, rates of acute exacerbations and the number of deaths, compared with placebo; however, the primary outcome of annual rate of decline in FVC was not statistically significantly different in this 54-month study. Treatment with NAC was evaluated in three studies: in combination with azathioprine and prednisolone in two and as a single agent in an inhaled format in one. Follow-up was approximately 12 months in these studies. Study results were mixed; there was no benefit from triple therapy for FVC compared with placebo in one study, but there was a benefit for VC when compared with double therapy in another study. Inhaled single-therapy NAC was not statistically significantly different from a control intervention. Secondary outcomes were reported, similarly with mixed results across the three studies. The two studies with triple-therapy interventions had a low risk of bias; however, the study using nebulised NAC had an unclear risk of bias. Pirfenidone was studied in four RCTs, and meta-analysis of these shows that pirfenidone appears to demonstrate a favourable effect on FVC when compared with placebo treatment. However, caution is required in interpreting these data as outcomes pooled were different and, as a consequence, a standardised mean difference analysis was undertaken; in addition, the timing of assessment of these outcomes varied (from 48 weeks to 72 weeks). Results of secondary outcomes were less certain. In a small crossover study, thalidomide appeared to improve cough, cough-related quality life and respiratory-related QoL, compared with treatment with placebo.

One study assessed sildenafil for those with moderate to severe IPF; some of the participants in this study might also have had pulmonary hypertension. Results on the primary outcome, a 20% improvement on the 6MWT, were not statistically significant between the sildenafil and placebo groups. Results on secondary outcomes were mixed, with some favourable to sildenafil and others favouring placebo. This study followed participants for 12 weeks.

Adverse events were generally mild to moderate and were reasonably well balanced between the treatments and the placebo arms across the studies. The exception was thalidomide, where adverse events were more common. Severe adverse events appeared to be more common in one study in those treated with triple therapy.

Three studies evaluated non-pharmacological treatments for people with IPF. Two compared pulmonary rehabilitation with a control and the other compared a disease management approach with a control. Results are uncertain as to the effects of pulmonary rehabilitation interventions; differences favouring pulmonary rehabilitation were seen on some outcomes but not others. There was an uncertain risk of bias in one study and a high risk of bias in another. There were also baseline differences between groups on many key outcomes in one study. In both studies, outcomes were assessed immediately after the cessation of the pulmonary rehabilitation programme (range 10–12 weeks). The third study reported limited evidence on the effects of a disease management programme in IPF; there were no statistically significant differences in dyspnoea and QoL results were mixed. This study has an uncertain risk of bias and follow-up was at 6 weeks, immediately after the programme had completed.

A NMA focusing on pharmacological treatments for IPF and assessing FVC end points was undertaken on six interventions. The FVC end point was measured on two continuous scales [FVC per cent predicted and absolute change from baseline (litres)] and, as such, the NMA used the standardised mean difference approach. This converts the outcome measures to a common scale. Only the fixed effect results for BIBF 1120 and pirfenidone were statistically significant, reducing the rate of decline in FVC compared with placebo. A head-to-head comparison of BIBF 1120 versus pirfenidone showed a trend favouring BIBF, although this was not statistically significant. Caution is required in the interpretation of the results of the NMA, as differences in study design and patient populations may cause the results to be biased, and there was little capacity to formally explore heterogeneity because of the small number of comparators.

Cost-effectiveness

A systematic search of the literature found one full economic evaluation of treatment for patients with IPF, which examined the benefits of a testing strategy prior to treatment with triple therapy. It did not examine the cost-effectiveness of any IPF treatment. A systematic review of studies of QoL for patients with IPF identified 23 relevant studies. Results were mixed but appear to show that IPF has an adverse effect on HRQoL compared with population norms, and that this is likely to diminish as IPF become more severe.

A new decision-analytic model was developed to estimate the cost-effectiveness of six pharmacological treatments for IPF. Results from this model were presented for a base-case analysis and a number of sensitivity and scenario analyses.

The model base-case results show increased survival for five of the treatments compared with BSC, at increased cost. Only one treatment, inhaled NAC, is cost-effective at a WTP threshold of £30,000, but its treatment effect does not achieve statistical significance in the single primary study identified,[69] or in NMA. The clinical effectiveness review found the primary study to have an unclear risk of bias. PSA results for the six treatments and BSC show that inhaled NAC has a probability of 65% of being cost-effective at a WTP of £20,000. Only pirfenidone and BIBF 1120 achieve a statistically significant treatment effect in NMA but each has a probability of 0% of being cost-effective at a WTP of £30,000 per QALY.

The monthly cost of BIBF 1120 is not yet available from conventional sources (e.g. BNF), and was not provided by the manufacturer. The monthly cost assumed in the model base case is £3274. One-way sensitivity analysis indicated that BIBF 1120 must cost less than £736 per month to be considered the cost-effective treatment option compared with BSC at a WTP threshold of £30,000 per QALY.

If estimates of the inhaled NAC treatment effect can achieve the level of statistical significance demanded by clinical practice, and are consistent with earlier estimates,[69] then inhaled NAC would represent a clinically acceptable and cost-effective treatment option for IPF at a WTP threshold of £20,000. VOI analysis was undertaken to determine the expected value of this uncertainty. Considering only BSC and inhaled NAC, the EVPI is £19.4M at a WTP threshold of £20,000 per QALY. The EVPPI, for the treatment effect of inhaled NAC alone, is £15.8M at a WTP threshold of £20,000. This indicates that a large part of the overall inhaled NAC versus BSC EVPI stems from uncertainty in the inhaled NAC treatment effect, and that, consequently, further research into this effect should be a research priority.

General discussion

The findings of this evidence synthesis provide evidence on the effectiveness and cost-effectiveness of a range of interventions which are currently used, or proposed to be used, in the management of IPF in the UK. No previous systematic reviews have included all potentially relevant treatments for IPF, and there have been only limited economic evaluations in this area. The results of this evidence synthesis will, therefore, be of value to clinicians and patients, and complement recent national guidance by NICE in the UK.[31] The past few years have seen an increasing interest in the management of IPF, with pharmacological companies evaluating a range of potential interventions, and a number of influential

bodies producing guidelines. The current state of the evidence suggests that there are few treatments that have any effect on surrogate outcomes which can be linked through evidence to patient-related outcomes such as mortality (some specific issues related to some of the interventions reviewed are discussed below). In terms of a cure, it is considered that lung transplantation is the only intervention available which has curative intention; however, no evidence on lung transplant was eligible for inclusion in this evidence synthesis and so this could not be evaluated formally.

Overall, there was a scarcity of studies on symptom control. Good practice guidelines suggest that, as IPF is an incurable disease, patients diagnosed with it should be provided with specialist palliative care from the point of diagnosis, and yet in terms of evidence that would meet the inclusion criteria of this review, there is little documented about the most beneficial approaches to palliation in IPF. A palliative care approach may be helpful, and often the reason that there is limited evidence is because it has not been tested in a specific group, in which case it may be appropriate to use evidence from similar conditions where the evidence base is more complete. In this synthesis, evidence, albeit limited, for symptomatic control of cough suggests that thalidomide may have the potential to improve QoL. Non-pharmacological interventions to improve control, anxiety and QoL would appear to have some potential benefit but, again, the quality of the evidence base is generally poor. Despite the limited evidence on pulmonary rehabilitation, it is likely that, overall, there may be potential benefits that are less easy to quantify in an intervention study. While it is important for clinicians to discuss the limitations in the evidence base for treatments such as pulmonary rehabilitation with patients, the evidence base in other areas of respiratory medicine is more certain and this information may allow patients the opportunity to access these treatments at a time when there are very few treatments available, and until the evidence base is more complete.

One included intervention, azathioprine, is no longer indicated in IPF but was included by the advisory group because there are some patients who still receive azathioprine treatment in restricted circumstances. The included study used azathioprine in combination with prednisolone; as noted in the results section (see *Assessment of effectiveness of pharmacological interventions for mild to moderate idiopathic pulmonary fibrosis*), this was a small study, and there was the possibility that a proportion of patients might actually have had NSIP, which has generally been shown to be more responsive to treatment. The treatment effect shown in this trial may well have been due to this factor and, therefore, results should be interpreted cautiously.

At present, BIBF 1120 does not have marketing authorisation for this indication. BIBF 1120 is a potential treatment for IPF which, at the time of writing, has only been studied in a dose-ranging RCT where the primary outcome measure was not statistically significant. Ongoing RCTs of BIBF 1120 will report in the coming years and the effects of this treatment can be evaluated more fully at this time.

The evidence for NAC therapy is complicated to interpret given the differences in the three studies included in the systematic review. To work out the standalone effect of NAC overall is difficult, given that two studies compare NAC in combination with azathioprine and prednisolone (triple therapy) and in the other NAC in an inhaled format was used. The comparators also differed, with one comparing triple therapy with placebo, one comparing triple therapy with azathioprine and prednisolone, and the other comparing inhaled NAC with a control. In addition, the doses of prednisolone used varied between the studies; for example, in the Raghu and colleagues IPFCRN 2012[68] trial, higher doses were used than in the Demedts and colleagues study[67] and than are likely to be used in current practice. The NAC versus placebo results from the Raghu and colleagues ongoing trial will, in the future, provide evidence of the clinical effectiveness of single-therapy NAC. Recent guidelines[31] suggest that triple therapy is not indicated because of the safety issues; however, many patients who take NAC believe that they have benefited and may find it difficult to stop taking it. The recent NICE guideline states that the benefits of NAC therapy are unknown but that if it is used as a treatment this should be discussed with patients. This would seem to be an appropriate decision based on the interpretation of the evidence made in this review.

The NMA and economic evaluation consider inhaled NAC separately from triple therapy as it has a different method of administration, and NMA results give a treatment effect estimate for inhaled NAC that is better than the estimate for triple therapy (although not significantly so).

Pirfenidone was subject to a NICE technology appraisal in 2012.[23] Pirfenidone was recommended as a treatment option in IPF where an individual has an FVC between 50% and 80% predicted. In addition, this is only when the treatment is provided at a previously agreed (undisclosed) discount price. The results on FVC outcomes were statistically significant when the studies were combined in a meta-analysis. Across the individual trials, however, one trial showed a different pattern with non-significant results on FVC measures. It is unclear why this would be the case, especially when considered in the light of the very similarly conducted trial by the same group. The results of the ongoing trial will be key in the consideration of the clinical effectiveness of pirfenidone.

Thalidomide is a treatment which has a different proposed indication as it is suggested to be used for the treatment of cough, which is a common symptom in IPF. Therefore, no attempt to improve lung function was made in the included trial and as such it was not included in the NMA or economic evaluation.

The proposed indication for the use of sildenafil is for treatment at the more severe end of the spectrum, and many patients in the severe stages are considered likely to have pulmonary hypertension. Current NICE guidance does not recommend sildenafil.

Discussion of network meta-analysis

We have used standardised mean differences to synthesise two different measurements of a common FVC end point. Hedges' *g* formulation was preferred over Cohen's because of the small sample size in one of the included trials. Alternatively, a bivariate meta-analytic framework could have been used to borrow strength across these end points.[164] While increasingly used in the pairwise NMA, this methodology is relatively novel in NMA and may require access to individual patient-level data.[165]

Converting the standardised mean differences to log-odds ratios using an established methodology facilitated including the widest range of comparators and enabled use of these data in our de novo cost-effectiveness model (see *Chapter 4*, *Independent economic evaluation*). In our NMA, only two treatments (BIBF 1120 and pirfenidone) showed important differences in slowing the decline in FVC compared with placebo under a fixed-effect model.

However, there remain concerns with the data. With only 10 studies and seven comparators, there is little capability to explore heterogeneity within these data using meta-regression techniques. Differences in study design and patient populations that could modify relative treatment effects may invalidate the assumption of exchangeability and cause the results of indirect comparisons to be biased. There was potential heterogeneity in terms of baseline FVC per cent predicted but insufficient data to allow us to control for this in the analysis. While the Zisman and colleagues IPFCRN 2010 study[72] was notably in a more severe population (defined as $DL_{CO} < 35\%$), its inclusion in the analysis did not affect the estimates for the other treatment effects as it connected only to placebo. There is also the disagreement between the estimates for azathioprine (Raghu and colleagues[70] and Demedts and colleagues[67]). While the Raghu and colleagues trial is very small (19 subjects) and thus subject to high variability, the differences could be attributable to trial design, unreported heterogeneity between the studies (e.g. proportion of NSIP subjects), or a placebo effect.[166]

No published NMA in IPF have been identified. One pairwise meta-analysis of pirfenidone versus placebo has been presented in abstract form[97] and was cited in the manufacturer submission to NICE[167] (see also *Chapter 3*, *Existing systematic reviews*). The authors also used standardised mean differences and their results were similar to those observed here, at around 0.2 SDs. However, owing to longitudinal data availability in the sponsor's trials, FVC was modelled at distinct time intervals with the observed treatment effect maintained up to 72 weeks in the Capacity 004 study but dropping off in the Capacity 006 study.[62]

No explanation was offered for this. It would have been interesting if the data had been modelled jointly in a longitudinal framework (e.g. Ding and Fu[168]). Furthermore, despite the same data input, the Capacity 004 study appeared to produce slightly different effect estimates from those reported here (at least from the graph in the abstract), although no data are provided on whether fixed or random effects were presented or which standardised mean difference formulation was chosen.

Discussion of the modelling approach with respect to diagnosis of idiopathic pulmonary fibrosis

As discussed in *Chapter 1* (see *Diagnosis*), IPF can be difficult to diagnose and patients can experience delay in diagnosis from the time of symptomatic presentation. This delay in diagnosis can potentially lead to variation in the ability to benefit from treatments. Our economic model did not commence at the time of diagnosis as it was based on the population in the Capacity trial, with patients starting treatment at a recently diagnosed baseline value. The model, therefore, reflects delays in diagnosis which occurred in the Capacity trial and an assumption was made that diagnostic delays were the same across all treatment arms.

There are two separate issues with respect to timing of diagnosis of IPF: time since diagnosis which is reported, albeit inconsistently, in the clinical trials; and the delay in diagnosis observed in clinical practice. Time since diagnosis in our model reflects the Capacity trial. With better-reported data we could potentially have included time since diagnosis as a predictor in the meta-analysis. If it was found to be a relative treatment effect modifier, this could impact on results. The impact of delay in diagnosis, which is likely to lead to a variation in health at point of diagnosis, is less clear. We have essentially modelled a more severe patient subpopulation than those patients who would have been included at time of disease occurrence. We have no information with which to model this delay in diagnosis or its likely impact on patients' ability to benefit or duration of therapy. It may be patients with milder symptoms who are being missed or not presenting to GPs. It could be argued that more severe patients may respond less well to treatment or, conversely, they may have more capacity to benefit from treatment. In any case, a consistent approach has been adopted by excluding this from our model. Our focus on a more severe subpopulation of IPF equates to less variation (uncertainty) in the model than if we included a factor for delay in diagnosis. It follows, therefore, that the VOI results are likely to be a more conservative estimate than would be the case if we were to model delay in diagnosis.

Strengths and limitations of the assessment

The evidence synthesis has the following strengths:

- It is independent of vested interests.
- It has been undertaken following the principles for conducting systematic reviews and economic evaluations. The methods were set out in a research protocol, which defined the research question, study selection criteria, quality assessment criteria, data extraction process and the process by which the methods will be employed at different stages in the systematic review and economic evaluations.
- An advisory group has informed the evidence synthesis from its initiation. The research protocol and a copy of the draft final report were sent to the advisory group for review and comment. The interventions to be included were decided by the advisory group to ensure coverage of any treatment relevant to current management in the NHS.
- The systematic review brings together the evidence for the clinical effectiveness and cost-effectiveness of treatments for IPF and of HRQoL of people with IPF. This evidence has been critically appraised and presented in a consistent and transparent manner.
- A new economic model has been developed following recognised guidelines and systematic searches have been conducted to identify data to populate the different parameters. The main results have been summarised and presented.

In contrast, the evidence synthesis has certain limitations:

- Although only one non-English language study was identified and included in the systematic review of clinical effectiveness, restrictions on translation meant that a full assessment and critique of this study could not be made. The authors of this primary study were contacted to request additional information where it was deemed that this might have provided helpful data; however, no responses were received.
- Many of the included studies compared treatments with placebo; there were few studies directly comparing different interventions. An indirect comparison through a NMA was performed; however, there are known limitations in the use of indirect comparisons and, therefore, caution is recommended in the interpretation of these results.
- A meta-analysis and NMA used the standardised mean difference, which standardises the study findings so that they are expressed on a common scale. This was our assumption when combining mean change in FVC per cent predicted with absolute change in FVC, albeit the former was adjusted for certain baseline characteristics. Interpretation of the results of a meta-analysis using the standardised mean difference is also not intuitive. A threshold of the magnitude of the effect can be defined a priori, or, as in this case, the standardised mean differences can be converted to log-odds ratios using established methodologies. These should be considered when interpreting the results.
- There were insufficient data to explore the heterogeneity between trials in the NMA. Differences in study design or patient populations, reported or unreported, that could impact on relative treatment effects could invalidate the assumptions underlying the NMA. In addition, it was not possible to rank the treatments in the NMA owing to the limited data.
- The economic model uses a relative measure, absolute decline in FVC per cent predicted, as its measure of disease progression. It is possible that some bias is introduced by use of such a measure as patients with very different FVC per cent predicted may be considered together in the same disease state, with an unknown distribution. However, absolute decline in FVC per cent predicted is a standard measure of disease progression in IPF and insufficient evidence was available to inform rates of disease progression by treatment at different starting FVCs.
- BIBF 1120 is a new treatment and at the time of writing was not being marketed. An assumed price was, consequently, used in the economic model. Base-case results will require revision once an actual price is known.
- The model assumes that a pharmacological treatment has a constant effect on relative rate of FVC per cent decline compared with BSC, irrespective of the time for which the treatment is taken or the FVC achieved. This is potentially a generous assumption, as a treatment may become less effective with time and/or as IPF progresses. Nevertheless, even with this assumption, only one treatment was found to be cost-effective at a WTP of £30,000 per QALY.
- Model utility values post lung transplant are lower than the utility found for the progressed IPF state. This suggests no QoL benefit to lung transplant in a progressed IPF population. The post-lung transplant utility data were obtained more than 10 years ago and may require updating, while further work on QoL in progressed IPF would help to provide more robust utility estimates for this health state.

Suggested research priorities

Key research recommendations in the recent NICE guideline[31] include a recommendation to assess the effects of pulmonary rehabilitation in IPF. The guideline also recommends research into the effects of ambulatory oxygen and antireflux therapy. Given the limited evidence base on palliative care measures in IPF identified in the present evidence synthesis, we are in agreement with these recommendations. Well-designed studies, preferably RCTs with adequately described interventions, are required, particularly in the area of pulmonary rehabilitation.

Inhaled NAC has been studied in only one small RCT in a Japanese population. Although results favoured treatment with inhaled NAC, the effect was not statistically significant. However, given the limitations of this study, and the results of the NMA and EVPI, a well-designed RCT of inhaled NAC should be considered. According to our search of ongoing RCTs, none are currently under way.

One small RCT has been identified that investigates thalidomide for cough in IPF. As a potential symptomatic treatment, the results of the study show that thalidomide may have some benefits; however, this needs to be considered in the light of the adverse events. A well-conducted RCT of thalidomide is recommended.

Lung transplant is seen as a curative treatment for IPF; however, no studies of lung transplant that met the criteria of the present evidence synthesis have been undertaken in IPF. Although lung transplant in people with IPF is unlikely to be appropriate for studying in a RCT, a well-conducted prospective cohort study may provide further evidence of its benefits.

Most of the included pharmacological studies were placebo controlled. A head-to-head trial of BIBF 1120 and pirfenidone may be appropriate, although there are a number of issues that would need to be considered in any trial design (e.g. sample size, recruitment) and these may mean that a trial is not feasible. In addition, our economic evaluation suggests that neither treatment is currently cost-effective.

There are a number of ongoing studies that have been identified in the present report; for example, trials of pirfenidone, BIBF 1120 and single-therapy NAC are ongoing. When these conclude and report their findings, an update of the present evidence synthesis is recommended.

Conclusions

This evidence synthesis has identified limited evidence of the clinical effectiveness of a number of available treatments for IPF. Pirfenidone and BIBF 1120 appear to be effective; however, general recommendations cannot be made in terms of their cost-effectiveness due to limitations in the evidence base. Further research is required in a number of areas as outlined above.

Acknowledgements

We would like to thank members of our advisory group who provided expert advice and comments on the protocol and/or a draft of this report:

Note: this document and any associated economic model are protected by intellectual property rights, which are owned by the University of Southampton. Anyone wishing to modify, adapt, translate, reverse engineer, decompile, dismantle or create derivative work based on the economic model must first seek the agreement of the property owners.

Contributions of authors

Emma Loveman (Senior Research Fellow) developed the original research application, developed the research protocol, contributed to the background section, assisted in the development of the search strategy, assessed studies for inclusion, extracted data from and quality assessed included studies, synthesised evidence, drafted and edited the final report and project managed the study.

Vicky R Copley (Senior Research Fellow, Health Economics) assessed studies for inclusion, extracted data from and quality assessed included studies, synthesised evidence, developed the economic evaluation and drafted the report.

Jill Colquitt (Senior Research Fellow) developed the original research application, contributed to the development of the research protocol, contributed to the background section, assessed studies for inclusion, extracted data from and quality assessed included studies, synthesised evidence and drafted the report.

Publications

Loveman E, Copley VR, Colquitt JL, Scott DA, Clegg AJ, Jones J, *et al.* The effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: systematic review, network meta-analysis and health economic evaluation. *BMC Pharmacol Toxicol* 2014;**15**:63.

References

1. Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. *Orphanet J Rare Dis* 2008;**3**:8. http://dx.doi.org/10.1186/1750-1172-3-8

2. du Bois RM. An earlier and more confident diagnosis of idiopathic pulmonary fibrosis. *Eur Respir J* 2012;**21**:141–6. http://dx.doi.org/10.1183/09059180.00000812

3. Guenther A, Korfei M, Mahavadi P, von der Beck D, Ruppert C, Markart P. Unravelling the progressive pathophysiology of idiopathic pulmonary fibrosis. *Eur Respir J* 2012;**21**:152–60. http://dx.doi.org/10.1183/09059180.00001012

4. Swigris JJ, Stewart AL, Gould MK, Wilson SR. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;**3**:61. http://dx.doi.org/10.1186/1477-7525-3-61

5. Dempsey OJ. Clinical review: idiopathic pulmonary fibrosis – past, present and future. *Respir Med* 2006;**100**:1871–85. http://dx.doi.org/10.1016/j.rmed.2006.08.017

6. Collard HR, Moore BB, Flaherty KR, Brown KK, Kaner RJ, King TE Jr, *et al.* Acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2007;**176**:636–43. http://dx.doi.org/10.1164/rccm.200703-463PP

7. 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* 2000;**161**:646–64. http://dx.doi.org/10.1164/ajrccm.161.2.ats3-00

8. American Thoracic Society. American Thoracic Society/European Respiratory Society International multidisciplinary consensus classification of idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2002;**165**:277–304.

9. Wells AU, Hirani N, on behalf of the BTS interstitial lung disease guideline group. Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. *Thorax* 2008;**63**:v1–58. http://dx.doi.org/10.1136/thx.2008.101691

10. American Thoracic Society. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;**183**:788–824. http://dx.doi.org/10.1164/rccm.2009-040GL

11. Gribbin J, Hubbard RB, Jeune IL, Smith CJ, West J, Tata LJ. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Thorax* 2006;**61**:980–5. http://dx.doi.org/10.1136/thx.2006.062836

12. Mejía M, Carrillo G, Rojas-Serrano J, Estrada A, Suárez T, Alonso D, *et al.* Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. *Chest* 2009;**136**:10–15. http://dx.doi.org/10.1378/chest.08-2306

13. Navaratnam V, Fleming KM, West J, Smith CJ, Jenkins RG, Fogarty A, *et al.* The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* 2011;**66**:462–7. http://dx.doi.org/10.1136/thx.2010.148031

14. Fernández Pérez ER, Daniels CE, Schroeder DR, St. Sauver J, Hartman TE, Bartholmai BJ, *et al.* Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest* 2010;**137**:129–37. http://dx.doi.org/10.1378/chest.09-1002

15. Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2006;**174**:810–6. http://dx.doi.org/10.1164/rccm.200602-163OC

16. Costabel U. Idiopathic pulmonary fibrosis: recent milestones in disease management. *Eur Respir J* 2012;**21**:140. http://dx.doi.org/10.1183/09059180.00000712

17. Song JW, Hong SB, Lim CM, Koh Y, Kim DS. Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *Eur Respir J* 2011;**37**:356–63. http://dx.doi.org/10.1183/09031936.00159709

18. Cottin V. Changing the idiopathic pulmonary fibrosis treatment approach and improving patient outcomes. *Eur Respir J* 2012;**21**:161–7. http://dx.doi.org/10.1183/09059180.00001112

19. Agarwal R, Jindal SK. Acute exacerbation of idiopathic pulmonary fibrosis: a systematic review. *Eur J Intern Med* 2008;**19**:227–35. http://dx.doi.org/10.1016/j.ejim.2007.04.024

20. Mura M, Zompatori M, Pacilli AM, Fasano L, Schiavina M, Fabbri M. The presence of emphysema further impairs physiologic function in patients with idiopathic pulmonary fibrosis. *Respir Care* 2006;**51**:257–65.

21. Nathan SD, Noble PW, Tuder RM. Idiopathic pulmonary fibrosis and pulmonary hypertension: connecting the dots. *Am J Respir Crit Care Med* 2007;**175**:875–80. http://dx.doi.org/10.1164/rccm.200608-1153CC

22. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, *et al.* Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. *Chest* 2011;**140**:221–9. http://dx.doi.org/10.1378/chest.10-2572

23. National Institute for Health and Care Excellence. *Pirfenidone for Treating Idiopathic Pulmonary Fibrosis*. Technology appraisal TA282. London: NICE; 2013.

24. Bajwah S, Higginson IJ, Ross JR, Wells AU, Birring SS, Patel A, *et al.* Specialist palliative care is more than drugs: a retrospective study of ILD patients. *Lung* 2012;**190**:215–20. http://dx.doi.org/10.1007/s00408-011-9355-7

25. Clark M, Cooper B, Singh S, Cooper M, Carr A, Hubbard R. A survey of nocturnal hypoxaemia and health related quality of life in patients with cryptogenic fibrosing alveolitis. *Thorax* 2001;**56**:482–6. http://dx.doi.org/10.1136/thorax.56.6.482

26. Ryerson CJ, Donesky D, Pantilat SZ, Collar, Collard HR. Dyspnea in idiopathic pulmonary fibrosis: a systematic review. *J Pain Symptom Manage* 2012;**43**:771–82. http://dx.doi.org/10.1016/j.jpainsymman.2011.04.026

27. Lewis D, Scullion J. Palliative and end-of-life care for patients with idiopathic pulmonary fibrosis: challenges and dilemmas. *Int J Palliat Nurs* 2012;**18**:331–7. http://dx.doi.org/10.12968/ijpn.2012.18.7.331

28. Booth S, Bausewein C, Higginson IJ, Moosavi SH. Pharmacological treatment of refractory breathlessness. *Expert Rev Respir Med* 2009;**3**:21–36. http://dx.doi.org/10.1586/17476348.3.1.21

29. Martinez FJ, Safrin S, Weycker D, Starko KM, Bradford WZ, King TE Jr, *et al.* The clinical course of patients with idiopathic pulmonary fibrosis. *Ann Intern Med* 2005;**142**:963–7. http://dx.doi.org/10.7326/0003-4819-142-12_Part_1-200506210-00005

30. Health and Social Care Information Centre. *Hospital Episode Statistics: Inpatient Data, Main Procedures and Interventions*. London: Health and Social Care Information Centre; 2009. URL: www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=192 (accessed 14 June 2013).

31. National Institute for Health and Care Excellence. *Diagnosis and Management of Suspected Idiopathic Pulmonary Fibrosis*. Clinical guideline CG163. London: NICE; 2013.

32. Spagnolo P, Tonelli R, Cocconcelli E, Stefani A, Richeldi L. Idiopathic pulmonary fibrosis: diagnostic pitfalls and therapeutic challenges. *Multidiscip Respir Med* 2012;**7**:42. http://dx.doi.org/10.1186/2049-6958-7-42

33. Cottin V, Cordier JF. Velcro crackles: the key for early diagnosis of idiopathic pulmonary fibrosis? *Eur Respir J* 2012;**40**:519–21. http://dx.doi.org/10.1183/09031936.00001612

34. Lamas DL, Kawut SM, Bagiella E, Phillip N, Arcasoy SM. Delayed access and survival in idiopathic pulmonary fibrosis. A cohort study. *Am J Respir Crit Care Med* 2011;**184**:842–7. http://dx.doi.org/10.1164/rccm.201104-0668OC

35. Collard HR, Tino G, Noble PW, Shreve MA, Michaels M, Carlson B, *et al.* Patient experiences with pulmonary fibrosis. *Respir Med* 2007;**101**:1350–54. http://dx.doi.org/10.1016/j.rmed.2006.10.002

36. Schoenheit G, Becattelli I, Cohen AH. Living with idiopathic pulmonary fibrosis: an in-depth qualitative survey of European patients. *Chron Respir Dis* 2011;**8**:225–31. http://dx.doi.org/10.1177/1479972311416382

37. Lacasse Y, Goldstein R, Lasserson TJ, Martin. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2006;**4**:CD003793.

38. Christie JD, Edwards LB, Kucheryavaya AY, Benden C, Dipchand AI, Dobbels F, *et al.* The Registry of the International Society for Heart and Lung Transplantation: 29th official adult heart transplant report 2012. *J Heart Lung Transplant* 2013;**31**:1073. http://dx.doi.org/10.1016/j.healun.2012.08.004

39. George TJ, Arnaoutakis GJ, Shah AS. Lung transplant in idiopathic pulmonary fibrosis. *Arch Surg* 2011;**146**:1204–9. http://dx.doi.org/10.1001/archsurg.2011.239

40. Kilic A, Merlo CA, Conte J, Shah. Lung transplantation in patients 70 years old or older: have outcomes changed after implementation of the lung allocation score? *J Thorac Cardiovasc Surg* 2012;**144**:1133–8. http://dx.doi.org/10.1016/j.jtcvs.2012.07.080

41. Orens JB, Estenne M, Arcasoy S, Conte JV, Corris P, Egan JJ, *et al.* International guidelines for the selection of lung transplant candidates: 2006 update–a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2006;**25**:745–55. http://dx.doi.org/10.1016/j.healun.2006.03.011

42. Allen JG, Arnaoutakis GJ, Weiss ES, Merlo CA, Conte JV, Shah AS. The impact of recipient body mass index on survival after lung transplantation. *J Heart Lung Transplant* 2010;**29**:1026–33. http://dx.doi.org/10.1016/j.healun.2010.05.005

43. Kazerooni EA, Martinez FJ, Flint A, Jamadar DA, Gross BH, Spizarny DL, *et al.* Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *Am J Roentgenol* 1997;**169**:977–83. http://dx.doi.org/10.2214/ajr.169.4.9308447

44. Thabut G, Mal H, Castier Y, Groussard O, Brugiere O, Marrash-Chahla R, *et al.* Survival benefit of lung transplantation for patients with idiopathic pulmonary fibrosis. *J Thorac Cardiovasc Surg* 2003;**126**:469–75. http://dx.doi.org/10.1016/S0022-5223(03)00600-7

45. Mason DP, Brizzio ME, Alster JM, McNeill AM, Murthy SC, Budev MM, *et al.* Lung transplantation for idiopathic pulmonary fibrosis. *Ann Thorac Surg* 2013;**84**:1121–8. http://dx.doi.org/10.1016/j.athoracsur.2007.04.096

46. Grossman RF, Frost A, Zamel N, Patterson GA, Cooper JD, Myron PR, *et al.* Results of single-lung transplantation for bilateral pulmonary fibrosis. The Toronto Lung Transplant Group. *NEJM* 1990;**322**:727–33. http://dx.doi.org/10.1056/NEJM199003153221104

47. De Oliveira NC, Osaki S, Maloney J, Cornwell RD, Meyer KC. Lung transplant for interstitial lung disease: outcomes for single versus bilateral lung transplantation. *Interact Cardiovasc Thorac Surg* 2012;**14**:263–7. http://dx.doi.org/10.1093/icvts/ivr085

48. Lanken PN, Terry PB, DeLisser HM, Fahy BF, Hansen-Flaschen J, Heffner JE, *et al.* An official American Thoracic Society clinical policy statement: palliative care for patients with respiratory diseases and critical illnesses. *Am J Respir Crit Care Med* 2008;**177**:912–27. http://dx.doi.org/10.1164/rccm.200605-587ST

49. Lee JS, McLaughlin S, Collard HR. Comprehensive care of the patient with idiopathic pulmonary fibrosis. *Curr Opin Pulm Med* 2011;**17**:384–54. http://dx.doi.org/10.1097/MCP.0b013e328349721b

50. Bradley B, Branley HM, Egan JJ, Greaves MS, Hansell DM, Harrison NK, *et al.* Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. *Thorax* 2008;**63**:1–58. http://dx.doi.org/10.1136/thx.2008.101691

51. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991;**85**:25–31. http://dx.doi.org/10.1016/S0954-6111(06)80166-6

52. Williams JE, Singh SJ, Sewell L, Guyatt GH, Morgan MD. Development of a self-reported Chronic Respiratory Questionnaire (CRQ-SR). *Thorax* 2001;**56**:954–59. http://dx.doi.org/10.1136/thorax.56.12.954

53. American Thoracic Society. ATS Statement: guideline for the six-minute walk test. *Am J Respir Crit Care Med* 2002;**166**:111–17. http://dx.doi.org/10.1164/ajrccm.166.1.at1102

54. Singh SJ, Morgan MD, Scott S, Walters D, Hardman AE. Development of a shuttle walking test of disability in patients with chronic airways obstruction. *Thorax* 1992;**47**:1019–24. http://dx.doi.org/10.1136/thx.47.12.1019

55. Horton MR, Santopietro V, Matthew L, Horton KM, Polito AJ, Liu MC, *et al.* Thalidomide for the treatment of cough in idiopathic pulmonary fibrosis. *Ann Intern Med* 2012;**157**:398–406. http://dx.doi.org/10.7326/0003-4819-157-6-201209180-00003

56. European Medicines Agency. *Public Summary of Opinion on Orphan Designation: Nintedanib for the Treatment of Idiopathic Pulmonary Fibrosis*. London: European Medicines Agency, Committee for Orphan Medicinal Products; 2013. URL: www.ema.europa.eu/docs/en_GB/document_library/Orphan_designation/2013/05/WC500143247.pdf (accessed 20 May 2013).

57. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. 3rd edn. York: CRD, University of York; 2009.

58. Higgins JP, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011)*. London: The Cochranes; 2011. URL: www.cochrane-handbook.org (accessed 14 June 2013).

59. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;**313**:275–83. http://dx.doi.org/10.1136/bmj.313.7052.275

60. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al.* Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess* 2004;**8**(36). http://dx.doi.org/10.3310/hta8360

61. Jastrzebski D, Kozielski J, Zebrowska A. [Pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis with inspiratory muscle training.] *Pneumonologia i Alergologia Polska* 2008;**76**:131–41.

62. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011;**377**:1760–9. http://dx.doi.org/10.1016/S0140-6736(11)60405-4

63. Dowman L, McDonald CF, Hill C, Lee A, Barker K, Boote C, *et al.* The benefits of exercise training in interstitial lung disease: protocol for a multicentre randomised controlled trial. *BMC Pulm Med* 2013;**13**:8. http://dx.doi.org/10.1186/1471-2466-13-8

64. Gomez O, Ramos CF, Cardenas D, Gaunaurd I, Eustis N, Cohen M, *et al.* Exercise training during pulmonary rehabilitation increases activity levels of IPF patients. *Am J Respir Crit Care Med* 2012;**185**:A4885.

65. Taniguchi H, Ebina M, Kondoh Y, Ogura T, Azuma A, Suga M, *et al.* Pirfenidone in idiopathic pulmonary fibrosis. *Eur Respir J* 2010;**35**:821–9. http://dx.doi.org/10.1183/09031936.00005209

66. Azuma A, Nukiwa T, Tsuboi E, Suga M, Abe S, Nakata K, *et al.* Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005;**171**:1040–7. http://dx.doi.org/10.1164/rccm.200404-571OC

67. Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM, *et al.* High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005;**353**:2229–42. http://dx.doi.org/10.1056/NEJMoa042976

68. Raghu G, Anstrom KJ, King J, Lasky JA, Martinez FJ. Prednisone, azathioprine, and *N*-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 2012;**366**:1968–77. http://dx.doi.org/10.1056/NEJMoa1113354

69. Homma S, Azuma A, Taniguchi H, Ogura T, Mochiduki Y, Sugiyama Y, *et al.* Efficacy of inhaled *N*-acetylcysteine monotherapy in patients with early stage idiopathic pulmonary fibrosis. *Respirology* 2012;**17**:467–77. http://dx.doi.org/10.1111/j.1440-1843.2012.02132.x

70. Raghu G, Depaso WJ, Cain K, Hammar SP, Wetzel CE, Dreis DF, *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* 1991;**144**:291–6. http://dx.doi.org/10.1164/ajrccm/144.2.291

71. Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, *et al.* Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011;**365**:1079–87. http://dx.doi.org/10.1056/NEJMoa1103690

72. 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* 2010;**363**:620–8. http://dx.doi.org/10.1056/NEJMoa1002110

73. Nishiyama O, Kondoh Y, Kimura T, Kato K, Kataoka K, Ogawa T, *et al.* Effects of pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. *Respirology* 2008;**13**:394–9. http://dx.doi.org/10.1111/j.1440-1843.2007.01205.x

74. Lindell KO, Olshansky E, Song MK, Zullo TG, Gibson KF, Kaminski N, *et al.* Impact of a disease-management program on symptom burden and health-related quality of life in patients with idiopathic pulmonary fibrosis and their care partners. *Heart Lung* 2010;**39**:304–13. http://dx.doi.org/10.1016/j.hrtlng.2009.08.005

75. Horton MR, Santopeitro V, Matthew L, Richardson B, Danoff SK, Polito AJ. Thalidomide suppresses cough and improves quality of life in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012;**185**:A3635.

76. Caldwell D, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;**331**:897–900. http://dx.doi.org/10.1136/bmj.331.7521.897

77. Hasselblad V. Meta-analysis of multitreatment studies. *Med Decis Making* 1998;**18**:37–43. http://dx.doi.org/10.1177/0272989X9801800110

78. National Institute for Health and Clinical Excellenece. *Guide to the Methods of Technology Appraisal*. London: NICE; 2004.

79. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappalleri JC, *et al.* Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on indirect treatment comparisons good research practices – part 2. *Value Health* 2011;**14**:429–37. http://dx.doi.org/10.1016/j.jval.2011.01.011

80. Albera C. Challenges in idiopathic pulmonary fibrosis trials: the point on end-points. *Eur Respir Rev* 2011;**20**:195–200. http://dx.doi.org/10.1183/09059180.00001711

81. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;**17**:279–301. http://dx.doi.org/10.1177/0962280207080643

82. Egger M, Davey Smith G, Altman DG. *Systematic Reviews in Health Care: Meta-Analysis in Context.* London: BMJ Books; 2001. http://dx.doi.org/10.1002/9780470693926

83. Dias S, Welton NJ, Sutton AJ, Ades AE. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials 2011*. URL: www.nicedsu.org.uk (accessed 14 June 2013).

84. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 2002;**64**:583–639. http://dx.doi.org/10.1111/1467-9868.00353

85. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;**7**:434–55.

86. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A* 2009;**172**:137–59. http://dx.doi.org/10.1111/j.1467-985X.2008.00552.x

87. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med* 2000;**19**:3127–31. http://dx.doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

88. Dias S, Welton NJ, Sutton AJ, Caldwell D, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making* 2013;**33**:641–56. http://dx.doi.org/10.1177/0272989X12455847

89. Bucher AC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomised controlled trials. *J Clin Epidemiol* 1997;**50**:683–90. http://dx.doi.org/10.1016/S0895-4356(97)00049-8

90. Dias S, Welton NJ, Caldwell D, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;**29**:932–44. http://dx.doi.org/10.1002/sim.3767

91. Dequen P, Sutton AJ, Scott DA, Abrams KR. Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery. *Value Health* 2014;**17**:416–23. http://dx.doi.org/10.1016/j.jval.2014.02.013

92. Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol* 2013;**13**:35. http://dx.doi.org/10.1186/1471-2288-13-35

93. Cox DR, Snell EJ. *Analysis of Binary Data*. London: Chapman & Hall; 1989.

94. Spagnolo P, Del GC, Luppi F, Cerri S, Balduzzi S, Walters EH, *et al.* Non-steroid agents for idiopathic pulmonary fibrosis. *Cochrane Database Syst Rev* 2010;**9**:CD003134. http://dx.doi.org/doi:10.1002/14651858.CD003134.pub2

95. Richeldi L, Davies HR, Ferrara G, Franco F. Corticosteroids for idiopathic pulmonary fibrosis. *Cochrane Database Syst Rev* 2003;**3**:CD002880.

96. Bajwa EK, Ayas NT, Schulzer M, Mak E, Ryu JH, Malhotra A. Interferon-gamma1b therapy in idiopathic pulmonary fibrosis: a meta-analysis. *Chest* 2005;**128**:203–6. http://dx.doi.org/10.1378/chest.128.1.203

97. King TE Jr, Albera C, du Bois RM, Bradford W, Costabel U, Noble PW, *et al.* The effect of treatment with pirfenidone on longitudinal change In lung volume in patients with idiopathic pulmonary fibrosis (IPF): a meta-analysis of outcomes in four randomized controlled clinical trials. *Am J Respir Crit Care Med* 2011;**183**:A5302.

98. Ryerson CJ, Donesky DA, Pantilat SZ, Collard HR. Dyspnea in idiopathic pulmonary fibrosis: a systematic review. *J Pain Symptom Manage* 2012;**43**:771–82. http://dx.doi.org/10.1016/j.jpainsymman.2011.04.026

99. Hagaman JT, Kinder BW, Eckman MH. Thiopurine *S*-methyltranferase testing in idiopathic pulmonary fibrosis: a pharmacogenetic cost-effectiveness analysis. *Lung* 2010;**188**:125–32. http://dx.doi.org/10.1007/s00408-009-9217-8

100. Cooper K, Mendes D, Picot J, Loveman E. *Pirfenidone for the Treatment of Idiopathic Pulmonary Fibrosis: A Single Technology Appraisal*. SHTAC; 2012. URL: www.nice.org.uk/guidance/ta282/documents/idiopathic-pulmonary-fibrosis-pirfenidone-evidence-review-group-report2 (accessed 1 December 2014).

101. Gisbert JP, Gomollon F, Cara C, Luna M, González-Lama Y, Pajares JM, *et al.* Thioprine methyltransferase activity in Spain: a study of 14,545 patients. *Dig Dis Sci* 2007;**52**:1262–9. http://dx.doi.org/10.1007/s10620-006-9119-z

102. Schwab M, Schaffeler E, Marx C, Fischer C, Lang T, Behrens C, *et al.* Azathioprine therapy and adverse drug reactions in patients with inflammatory bowel disease: impact of thiopurine *S*-methyltransferase polymorphism. *Pharmacogenetics* 2002;**12**:429–36. http://dx.doi.org/10.1097/00008571-200208000-00003

103. Ansari A, Hassan C, Duley J, Marinaki A, Shobowale-Bakre EM, Seed P, *et al.* Thiopurine methyltransferase activity and the use of azathioprine in inflammatory bowel disease. *Aliment Pharmacol Ther* 2002;**16**:1743–50. http://dx.doi.org/10.1046/j.1365-2036.2002.01353.x

104. Perri D, Cole DE, Friedman O, Piliotis E, Mintz S, Adhikari NK. Azathioprine and diffuse alveolar haemorrhage: the pharmacogenetics of thiopurine methyltransferase. *Eur Respir J* 2007;**30**:1014–7. http://dx.doi.org/10.1183/09031936.00026107

105. Meggitt SJ, Reynolds NJ. Azathioprine for atopic dermatitis. *Clin Exp Dermatol* 2001;**26**:369–75. http://dx.doi.org/10.1046/j.1365-2230.2001.00837.x

106. Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010;**65**:921–6. http://dx.doi.org/10.1136/thx.2010.139121

107. Swigris J, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;**8**:77. http://dx.doi.org/10.1186/1477-7525-8-77

108. Patel AS, Siegert R, Brignall K, Gordon P, Steer S, Desai SR, *et al.* The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012;**67**:804–10. http://dx.doi.org/10.1136/thoraxjnl-2012-201581

109. Martinez TY, Pereira CA, dos Santos ML, Ciconelli RM, Guimarães SM, Martinez JA. Evaluation of the short-form 36-item questionnaire to measure health-related quality of life in patients with idiopathic pulmonary fibrosis. *Chest* 2000;**117**:1627–32. http://dx.doi.org/10.1378/chest.117.6.1627

110. Ozalevli S, Karaali HK, Ilgin D, Ucan ES. Effect of home-based pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. *Multidiscip Respir Med* 2010;**5**:31–7. http://dx.doi.org/10.1186/2049-6958-5-1-31

111. Baddini Martinez JA, Martinez TY, Lovetro Galhardo FP, de Castro Pereira CA. Dyspnea scales as a measure of health-related quality of life in patients with idiopathic pulmonary fibrosis. *Med Sci Monit* 2002;**8**:CR405–10.

112. Jastrzebski D, Kozielski J, Banas A, Cebula T, Gumola A, Ziora D, *et al.* Quality of life during one-year observation of patients with idiopathic pulmonary fibrosis awaiting lung transplantation. *J Physiol Pharmacol* 2005;**56**(Suppl. 4):99–105.

113. Kozu R, Jenkins S, Senjyu H. Effect of disability level on response to pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. *Respirology* 2011;**16**:1196–202. http://dx.doi.org/10.1111/j.1440-1843.2011.02029.x

114. Kozu R, Senjyu H, Jenkins SC, Mukae H, Sakamoto N, Kohno S. Differences in response to pulmonary rehabilitation in idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease. *Respiration* 2011;**81**:196–205. http://dx.doi.org/10.1159/000315475

115. Tomioka H, Imanaka K, Hashimoto K, Iwasaki H. Health-related quality of life in patients with idiopathic pulmonary fibrosis–cross-sectional and longitudinal study. *Intern Med* 2007;**46**:1533–42. http://dx.doi.org/10.2169/internalmedicine.46.6218

116. Mermigkis C, Bouloukaki I, Antoniou KM, Mermigkis D, Psathakis K, Giannarakis I, *et al.* CPAP therapy in patients with idiopathic pulmonary fibrosis and obstructive sleep apnea: does it offer a better quality of life and sleep? *Sleep Breath* 2013;**17**:1137–43. http://dx.doi.org/10.1007/s11325-013-0813-8

117. De Vries J, Seebregts A, Drent M. Assessing health status and quality of life in idiopathic pulmonary fibrosis: which measure should be used? *Respir Med* 2000;**94**:273–8. http://dx.doi.org/10.1053/rmed.1999.0736

118. Peng S, Li Z, Kang J, Hou X. Cross-sectional and longitudinal construct validity of the Saint George's Respiratory Questionnaire in patients with IPF. *Respirology* 2008;**13**:871–9. http://dx.doi.org/10.1111/j.1440-1843.2008.01359.x

119. Nishiyama O, Taniguchi H, Kondoh Y, Kimura T, Ogawa T, Watanabe F, *et al.* Health-related quality of life in patients with idiopathic pulmonary fibrosis. What is the main contributing factor? *Respir Med* 2005;**99**:408–14. http://dx.doi.org/10.1016/j.rmed.2004.09.005

120. Tzanakis N, Samiou M, Lambiri I, Antoniou K, Siafakas N, Bouros D. Evaluation of health-related quality-of-life and dyspnea scales in patients with idiopathic pulmonary fibrosis. Correlation with pulmonary function tests. *Eur J Intern Med* 2005;**16**:105–12. http://dx.doi.org/10.1016/j.ejim.2004.09.013

121. Lechtzin N, Hilliard ME, Horton MR. Validation of the Cough Quality-of-Life Questionnaire in patients with idiopathic pulmonary fibrosis. *Chest* 2013;**143**:1745–9. http://dx.doi.org/10.1378/chest.12-2870

122. Nishiyama O, Taniguchi H, Kondoh Y, Kimura T, Kataoka K, Nishimura K, *et al.* Health-related quality of life does not predict mortality in idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2012;**29**:113–18.

123. Verma G, Marras T, Chowdhury N, Singer L. Health-related quality of life and 6 min walk distance in patients with idiopathic pulmonary fibrosis. *Can Respir J* 2011;**18**:283–7.

124. Zimmermann CS, Carvalho CR, Silveira KR, Yamaguti WP, Moderno EV, Salge JM, *et al.* Comparison of two questionnaires which measure the health-related quality of life of idiopathic pulmonary fibrosis patients. *Braz J Med Biol Res* 2007;**40**:179–87. http://dx.doi.org/10.1590/S0100-879X2007000200004

125. Raghu G, King TE Jr., Behr J, Brown KK, du Bois RM, Leconte I, *et al.* Quality of life and dyspnoea in patients treated with bosentan for idiopathic pulmonary fibrosis (BUILD-1). *Eur Respir J* 2010;**35**:118–23. http://dx.doi.org/10.1183/09031936.00188108

126. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, *et al.* The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med* 2010;**104**:296–304. http://dx.doi.org/10.1016/j.rmed.2009.09.006

127. Lutogniewska W, Jastrzebski D, Wyrwol J, Ksiazek B, Ochman M, Kowalski K, *et al.* Dyspnea and quality of life in patients referred for lung transplantation. *Eur J Med Res* 2010;**15**(Suppl. 2):76–8.

128. King J, Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA, *et al.* BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008;**177**:75–81. http://dx.doi.org/10.1164/rccm.200705-732OC

129. Huang IC, Willke RJ, Atkinson MJ, Lenderking WR, Frangakis C, Wu AW. US and UK versions of the EQ-5D preference weights: does choice of preference weights make a difference? *Qual Life Res* 2007;**16**:1065–72. http://dx.doi.org/10.1007/s11136-007-9206-4

130. Cheung K, Oemar M, Oppe M, Rabin R. *EQ-5D User Guide: Basic Information on How to Use the EQ-5D*. 2009. URL: www.euroqol.org/fileadmin/user_upload/Documenten/PDF/User_Guide_v2_March_2009.pdf (accessed 14 May 2013).

131. Ware JE, Sherbourne CD. The MOS 36 item short-form health survery (SF-36). *Med Care* 1992;**30**:473–80. http://dx.doi.org/10.1097/00005650-199206000-00002

132. Ware JE, Gandek B, for the IQOLA project. Overview of the SF-36 health survey and the International Quality of Life Assessment (IQOLA) project. *J Clin Epidemiol* 1998;**51**:903–12. http://dx.doi.org/10.1016/S0895-4356(98)00081-X

133. Prescott-Clarke P, Primatesta P, editors. *Health Survey for England 1996*. London: The Stationery Office; 1998.

134. Thomas M, Goddard D, Hickman M, Hunter P. *General Household Survey 1992.* Office of Population Censuses and Surveys. London: HMSO; 1994.

135. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. *Am Rev Respir Dis* 1992;**145**:1321–7. http://dx.doi.org/10.1164/ajrccm/145.6.1321

136. Jones PW, Forde Y. *St George's Respiratory Questionnaire Manual*. London: St George's, University of London; 2009.

137. Jastrzebski D, Gumola A, Gawlik R, Kozielski J. Dyspnea and quality of life in patients with pulmonary fibrosis after six weeks of respiratory rehabilitation. *J Physiol Pharmacol* 2006;**57**:139–48.

138. HM Treasury. *The Green Book: Appraisal and Evaluation in Central Government*. URL: www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government (accessed 14 March 2013).

139. Richeldi L, Ryerson CJ, Lee JS, Wolters PJ, Koth LL, Ley B, *et al.* Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 2012;**67**:407–11. http://dx.doi.org/10.1136/thoraxjnl-2011-201184

140. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, *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* 2011;**183**:788–824. http://dx.doi.org/10.1164/rccm.2009-040GL

141. Shulgina L, Cahn A, Chilvers E, Parfrey H, Clark A, Wilson E, *et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole. *Thorax* 2011;**68**:155–62. http://dx.doi.org/10.1136/thoraxjnl-2012-202403

142. Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. *Respir Med* 2009;**103**:1430–5. http://dx.doi.org/10.1016/j.rmed.2009.04.024

143. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd edn. Oxford: Oxford University Press; 2005.

144. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006.

145. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol* 2012;**12**:9. http://dx.doi.org/10.1186/1471-2288-12-9

146. Latimer NR. Survival analysis for economic evaluations alongside clinical trials–extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013;**33**:743–54. http://dx.doi.org/10.1177/0272989X12472398

147. Collard HR, King J, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;**168**:538–42. http://dx.doi.org/10.1164/rccm.200211-1311OC

148. Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, *et al.* Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J* 2010;**35**:830–5. http://dx.doi.org/10.1183/09031936.00155108

149. Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. *Med Decis Making* 1994;**14**:52–8. http://dx.doi.org/10.1177/0272989X9401400107

150. NHS Choices. *Lung Transplant*. NHS Choices; 2013. URL: www.nhs.uk/Conditions/Lung-transplant/Pages/Introduction.aspx (accessed 14 June 2013).

151. Emin A, Rogers C, Taylor R, Barber K, van der Meulen J, Parameshwar J, *et al. UK Cardiothoracic Transplant Audit: In Patients who Received a Transplant Between 1st July 1995 and 31st March 2011*. London: RCS, Clinical Effectiveness Unit; 2011.

152. Christie JD, Edwards LB, Kucheryavaya AY, Benden C, Dobbels F, Kirk R, *et al.* The Registry of the International Society for Heart and Lung Transplantation: twenty-eighth Adult Lung and Heart-Lung Transplant Report–2011. *J Heart Lung Transplant* 2011;**30**:1104–22. http://dx.doi.org/10.1016/j.healun.2011.08.004

153. Office for National Statistics. *Interim Life Tables, England and Wales, 2009–2011*. London: Office for National Statistics; 2013. URL: www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2009-2011/stb-2009-2011.html (accessed 16 June 2013).

154. Anyanwu AC, McGuire A, Rogers CA, Murday AJ. Assessment of quality of life in lung transplantation using a simple generic tool. *Thorax* 2001;**56**:218–22. http://dx.doi.org/10.1136/thorax.56.3.218

155. Lloyd A, Price D, Brown R. The impact of asthma exacerbations on health-related quality of life in moderate to severe asthma patients in the UK. *Prim Care Respir J* 2007;**16**:22–7. http://dx.doi.org/10.3132/pcrj.2007.00002

156. Ara R, Brazier J. Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). *Value Health* 2008;**11**:1131–43. http://dx.doi.org/10.1111/j.1524-4733.2008.00352.x

157. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health* 2013;**16**:202–10. http://dx.doi.org/10.1016/j.jval.2012.10.010

158. Menn P, Weber N, Holle R. Health-related quality of life in patients with severe COPD hospitalized for exacerbations – comparing EQ-5D, SF-12 and SGRQ. *Health Qual Life Outcomes* 2010;**8**:39. http://dx.doi.org/10.1186/1477-7525-8-39

159. Department of Health. *National Health Service England and Wales Electronic Drug Tariff May 2013*. 2013. URL: www.ppa.org.uk/ppa/edt_intro.htm (accessed 14 June 2013).

160. Joint Formulary Committee. *British National Formulary*. 65 edn. London: BMJ Group and Pharmaceutical Press; 2013.

161. Department of Health. *NHS Reference Costs 2011–2012.* URL: www.gov.uk/government/publications/nhs-reference-costs-financial-year-2011-to-2012 (accessed 8 November 2012).

162. Department of Health. *Costing Tool for Home Oxygen: Assessment and Review*. URL: www.gov.uk/government/publications/commissioning-toolkit-for-respiratory-services (accessed 14 June 2013).

163. Department of Health. *Service Specification: Home Oxygen Assessment and Review Service*. www.gov.uk/government/publications/commissioning-toolkit-for-respiratory-services (accessed 14 June 2013).

164. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med* 2007;**26**:78–97. http://dx.doi.org/10.1002/sim.2524

165. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;**7**:3. http://dx.doi.org/10.1186/1471-2288-7-3

166. Hawkins N, Scott DA. Cost-effectiveness analysis: discount the placebo at your peril. *Med Decis Making* 2010;**30**:536–43. http://dx.doi.org/10.1177/0272989X10362106

167. National Institute for Health and Care Excellence. *Pirfenidone for the Treatment of Idiopathic Pulmonary Fibrosis:. Manufacturer's Single Technology Appraisal (STA) Submission to NICE*. 2011. URL: www.nice.org.uk/nicemedia/live/13039/61701/61701.pdf (accessed 2 May 2013).

168. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med* 2013;**32**:2613–28. http://dx.doi.org/10.1002/sim.5688

169. Wells AU. FVC as a primary end-point in IPF treatment trials: making a silk purse from a sow's ear. *Thorax* 2012;**0**:1–2. http://dx.doi.org/doi: 10.1136/thoraxjnl-2012-202640

170. Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, *et al.* Prognostic implications of physiologic and radiologic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003;**168**:543–8. http://dx.doi.org/10.1164/rccm.200209-1112OC

171. King J, Safrin S, Starko KM, Brown KK, Noble PW, Raghu G, *et al.* Analyses of efficacy end points in a controlled trial of interferon-gamma1b for idiopathic pulmonary fibrosis. *Chest* 2005;**127**:171–7. http://dx.doi.org/10.1378/chest.127.1.171

172. Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, *et al.* Fibrotic idiopathic interstitial pneumonia. The prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 2003;**168**:531–7. http://dx.doi.org/10.1164/rccm.200210-1245OC

173. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;**184**:1382–9. http://dx.doi.org/10.1164/rccm.201105-0840OC

174. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;**183**:1231–7. http://dx.doi.org/10.1164/rccm.201007-1179OC

175. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;**184**:459–66. http://dx.doi.org/10.1164/rccm.201011-1790OC

176. Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;**60**:588–94. http://dx.doi.org/10.1136/thx.2004.035220

Appendix 1 Search strategy

All databases searched for the systematic review are presented below. Searches were updated on 26 June 2013.

Database searched	Clinical effectiveness searches	Cost-effectiveness and QoL searches
Cochrane Central Register of Controlled Trials	All available years	All available years
Cochrane Database of Systematic Reviews	All available years	All available years
Database of Abstracts of Reviews of Effects	All available years	All available years
EMBASE	All available years	All available years
Health Technology Assessment Database (Centre for Reviews and Dissemination)	All available years	All available years
Medline (Ovid)	All available years	All available years
Medline in-Process & Other Non-Indexed Citations	Searched 26 June 2013	
NHS Economic Evaluation Database		All available years
Web of Science (Science Citation Index Expanded and Conference Proceedings Citation Index)	1990–2013	1990–2013
Web of Science (Bioscience Information Centre Previews)	All available years	All available years
Searched for ongoing trials		
National Institute for Health Research Clinical Research Network (NIHR CRN Portfolio, formally UKCRN website)		
Current Controlled Trials		
Clinical trials.gov		
World Health Organization International Clinical Trials Registry Platform		
PROSPERO, Centre for Reviews and Dissemination		

The MEDLINE search strategy (presented below) for the systematic review of clinical effectiveness was adjusted as necessary for other electronic databases for both clinical effectiveness and cost-effectiveness (including quality of life information) searches. Search strategies for the systematic review are available from the authors on request. Citations identified by the searches were added to a Reference Manager database.

Medline search strategy

1. idiopathic pulmonary fibrosis/ (548)
2. IPF.tw. (1925)
3. (idiopath$ and pulmonary and fibro$).tw. (3334)
4. (idiopath$ and (lung and fibro$)).tw. (2502)
5. (((usual or ordinary) adj3 interstiti$) and pneumo$).tw. (698)
6. (((nonspecific or "non specific") adj3 interstitial) and pneumo$).tw. (583)
7. (idiopath$ and interstiti$ and pneumoni$).tw. (1383)
8. ("usual interstiti$" adj5 (lung or pulmonary or alveoli$)).tw. (191)
9. ("nonspecific interstiti$" adj5 (lung or pulmonary or alveoli$)).tw. (83)
10. ("non specific interstiti$" adj5 (lung or pulmonary or alveoli$)).tw. (38)
11. (cryptog$ and fibro$ and alveoli$).tw. (327)
12. or/1-11 (5331)
13. ("lung disease$" adj5 (interstiti$ or fibrosis or fibrotic)).tw. (5421)
14. pulmonary fibrosis/ (15,120)
15. exp lung diseases interstitial/ (42,412)
16. (pulmonary adj5 (fibrosis or fibrotic)).tw. (10,738)
17. (interstiti$ adj5 (pneumonia or lung or pulmonary or alveoli$)).tw. (12,889)
18. "diffuse parenchymal lung disease".tw. (82)
19. or/13-18 (65,328)
20. (idiopathic or unexplained or nonspecific or "non specific").tw. (178,936)
21. (((unknow$ or uncertain$) adj4 (origin$ or cause$ or aetiol$ or etiol*)) or idiopa$).tw. (102,573)
22. 20 or 21 (207,520)
23. 19 and 22 (6196)
24. 12 or 23 (7352)
25. (cystic adj fibro$).mp. (33,749)
26. 24 not 25 (7191)
27. limit 26 to humans (6808)
28. limit 26 to animals (785)
29. 26 not 27 not 28 (81)
30. 27 or 29 (6889)
31. Randomized Controlled Trials as Topic/ (77,921)
32. randomized controlled trial.pt. (322,018)
33. controlled clinical trial.pt. (83,725)
34. Controlled Clinical Trial/ (83,725)
35. placebos/ (30,626)
36. random allocation/ (73,596)
37. Double-Blind Method/ (113,512)
38. Single-Blind Method/ (15,853)
39. (random* adj2 allocat*).tw. (17,050)
40. placebo*.tw. (133,778)
41. ((singl* or doubl* or trebl* or tripl*) adj (blind* or mask*)).tw. (111,089)
42. crossover studies/ (29,092)
43. (crossover* or (cross adj over*)).tw. (49,865)
44. Research Design/ (65,242)
45. ((random* or control*) adj5 (trial* or stud*)).tw. (424,830)
46. Clinical Trials as Topic/ (158,570)
47. trial.ti. (97,491)
48. randomly.ab. (164,073)
49. (randomized or randomised).ab. (271,090)

50. or/31-49 (1,044,973)

51. 3 and 50 (257)

52. limit 30 to (controlled clinical trial or randomized controlled trial) (102)

53. 51 or 52 (292)

54. limit 30 to meta analysis (7)

55. 53 or 54 (294)

56. (rat or rats).ti. (676,338)

57. 55 not 56 (292)

58. (official and "idiopathic pulmonary fibrosis").ti. (1)

59. 57 OR 58 (293)

Reference lists

The reference lists of retrieved articles were examined for additional studies.

Other searches

The experts advisory group were contacted in order to obtain information about additional references and any ongoing studies.

British Societies and Conferences

Pulmonary Fibrosis Organisation, ASCEND trial, National Organisation for Rare Diseases, British Lung Foundation, National Heart Lung and Blood Institute, IPFnet, Coalition for IPF, ATS, British Thoracic Society and International Meeting on Pulmonary Rare Diseases and Orphan Drugs.

Appendix 2 Data extraction and quality assessment template

Reviewer 1:	Reviewer 2:	
Study details	**Participant details**	**Date: 9/8/12 Version:**
First author *et al.*, year, ref ID	*Number of participants:* total and number per treatment group	
Country:	*Sample attrition/dropout:*	
Design:	*Sample crossovers:*	
Number of centres:	*Inclusion criteria:*	
Funding:	*Exclusion criteria:*	
Intervention details	**Outcomes**	
Intervention	Primary outcomes:	
1.	Secondary outcomes:	
2.	Method of assessing outcome:	
Dose details:	Length of follow-up:	
Dose modifications:		
Concurrent treatment:		
Duration of treatment:		

Participant characteristics (add column for *p*-values if reported)		
	Intervention 1, *n* =	**Intervention 2, *n* =**
List here		
Comments		

Results			
	Intervention 1, *n*=	**Intervention 2, *n*=**	**Difference, *p*-value**
Outcome 1			
Comments			
Outcome 2			
Comments			
Outcome 3			
Comments			

Methodology

- Allocation to treatment groups:
- Blinding:
- Comparability of treatment groups:
- Method of data analysis:
- Sample size/power calculation:
- Attrition/drop-out:

General comments

- Generalisability: state severity of disease if reported, or estimate from baseline FVC based on range $\geq$ 70% 'mild', 55–69% 'moderate', < 55% severe
- Outcome measures:
- Intercentre variability:
- Conflict of interests:

Quality assessment/risk of bias[a]	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	
2. Was the allocation adequately concealed?	
3. Were the groups similar at the outset of the study in terms of prognostic factors?	
4. Was the care provider blinded?	
5. Was the patient blinded?	
6. Were outcome assessors blinded to the treatment allocation?	
7. (i) Were there any unexpected imbalances in drop-outs between groups? (ii) If so, were they explained or adjusted for?	
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	

a Amend for non RCTs.

Appendix 3 Clinical trial end points

Most outcomes reported in trials of IPF are surrogate measures. Reliable, objective measures to assess disease progression and treatment response are poorly defined and not fully validated. Generally, lung function is the most valid. A number have been found to be associated with mortality. These include FVC, 6MWT distance, DL_{CO} and dyspnoea. The FVC and 6MWT are suggested to be the best predictors of mortality at baseline in IPF.[80]

Other outcomes include HRQoL, PFS, peripheral oxygen saturation (SpO_2) during the SMWT, respiratory hospitalisations, and time to disease progression.

There is heterogeneity in the primary end points used in trials of treatments for IPF.

Clinical significance on these outcomes is often considered using the MCID approach, which is the smallest difference in an outcome that patients would perceive to be important, either beneficial or harmful, and lead to a clinician to consider a change in a patients status/therapy.[80]

Forced vital capacity/vital capacity

Forced vital capacity is measured with a spirometer, and is a measure of the maximum volume of air that can be expired from the lungs during a forced and complete expiration from a position of full inspiration. VC is the volume measured at the mouth, between the positions of full inspiration and full expiration. In obstructive airways disease, the VC may be greater than the FVC as a result of compression and collapse of narrowed airways during a forced manoeuvre. However, in IPF these two measurements should be the same and can be assessed as such in trial evidence.

The FVC shows good reliability, and correlates reasonably well with measures of physiologic function, dyspnoea and HRQoL. It can be measured according to standard protocols.

Forced vital capacity can be reported as either absolute values (litres) or a percentage of a reference value (predicted) for someone of the same age, sex, height and ethnic group. Reference values are determined from large population studies. A FVC percentage of 80–120% of the reference value is generally considered to be within normal limits.

While there is no fixed cut-off for disease severity, Nathan and colleagues[22] categorised participants in their study into terciles of per cent predicted FVC for mild (≥ 70%), moderate (55–69%) and severe (< 55%) disease. These can be used as a rough idea of severity; however, caution is required because many people have comorbidities which contribute to their disease severity, and these are not always captured by FVC. Recently, these terciles were applied by the manufacturer of pirfenidone in their submission to NICE and these were accepted in the NICE guidance.[23]

In trials, FVC can be presented as continuous or categorical variables. The former is more sensitive, the latter more advantageous in terms of interpreting the meaning of any change seen.[169] There have been a number of studies investigating what would be a clinically significant threshold of change on the FVC. There is an important distinction between these studies: some refer to the threshold in a cohort, and others to an individual-patient threshold. The traditional threshold for a clinically significant effect of decline in per cent predicted FVC in an individual accounts for measurement variation in the FVC. A clinically significant effect in a cohort is likely to be lower as there is no net effect from measurement variation when the FVC is averaged out.

Interpreting forced vital capacity within an individual

A decline in per cent predicted FVC of ≥ 10% (absolute change, see below) is recognised as clinically significant and highly predictive of mortality. A progressive, sustained decrease from baseline in absolute FVC is recognised in international guidelines as being consistent with progressive disease.[10] The Food and Drug Administration also proposes that a categorical reduction of FVC of ≥ 10% is clinically meaningful.

This degree of decline has been shown to be an independent predictor of mortality, associated with a 1.7- to 4.8-fold increased risk of death over 1 year.[146,147,170–172]

Interpreting forced vital capacity from within a cohort

More recent studies[147,173,174] suggest that even small changes in FVC (e.g. > 5% decline) indicate a poor prognosis/risk of death within the context of symptomatic deterioration. du Bois and colleagues[173] showed that a 1-year risk of death was more than twofold higher ($p < 0.001$) in those with a 24-week decline in FVC between 5% and 10%. The estimated MCID in this study was estimated at 2–6%.

Relative versus absolute change on categorical change in forced vital capacity per cent predicted

A ≥ 10% decline can be a relative decline (e.g. from 60% to 54%) or an absolute decline (e.g. from 60% to 50%). How this has been calculated can have an impact on the frequency and prognostic value of declines in FVC. A recent study[139] showed that the frequency of a ≥ 10% decline over 12 months was almost twice as high using relative change than absolute change. This has important management implications for the clinician.

The study found that there was no difference in predicting transplant-free survival at 2 years using either method. The authors suggest, therefore, that using relative decline maximises the changes of identifying meaningful change without sacrificing prognostic accuracy.

Where trials report a threshold of decline on FVC per cent predicted, for example ≥ 10%, it should be noted whether this is relative or absolute change to aid interpretation.

Six-minute walk test

This is a measure of exercise tolerance and functional status where the individual is asked to walk on a flat surface for 6 minutes. It is a reliable measure and shows only small variation within an individual over short periods of time. It correlates reasonably well with other measures.

The distance walked in the 6MWT is used in a variety of pulmonary diseases and is predictive of mortality in IPF. At baseline, people with mild to moderate IPF would likely have a distance walked of approximately 400 m. A 24-week decrement of > 50 m has been associated with a fourfold increase in the risk of death over the next 12 months ($p < 0.001$).[174]

The MCID was estimated as a decline of between 24–45 m in one study,[174] 28 m in another[126] and 29–34 m in a third.[142]

Studies in IPF have been limited by small sample sizes and narrow patient subsets, and, therefore, can yield inconsistent results between studies.

Although the test is self-paced, it carries the possibility of interoperator bias.

Diffusing capacity of the lung for carbon monoxide

This measures the ability of the lungs to transfer gas from inhaled air to the red blood cells in pulmonary capillaries.

The traditional view is that a reduction from baseline of 15% is significant. A decline in DL_{CO} from baseline has been associated with decreased survival, although the evidence is less consistent. King and colleagues[171] found that decreases in per cent predicted DL_{CO} of $\geq$ 15% had no association with mortality. Flaherty and colleagues[170] found that 6- and 12-month decreases in per cent predicted DL_{CO} of $\geq$ 10% were not predictors of subsequent mortality. du Bois and colleagues,[175] in a large study found that per cent predicted DL_{CO} and 24-week change in per cent predicted DL_{CO} were independent predictors of all-cause mortality. However, they also report that DL_{CO} may not be incrementally informative in differentiating between IPF patients based on their mortality risk.

Changes in DL_{CO} are, therefore, not clear indicators for mortality. In addition, the complex procedure to measure DL_{CO} means that there is variability in clinical practice and the test is not always available.

Progression-free survival

Progression-free survival has been used in IPF studies using predictive end points such as categorical changes in FVC and 6MWD.[80]

Few studies have used PFS and different studies appear to use different composite end points. Caution is, therefore, required in their interpretation.

Peripheral oxygen saturation during the 6-minute walk test

Desaturation (a decline in oxygen saturation < 88%) during the 6MWT is seen as a marker for increased risk of mortality.[10] There are difficulties with reproducibility for this outcome.

Other lung function outcomes

There has been some suggestion that baseline TLC and alveolar–arterial oxygen difference in partial pressures ($PA - aO_2$) may be predictive of survival, but no clear threshold exists[147] and few trials report these outcomes.

Dyspnoea

Dyspnoea is a key issue in IPF and baseline dyspnoea correlates with QoL and survival in several studies.[10]

A variety of different metrics for dyspnoea have been used, including the BDI, the Borg Dyspnoea Index, the UCSDSBQ and the CRP dyspnoea score. These are validated in other lung diseases.

The Borg Dyspnoea Index measures perceived breathlessness on a scale of 0 (none) to 10 (maximum). It correlates well with diffusing capacity and blood oxygenation at rest and during exercise.

The UCSDSBQ indicates severity of dyspnoea on a scale from 0 to 5 on 21 activities of daily living, along with three ratings on limitations caused by dyspnoea or fear of dyspnoea, for a total score ranging from 0 to 120, with higher score indicating more dyspnoea. The minimally important difference for this is reported by some authors to be 5 points.[72]

Health-related quality of life

Idiopathic pulmonary fibrosis has an effect on physical health, general health, energy levels, respiratory symptoms and level of independence.[176] The HRQoL among patients has been found to be variable, however, and this cannot be fully explained by measures of dyspnoea or pulmonary function. There are no disease-specific measures of HRQoL for IPF. Generic instruments or non-IPF specific instruments have been used in clinical trials: EQ-5D, SF-36, SGRQ and WHOQOL (World Health Organization Quality of Life) instrument.

There is uncertainty over whether these are valid in the IPF population or reliable over time, or in response to specific treatments. As dyspnoea is an issue in IPF, it is uncertain if this is always picked up with the generic instruments well enough.

The EQ-5D is a widely used instrument developed to measure health utility. It comprises five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), which can be rated for severity. Scores from the health states are converted through value sets (or preference weights) obtained from general population studies to generate a single utility value.[129,130] This ranges from –0.59 to 1 and represents health on a continuum from worst to best. For more information see *Chapter 4*, *Systematic review of health-related quality of life studies*.

The SF-36 is a generic measure of health status. It is a 36-item survey and has eight scales covering functional health and well-being concepts.[131] Each score ranges from 0 to 100, with a higher score indicating better function. For more information see *Chapter 4*, *Systematic review of health-related quality of life studies*.

The SGRQ is validated for COPD and other chronic respiratory diseases, and appears to possess reasonable validity and reliability in IPF.[120] It has been found to correlate reasonably with FVC. The SGRQ is a disease-specific measure which aims to measure the impact of obstructive airways disease on overall health, daily life and perceived well-being.[135] It comprises 50 items within two parts, which cover three components: symptoms, activity and impact. A total score can also be calculated which summarises the impact of the disease on overall health status. Scores range from 0 to 100, where higher scores indicate greater impact on HRQoL. For more information on the SGRQ see *Chapter 1*, *Systematic review of health-related quality of life studies*.

There is an IPF-specific version of the SGRQ, the SGRQ-I,[106] which is deemed to have acceptable reliability and validity; however, this needs further testing. This does not appear to have been used in trials to date. Other disease-specific instruments such as the ATAQ-IPF[107] and the K-BILD[108] have also recently been developed.

Acute exacerbation
Patients with IPF may suffer periods of acute respiratory decline, often due to complications such as infection but not always from a known cause. These acute exacerbations are characterised by sudden worsening of respiratory symptoms accompanied by hypoxaemia and the outcome for those who have had an acute exacerbation is poor. There is uncertainty over what the definition of an acute exacerbation is and clinical trials do not provide definitions.

Respiratory hospitalisation
Respiratory hospitalisation has been found to be predictor of mortality, but definitions may vary.[80,175]

Appendix 4 List of excluded studies with rationale

Studies excluded at first screen of full papers

Antoniou KM, Alexandrakis MG, Sfiridaki K, Tsiligianni I, Perisinakis K, Tzortzaki EG, *et al.* Th1 cytokine pattern (IL-12 and IL-18) in bronchoalveolar lavage fluid (BALF) before and after treatment with interferon gamma-1b (IFN-gamma-1b) or colchicine in patients with idiopathic pulmonary fibrosis (IPF/UIP). *Sarcoidosis Vasc Diffuse Lung Dis* 2004;**21**:105–10. (Reason for exclusion: outcomes.)

Azuma A, Taguchi Y, Ogura T, Ebina M, Taniguchi H, Kondoh Y, *et al.* Exploratory analysis of a phase III trial of pirfenidone identifies a subpopulation of patients with idiopathic pulmonary fibrosis as benefiting from treatment. *Respir Res* 2011;**12**:143. (Reason for exclusion: study design post hoc subgroup analysis.)

Bando M, Hosono T, Mato N, Nakaya T, Yamasawa H, Ohno S, *et al.* Long-term efficacy of inhaled *N*-acetylcysteine in patients with idiopathic pulmonary fibrosis. *Intern Med* 2010;**49**:2289–96. (Reason for exclusion: study design.)

Ghofrani HA, Wiedemann R, Rose F, Schermuly RT, Olschewski H, Weissmann N, *et al.* Sildenafil for treatment of lung fibrosis and pulmonary hypertension: a randomised controlled trial. *Lancet* 2002;**360**:895–900. (Reason for exclusion: participants.)

Han MK, Bach DS, Hagan PG, Yow E, Flaherty KR, Toews GB, *et al.* Sildenafil preserves exercise capacity in patients with idiopathic pulmonary fibrosis and right-sided ventricular dysfunction. *Chest* 2013;**143**:1699–708. (Reason for exclusion: study design.)

Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Short term improvement in exercise capacity and symptoms following exercise training in interstitial lung disease. *Thorax* 2008;**63**:549–54. (Reason for exclusion: participants.)

Inoue Y, Azuma A, Taniguchi H, Ogura T, Tadayasu Y, Fujimoto T, *et al.* The pharmacokinetics of BIBF 1120 alone or in combination with pirfenidone in Japanese patients with idiopathic pulmonary fibrosis (IPF). *Respirology* 2011;Conference:318. (Reason for exclusion: outcomes.)

Kubo H, Yanai M, Azuma A. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis: a hidden subgroup? *Am J Respir Crit Care Med* 2013;**187**:1029–30. (Reason for exclusion: study design.)

Mackay A. Interferon gamma-1b does not improve survival for patients with idiopathic pulmonary fibrosis. *Thorax* 2010;**65**:186. (Reason for exclusion: study design.)

Polonski L, Krzywiecki A, Polonska A, Tendera M, Cwiertka P, Oklek K, *et al.* [Effects of long-term oxygen therapy in patients with idiopathic pulmonary fibrosis. I. Effect on the course of the primary disease and on pulmonary circulation.] *Polskie Archiwum Medycyny Wewnetrznej* 1995;**94**:331–6. (Reason for exclusion: participants.)

Polonski L, Kusnierz B, Krzywiecki A, Polonska A, Tendera M, Oklek K, *et al.* [Effects of long-term oxygen therapy in patients with idiopathic pulmonary fibrosis. II. Effect of oxygen therapy on function of heart ventricles.] *Polskie Archiwum Medycyny Wewnetrznej* 1995;**94**:337–41. (Reason for exclusion: participants, outcomes.)

Raghu G, King TE Jr, Behr J, Brown KK, du Bois RM, Leconte I, *et al.* Quality of life and dyspnoea in patients treated with bosentan for idiopathic pulmonary fibrosis (BUILD-1). *Eur Respir J* 2010;**35**:118–23. (Reason for exclusion: study design post hoc subgroup analysis.)

Shulgina L, Cahn A, Chilvers E, Parfrey H, Clark A, Wilson E, *et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole. *Thorax* 2011;Conference:A63–4. (Reason for exclusion: participants.)

Shulgina L, Cahn AP, Chilvers ER, Parfrey H, Clark AB, Wilson EC, *et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: a randomised controlled trial. *Thorax* 2013;**68**:155–62. (Reason for exclusion: participants.)

Taniguchi H, Kondoh Y, Ebina M, Azuma A, Ogura T, Taguchi Y, *et al.* The clinical significance of 5% change in vital capacity in patients with idiopathic pulmonary fibrosis: extended analysis of the pirfenidone trial. *Respir Res* 2011;**12**:93. (Reason for exclusion: study design post hoc subgroup analysis.)

Varney V, Parnell H, Salisbury D, Ratneepan S, Tayar R. A double blind randomised placebo controlled pilot study of Septrin in the treatment of cryptogenic fibrosing alveolitis. *Thorax* 2002;**57**:S49. (Reason for exclusion: participants.)

Varney VA, Parnell HM, Salisbury DT, Ratnatheepan S, Tayar RB. A double blind randomised placebo controlled pilot study of oral co-trimoxazole in advanced fibrotic lung disease. *Pulm Pharmacol Ther* 2008;**21**:178–87. (Reason for exclusion: participants.)

Studies excluded owing to interventions following advisory group feedback

Alton EW, Johnson M, Turner-Warwick M. Advanced cryptogenic fibrosing alveolitis: preliminary report on treatment with cyclosporin A. *Respir Med* 1989;**83**:277–9.

Antoniou KM, Nicholson AG, Dimadi M, Malagari K, Latsi P, Rapti A, *et al.* Long-term clinical effects of interferon gamma-1b and colchicine in idiopathic pulmonary fibrosis. *Eur Respir J* 2006;**28**:496–504.

Behr J, Million-Rousseau R, Morganti A, Perchenet L, Raghu G. Efficacy and safety of macitentan in idiopathic pulmonary fibrosis (IPF) [abstract]. European Respiratory Society Annual Congress, Vienna, Austria, 1–5 September 2012;**40**:309s.

Bradford WZ, Starko K, Noble PW. Hospitalization of patients with idiopathic pulmonary fibrosis (IPF) in a phase 3, randomized, double-blind, placebo-controlled trial of interferon gamma-1B (IFN-gamma 1B). *Chest* 2003;**124**:193S.

Daniels CE, Lasky JA, Limper AH, Mieras K, Gabor E, Schroeder DR, *et al.* Imatinib treatment for idiopathic pulmonary fibrosis: Randomized placebo-controlled trial results. *Am J Respir Crit Care Med* 2010;**181**:604–10.

Douglas WW, Ryu JH, Swensen SJ, Offord KP, Schroeder DR, Caron GM, *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* 1998;**158**:220–5.

Fiorucci F, Lucantoni G, Paone G, Zotti M, Li BE, Serpilli M, *et al.* Colchicine, cyclophosphamide and prednisone in the treatment of mild-moderate idiopathic pulmonary fibrosis: comparison of three currently available therapeutic regimens. *Eur Rev Med Pharmacol Sci* 2008;**12**:105–11.

Gulsvik A, Kjelsberg F, Bergmann A, Froland SS, Rootwelt K, Vale JR. High-dose intravenous methylprednisolone pulse therapy as initial treatment in cryptogenic fibrosing alveolitis. A pilot study. *Respiration* 1986;**50**:252–7.

Jackson RM, Glassberg MK, Ramos CF, Bejarano PA, Butrous G, Gomez-Marin O. Sildenafil therapy and exercise tolerance in idiopathic pulmonary fibrosis. *Lung* 2010;**188**:115–23.

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* 1989;**44**:280–8.

Keogh BA, Bernardo J, Hunninghake GW, Line BR, Price DL, Crystal RG. Effect of intermittent high dose parenteral corticosteroids on the alveolitis of idiopathic pulmonary fibrosis. *Am Rev Respir Dis* 1983;**127**:18–22.

King J, Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA, *et al.* BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008;**177**:75–81.

King J, Brown KK, Raghu G, du Bois RM, Lynch DA, Martinez F, *et al.* BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;**184**:92–9.

King J, Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L, *et al.* Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* 2009;**374**:222–8.

Krowka MJ, Ahmad S, de Andrade JA, Frost A, Glassberg MK, Lancaster LH, *et al.* A randomized, double-blind, placebo-controlled study to evaluate the safety and efficacy of iloprost inhalation in adults with abnormal pulmonary arterial pressure and exercise limitation associated with idiopathic pulmonary fibrosis. *Chest* 2007;**132**:633S.

Kubo H, Nakayama K, Yanai M, Suzuki T, Yamaya M, Watanabe M, *et al.* Anticoagulant therapy for idiopathic pulmonary fibrosis. *Chest* 2005;**128**:1475–82.

Malouf MA, Hopkins P, Snell G, Glanville AR, Everolimus in IPF Study Investigators. An investigator-driven study of everolimus in surgical lung biopsy confirmed idiopathic pulmonary fibrosis. *Respirology* 2011;**16**:776–83.

Noth I, Anstrom KJ, Calvert B, de Andrade J, Flaherty KR, Glazer C, *et al.* A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012;**186**:88–95.

Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, *et al.* A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2004;**350**:125–33.

Raghu G, Lasky JA, Costabel U, Brown KK, Cottin V, Thomeer M, *et al.* A randomized placebo controlled trial assessing the efficacy and safety of etanercept in patients with idiopathic pulmonary fibrosis (IPF). *Chest* 2005;**128**:496S.

Raghu G, Behr J, Brown KK, Egan J, Kawut SM, Flaherty KR, *et al.* ARTEMIS-IPF: a placebo-controlled trial of ambrisentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012;**185**:A3632.

Raghu G, Million-Rousseau R, Morganti A, Perchenet L, Behr J. Efficacy and safety of macitentan in idiopathic pulmonary fibrosis: results of a prospective, randomized, double-blind, placebo-controlled trial. *Am J Respir Crit Care Med* 2012;**185**:A3631.

Raghu G, Behr J, Brown KK, Egan JJ, Kawut SM, Flaherty KR, *et al.* Treatment of idiopathic pulmonary fibrosis with ambrisentan: a parallel, randomized trial. *Ann Intern Med* 2013;**158**:641–9.

Raghu G, Brown KK, Costabel U, Cottin V, du Bois RM, Lasky JA *et al.* Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. *Am J Respir Crit Care Med* 2008;**178**:948–55.

Roig V, Herrero A, Arroyo-Cozar M, Vielba D, Juarros S, Macias E. [Comparative study between oral azathioprine and intravenous cyclophosphamide pulses in the treatment of idiopathic pulmonary fibrosis.] *Archivos de Bronconeumologia* 2010;**46**:15–19.

Selman M, Carrillo G, Salas J, Padilla RP, Perez-Chavira R, Sansores R, *et al.* Colchicine, D-penicillamine, and prednisone in the treatment of idiopathic pulmonary fibrosis: a controlled clinical trial. *Chest* 1998;**114**:507–12.

Strieter RM, Starko KM, Enelow RI, Noth I, Valentine VG, Idiopathic Pulmonary Fibrosis Biomarkers Study Group. Effects of interferon-gamma 1b on biomarker expression in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2004;**170**:133–40.

Tomioka H, Kuwata Y, Imanaka K, Hashimoto K, Ohnishi H, Tada K, *et al.* A pilot study of aerosolized *N*-acetylcysteine for idiopathic pulmonary fibrosis. *Respirology* 2005;**10**:449–55.

Tzortzaki EG, Antoniou KM, Zervou MI, Lambiri I, Koutsopoulos A, Tzanakis N, *et al.* Effects of antifibrotic agents on TGF-beta1, CTGF and IFN-gamma expression in patients with idiopathic pulmonary fibrosis. *Respir Med* 2007;**101**:1821–9.

Xifteri A, Drakopanagiotakis F, Tzouvelekis A, Tsiambas E, Karameris A, Tsakanika K, *et al.* Expression of CTGF and TNFa in alveolar macrophages of patients with idiopathic pulmonary fibrosis before and after treatment with azathioprine or interferon-1b. *Pneumon* 2011;**24**:149–56.

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* 1999;**341**:1264–9.

Appendix 5 Data extraction tables in alphabetical order

Azuma *et al.*[66]

Study details	Participant details
Azuma *et al.* 2005[66] *Country:* Japan *Design:* RCT *Number of centres:* 25 *Funding:* Shionogi & Co., Ltd (licence holder in Japan for pirfenidone)	*Number of participants:* 109: pirfenidone $n = 73$; placebo $n = 36$ *Sample attrition/dropout*: one from each group was excluded for violation of the inclusion criteria *Sample crossovers:* acute exacerbation rates in the placebo arm at 6 months meant that the safety board recommended participants be given pirfenidone. Unclear if this happened; as discussed below all participants were reported to continue with allocated treatments until 9 months *Inclusion criteria:* histological evidence of UIP not mandatory, HRCT evidence of definite or probable UIP required. Definite UIP: basal predominant, subpleural reticular abnormality with traction bronchiectasis and honeycomb cysts without atypical features of UIP. Probable UIP: same pattern as definite UIP but without traction bronchiectasis. Presence of bibasilar inspiratory crackles, abnormal PFTs, and increased serum levels of damaged-pneumocyte markers (KL-6 and surfactant proteins A and D). 20–75 years of age with adequate oxygenation at rest ($PaO_2 \geq 70$ mmHg) and demonstrated SpO_2 of $\leq 90\%$ during exertion while breathing air, within 1 month before enrolment *Exclusion criteria:* a decrease in symptoms during the preceding 6 months; use of immunosuppressive and/or oral prednisolone > 10 mg/day during the preceding 3 months; clinical suspicion of IIP other than IPF; coexisting emphysema (HRCT images of low attenuated areas in upper lung fields), pulmonary hypertension, asthma, tuberculosis, sarcoidosis, bronchiectasis other than traction associated, aspergillosis or respiratory infection; uncontrolled diabetes; comorbid conditions including malignancy; severe hepatic, renal or cardiac disease; pregnancy (or its pursuance); breastfeeding; previous use of pirfenidone; suspicion of poor compliance in adherence to the protocol; or being unable to understand protocol/written informed consent

Intervention details	Outcomes
Intervention 1. Pirfenidone 1800 mg (600 mg three times daily) 2. Placebo Dose details: A dose-titration schedule was used: oral tables at a dose of 200 mg three times daily for 2 days, 400 mg three times daily for 2 days, 600 mg three times daily (maximum dose) for 3 days. This was maintained in patients tolerating it during the study Dose modifications: patients not tolerating the maximum dose of 1800 mg/day received a reduced, prespecified regimen utilising the standards for Classification of Serious Adverse Drug Reactions. An adverse event of grade 2 or worse, reduced	Primary outcomes: change in the lowest SpO_2 during the 6MET Secondary outcomes: resting PFTs while breathing air (VC, TLC, DL_{CO} PaO_2), disease progression by HRCT patterns, episodes of acute exacerbation of IPF, change in serum markers of pneumocyte damage, changes in QoL measurements. Also reports serum KL-6, surfactant protein-D levels (not data extracted) Method of assessing outcome: Lowest SpO_2 during 6MET: walk on a treadmill at a constant speed while breathing air. The SpO_2 was continuously measured using a pulse oximeter. The initial speed was started at 60 m/minute and the investigator, while monitoring the SpO_2, adjusted the treadmill speed to determine an appropriate speed (40–80 m/minute) on the basis of the patient's comfort to perform the test while the lowest SpO_2 reached 90% or below. This speed was then kept constant in the individual patient at each follow-up visit. The test was stopped if the SpO_2 reached 80%

Study details	Participant details
from nine tablets per day to six tablets/day for 14 days; if persisted or increased reduced to 3 tablets/day for 14 days. If persisted or increased study medication was discontinued for 14 days. If resolved, the dose could be increased up to nine tablets/day again. For adverse event grade 1, dosage reduction was deferred to the investigator's clinical judgement	HRCT scans were performed in accordance with predetermined protocol at baseline and 6-month intervals. Three expert radiologists independently evaluated the pattern of lung fibrosis. Where there was disagreement (8%), the scans were re-examined by the radiologists and the Study Co-ordinating Committee to reach a consensus. Disease worsening was defined as progression in the extent of UIP pattern compared with baseline
Concurrent treatment: prednisolone ≤ 10 mg/day was allowed. Cyclophosphamide, azathioprine, methotrexate, d-penicallimine, colchicine, erythromycin, interferons, NAC, cyclosporine, tacrolimus and other experimental agents under investigation for IPF were not allowed	Acute exacerbation was defined as all of worsening; otherwise unexplained clinical features within 1 month; progression of dyspnoea over a few days to less than 5 weeks; new radiographic/HRCT parenchymal abnormalities without pneumothorax or pleural effusion (e.g. new, superimposed ground-glass opacities); a decrease in the PaO_2 by ≥ 10 mmHg; exclusion of apparent infection based on absence of antibodies in blood, urine and sputum cultures
Duration of treatment: 9 months	QoL was measured using the Chronic Respiratory Disease Questionnaire score and the Hugh-Jones Classification score. References provided
	Length of follow-up: minimum of 9 months. A decision based on analyses at 6 months recommended early termination of the study; however, owing to time taken to collect, analyse and report data at 6 months, all participants actually completed 9 months of treatment

Participant characteristics	Pirfenidone, $n = 72$	Placebo, $n = 35$	*p*-value
Sex (M/F), %	62/10 (86/14)	33/2 (94/6)	0.3295
Mean (SD) age, years	64.0 (7.1)	64.3 (7.6)	0.8735
Smoking history, %			
Current	7 (10)	3 (9)	0.6403
Former	57 (79)	30 (86)	
Never	8 (11)	2 (6)	
Years since diagnosis, %			
< 1 year	20 (28)	6 (17)	0.3830
1–3 years	17 (24)	10 (29)	
> 3 years	35 (49)	19 (54)	
Surgical lung biopsy, %	15 (21)	8 (23)	0.8066
Prior treatment with oral corticosteroids, %			
No	62 (86)	30 (86)	1.0000
Yes	10 (14)	5 (14)	
Mean (SD) VC, % predicted	81.6 (20.3)	78.4 (17.2)	0.3978
Mean (SD) TLC, % predicted	78.5 (17.9)	73.9 (16.4)	0.1935
Mean (SD) DL_{CO}, % predicted	57.6 (17.2)	57.7 (13.8)	0.9697
Mean (SD) PaO_2 at rest, mmHg	80.3 (7.7)	82.0 (7.6)	0.3063
Mean (SD) lowest SpO_2 during 6MET, %	87.1 (3.9)	87.1 (4.2)	0.9891
Mean (SD) log SpO_2 area	8.2 (0.3)	8.2 (0.4)	0.8462
Mean (SD) KL-6, U/ml	1427.5 (997.7)	1335.3 (863.3)	0.6243
Mean (SD) SP-D, ng/ml	264.6 (206.0)	233.6 (137.1)	0.3569
Comments: rates of definite/probable UIP on HRCT in the total group were 91/16			

Results	Pirfenidone, $n = 72$	Placebo, $n = 35$	Difference, p-value
Mean (SD) change lowest SpO_2 during 6MET, % at 6 months	0.6364 (3.5502)	−0.5484 (3.7933)	0.1489
Mean (SD) change lowest SpO_2 during 6MET, % at 9 months	0.4697 (3.8838)	−0.9355 (3.3559)	0.0722
Categorical analysis, 9 months	$n = 66$	$n = 33$	
Improved	16 (24%)	2 (6%)	0.0159
Stable	38 (58%)	20 (61%)	
Deteriorated	12 (18%)	11 (33%)	

Comments: a > 4% increase in the lowest SpO_2 during the 6MET was classified as 'improved', a ≤ 4% decrease as 'deteriorated' and the other values as 'stable'

Results	Pirfenidone, $n = 72$	Placebo, $n = 35$	Difference, p-value
Mean (SD) change VC, l, at 6 months	−0.01 (0.21)	−0.08 (0.19)	0.0995
Mean (SD) change VC, l, at 9 months	−0.03 (0.22)	−0.13 (0.19)	0.0366
Categorical analysis, 9 months	$n = 67$	$n = 33$	
Improved	6 (9%)	0	0.0028
Stable	52 (78%)	21 (64%)	
Deteriorated	9 (13%)	12 (36%)	

Comments: classified as 'improved', 'stable', 'deteriorated' (changes of 10%)

Results	Pirfenidone, $n = 72$	Placebo, $n = 35$	Difference, p-value
Mean (SD) change TLC, l, at 6 months	−0.02 (0.34)	0.00 (0.35)	0.7550
Mean (SD) change TLC, l, at 9 months	−0.05 (0.39)	−0.09 (0.45)	0.6154
Categorical analysis, 9 months	$n = 64$	$n = 31$	
Improved	6 (9%)	2 (6%)	0.0155
Stable	49 (77%)	17 (66%)	
Deteriorated	9 (14%)	12 (39%)	

Comments: classified as 'improved', 'stable', 'deteriorated' (changes of 10%)

Results	Pirfenidone, $n = 72$	Placebo, $n = 35$	Difference, p-value
Mean (SD) change DL_{CO} ml/minute/mmHg at 6 months	−0.50 (2.07)	−0.83 (2.16)	0.4894
Mean (SD) change DL_{CO} ml/minute/mmHg at 9 months	−0.57 (2.15)	−1.19 (2.30)	0.2120
Categorical analysis, 9 months	$n = 72$	$n = 32$	
Improved	10 (16%)	2 (6%)	0.1580
Stable	33 (50%)	15 (47%)	
Deteriorated	29 (36%)	15 (47%)	

Comments: classified as 'improved', 'stable', 'deteriorated' (changes of 15%)

Results	Pirfenidone, $n = 72$	Placebo, $n = 35$	Difference, p-value
Mean (SD) change resting PaO_2 mmHg at 6 months	−2.09 (9.71)	−3.19 (10.97)	0.6171
Mean (SD) change resting PaO_2 mmHg at 9 months	−2.48 (10.30)	−3.66 (10.43)	0.5981
Categorical analysis, 9 months	$n = 69$	$n = 34$	
Improved	18 (26%)	9 (26%)	0.2724
Stable	20 (29%)	4 (12%)	
Deteriorated	31 (45%)	21 (62%)	

Comments: classified as 'improved', 'stable', 'deteriorated' (changes of 4 mmHg)

Results	Pirfenidone, *n* = 72	Placebo, *n* = 35	Difference, *p*-value
HRCT 'improved'	10/65 (15%)	2/29 (7%)	0.0921

Comments: disease worsening on HRCT defined as progression in the extent of UIP pattern compared with baseline. Reduction of ground-glass and reticular opacities was recognised as improved patterns on HRCT

Acute exacerbation	0	14% (5/35)	0.0031
Dyspnoea	Data not reported	Data not reported	0.6367
QoL	Data not reported	Data not reported	0.8720

No significant changes in serum SP-D or KL-6 (data not reported)

Adverse events occurring in ≥ 10% at 6 months

	Pirfenidone, *n* = 72	Placebo, *n* = 35	Difference, *p*-value
Any	72 (98.6)	32 (88.9)	0.040
Photosensitivity	32 (43.8)	0	0.000
Stomach discomfort	22 (30.1)	3 (8.3)	0.0143
Anorexia	23 (31.5)	2 (5.6)	0.0030
Nausea	16 (21.9)	2 (5.6)	0.0314
Heartburn	12 (16.4)	1 (2.8)	0.0566
Drowsiness	17 (23.3)	6 (16.7)	0.4672
Fatigue	16 (21.9)	1 (2.8)	0.0102
URTI	12 (16.4)	3 (8.3)	0.3767
Fever	6 (8.2)	4 (11.1)	0.7271
Elevation of GOT	4 (5.5)	6 (16.7)	0.0785
Elevation of γ-GTP	20 (27.4)	3 (8.3)	0.0249
Urinary occult blood positive	6 (8.2)	4 (11.1)	0.7271
Elevation of C-reactive protein	15 (20.5)	10 (27.8)	0.4694

Reasons for discontinuation of medication at 9 months

	Pirfenidone, *n* = 72	Placebo, *n* = 35	Difference, *p*-value
Any adverse event	11 (15.1)	2 (5.6)	0.2132
Photosensitivity	5 (6.8)	0	0.1686
Vomiting	1 (1.4)	0	1.0000
Fever	1 (1.4)	0	1.0000
Abnormality of hepatic function	1 (1.4)	0	1.0000
Dizziness	1 (1.4)	0	1.0000
Facial paralysis	1 (1.4)	0	1.0000
Hepatoma	1 (1.4)	0	1.0000
Headache	0	1 (2.8)	0.3303
Bradycardia	0	1 (2.8)	0.3303
Acute exacerbation	0	5 (13.9)	0.0032
Patient request	3 (4.1)	0	0.5493
Progression of disease	1 (1.4)	1 (2.8)	1.0000
Protocol violation	1 (1.4)	0	1.0000

Methodology

Allocation to treatment groups: states randomised using a modified permuted-block randomisation method with block sizes of six

Blinding: states double blind. Outcome assessors for the HRCT scans were blinded to treatment assignment

Comparability of treatment groups: states baseline characteristics were similar, confirmed by *p*-values shown

Method of data analysis: change from baseline values used the Welch's *t*-test. Categorical variables were analysed with the Wilcoxon test. Analyses of incidences were performed with Fisher's exact test. For missing values the LOCF principle was used. A subgroup analysis of those able to complete the 6MET without the SpO_2 reaching < 80% was specified after initiation of the trial (but before breaking the code). This was because 27 participants had been unable to complete the 6MET at baseline

Sample size/power calculation: the prespecified sample size was 90 participants (pirfenidone 60; placebo 30). Based on a simulation study this minimum number provided statistical power > 0.8 to detect assumed efficacy at the significance level of 0.025

Attrition/drop-out: numbers and reasons provided (although numbers for categorical outcomes differ; unclear why)

General comments

Generalisability: severity of disease not stated; based on baseline characteristics participants would be considered to be mild IPF cases, although approximately half had been diagnosed for more than 3 years. States majority were corticosteroid naive

Outcome measures: appear valid; although QoL scales were not described limited data were reported on this outcome

Intercentre variability: not stated

Conflict of interests: all author conflicts are noted; all had received financial support from either Shionogi or InterMune. Data were held and analysed by trial sponsor. All authors participated fully in study design and had full access to analysed data. No restrictions were placed on authors for analyses and reporting

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Unclear
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in drop-outs between groups?	No
(ii) If so, were they explained or adjusted for?	
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis?	No
(ii) If so, was this defined?	
10. (i) Did the analysis account for missing data?	Yes
(ii) If so, were the methods appropriate?	Yes

γ-GTP, γ-guanosine triphosphate; 6MET, 6-minute exercise test; GOT, glutamic oxaloacetic transaminase; KL-6, serum levels of pneumocyte marker KL-6; PFT, pulmonary function test; SP-D, serum level of surfactant protein.

Demedts *et al.*[67]

Study details	Participant details
Demedts *et al.* 2005[67]	*Number of participants:* 182 randomised: NAC *n* = 92; placebo *n* = 90 (27 were excluded, leaving NAC *n* = 80, placebo *n* = 75)
Country: Belgium, France, Germany, Italy, Spain, the Netherlands	*Sample attrition/dropout:*
Design: RCT	Post randomisation, pre treatment:
Number of centres: 36	NAC *n* = 12 not included (UIP not confirmed, *n* = 10; withdrawal of consent, *n* = 2); placebo *n* = 15 not included (UIP not confirmed, *n* = 12; withdrawal of consent, *n* = 3)
Funding: sponsored by the Zambon group	Post treatment:
	NAC *n* = 16 (20%) withdrew (prohibited therapy, *n* = 3; withdrawn by investigator, *n* = 3; withdrew consent, *n* = 4; adverse events, *n* = 2; non-compliance, *n* = 1; other, *n* = 3); deaths *n* = 7 (disease progression, *n* = 3; respiratory tract infection, *n* = 3; heart failure, *n* = 1)
	Placebo *n* = 16 (21%) withdrew (prohibited therapy, *n* = 2; withdrawn by investigator, *n* = 2; withdrew consent, *n* = 4; adverse events, *n* = 2; non-compliance, *n* = 2; ineffective treatment or worsening of condition, *n* = 4); deaths *n* = 8 (disease progression, *n* = 4; respiratory tract infection, *n* = 1; cardiac arrest, *n* = 1; myocardial infarction, *n* = 1; cancer, *n* = 1)
	Sample crossovers: assume none from data presented
	Inclusion criteria: 18–75 years, histological or radiologic pattern typical of UIP after other causes of UIP were ruled out. HRCT suggestive or consistent with a probable diagnosis of UIP. In those younger than 50 years, open or thoracoscopic lung biopsy mandatory and showed a pattern of UIP; lung biopsy optional for older patients. In the absence of lung biopsy, a transbronchial biopsy was advocated to exclude alternative diagnoses. Bronchoalveloar lavage must also have been performed before inclusion and failed to show features supporting alternative diagnoses (diagnoses confirmed by independent committees of experts on the basis of published criteria). Duration of more than 3 months, bibasilar inspiratory crackles present, dyspnoea score at least 2 on a scale of 0 (minimum) to 20 (maximum), VC of no more than 80% predicted value or TLC < 90% predicted, and single-breath DL_{CO} < 80% predicted
	Exclusion criteria: standard regimen with prednisolone and azathioprine contraindicated or not justified for them, or if they presented with a known intolerance to NAC. Treatment with prednisolone of at least 0.5 mg/kg/day or azathioprine at least 2 mg/kg/day during the month before inclusion, or treatment with NAC of > 600 mg/day for more than 3 months in the previous 3 years. Concomitant or pre-existing diseases, abnormalities, or treatment at study entry or in the past with drugs (e.g. antioxidants and antifibrotics) that interfere with the diagnosis, severity, therapy or prognosis of IPF

Intervention details	Outcomes
Intervention 1. NAC 1800 mg (600 mg three times per day) 2. Placebo Dose details: NAC at 600 mg three times per day Dose modifications: states if standard therapy had to be adapted (e.g. for adverse events, poor compliance or clinical worsening), the participant was treated according to the preference of each centre, with a drug-exclusion criteria especially concerning antioxidants and antifibrotic drugs taken into consideration. Standard therapy was changed for 13 participants in the acetylcysteine arm and 11 in the placebo arm. Three in the acetylcysteine arm and 11 in the placebo arm started continuous O_2 therapy during the study period Concurrent treatment: both treatment and placebo arms also given prednisolone (starting dose, 0.5 mg/kg body weight per day; 0.4 mg/kg per day at month 2; 0.3 mg/kg at month 3; progressively reduced to 10 mg per day in months 4–6 and then maintained until month 12) and azathioprine (2 mg/kg per day) in addition to usual care Duration of treatment: not stated, assume 12 months	Primary outcomes: absolute changes in VC and DL_{CO} at 12 months Secondary outcomes: per cent predicted VC, per cent predicted DL_{CO}, alveolar volume change and per cent predicted CRP score, dyspnoea score, maximum exercise indexes [load (W_{max}) oxygen uptake (VO_{2max}) and ventilation (VE_{max}) not extracted], scores of ground-glass opacities and of fibrosis on HRCT (not extracted), SGRQ, adverse events, withdrawals and mortality. Also reports a post hoc categorical analysis of VC (not extracted) Although drug non-compliance not a stated outcome, paper states that it (defined by an intake of < 50% of the study medication) was determined by counting returned tablets Method of assessing outcomes: VC and DL_{CO} measured according to the ERS guidelines. CRP score and dyspnoea score: references provided. SGRQ: no details of validation or administration; for scoring see below. SGRQ administered every 6 months. Safety was continuously monitored until 1 month after the participant completed or withdrew from the study Length of follow-up: 12 months

Participant characteristics	NAC, $n = 80$	Placebo, $n = 75$
Sex (M/F), %	69/31	75/25
Mean (SD) age (years)	62 (9)	64 (9)
Smoking status, %		
Current	3.8	6.7
Former	57.5	62.7
Never	38.8	30.7
Mean (SD) months since diagnosis [median]	19.9 (28.3) [5.0]	18.9 (33.1) [3.0]
Diagnosed within previous 6 months, n (%)	39 (49)	42 (56)
Underwent surgical biopsy, n (%)	38 (48)	35 (47)
VC, n (%), with:		
> 60%	49 (61)	53 (71)
≤ 60%	31 (39)	22 (29)
TLC, volume, l, mean (SD)	3.72 (1.00)	3.72 (0.94)
TLC, % of predicted value, mean (SD)	62.1 (13.9)	61.6 (11.5)
$P_AO_2 - PaO_2$ at rest, mmHg, mean (SD)	31.6 (13.8)	30.1 (11.8)
PaO_2 at rest, mmHg, mean (SD)	70.1 (12.7)	72.0 (11.2)
Total SGRQ score	50 (18)	52 (16)
Receiving continuous O_2	8	2

Participant characteristics	NAC, $n = 80$	Placebo, $n = 75$
Previous treatments		
Prednisone	16	13
Azathioprine	8	5
Both	2	3
VC, mean (SD) litres	2.29 (0.68)	2.36 (0.74)
VC, per cent predicted (SD)	64.76 (15.41)	66.57 (14.42)
DL_{CO} mean (SD) mmol/minute/kPa	3.85 (1.41)	3.90 (1.39)
DL_{CO} per cent predicted (SD)	43.04 (13.10)	44.79 (15.15)
CRP total, mean (SD)	39.73 (15.01)	37.06 (15.88)
CRP without exercise, mean (SD)	32.88 (9.81)	31.69 (9.04)
Dyspnoea score	8.35 (4.44)	7.92 (3.99)

Results (all treated and LOCF data extracted where numbers are reported, otherwise only LOCF data)	Acetylcysteine, $n = 80$	Placebo, $n = 75$	Difference (95% CI); *p*-value
VC, mean (SD), l, at 12 months	$n = 55$	$n = 51$	NA see below
	2.31 (0.79)	2.26 (0.72)	
VC, mean (SD), l, at 12 months, LOCF	$n = 71$	$n = 68$	NA see below
	2.22 (0.77)	2.17 (0.71)	
VC, l, mean change at 12 months, LOCF (95% CI)	−0.06 (−0.14 to 0.02)	−0.19 (−0.29 to −0.09)	NA see below
VC, l, LS mean (SD), LOCF	$n = 71$	$n = 68$	0.18 (0.03 to 0.32); $p = 0.02$[a]
	2.27 (0.05)	2.10 (0.05)	
VC % predicted, mean (SD) at 12 months, LOCF	63.14 (19.98)	61.59 (15.17)	Not reported
VC % predicted, LS mean (SD) at 12 months, LOCF	65.13 (1.85)	60.34 (1.85)	4.79 (0.80 to 8.77); $p = 0.02$[b]
DL_{CO} mean (SD) mmol/minute/kPa at 12 months	$n = 48$	$n = 47$	NA see below
	4.20 (2.07)	3.46 (1.22)	
DL_{CO} mean (SD) mmol/minute/kPa at 12 months, LOCF	$n = 68$	$n = 63$	NA see below
	3.74 (1.99)	3.20 (1.26)	
DL_{CO} mmol/minute/kPa mean change at 12 months, LOCF (95% CI)	−0.11 (−0.47 to 0.25)	−0.70 (−0.95 to −0.45)	NA see below
DL_{CO} mmol/minute/kPa, LS mean (SD), LOCF	$n = 68$	$n = 63$	0.75 (0.27 to 1.23); $p = 0.003$[c]
	3.85 (0.17)	3.10 (0.18)	
DL_{CO} % predicted, mean at 12 months, LOCF	40.85 (14.85)	38.75 (14.75)	Not reported
DL_{CO} % predicted, LS mean at 12 months, LOCF	41.6 (1.35)	36.52 (1.45)	5.08 (1.17 to 8.99); $p = 0.01$[d]

Exercise

Reports maximum exercise load, maximum oxygen update and maximum exercise ventilation, but not extracted as not directly assessing exercise/function

Results (all treated and LOCF data extracted where numbers are reported, otherwise only LOCF data)	Acetylcysteine, $n = 80$	Placebo, $n = 75$	Difference (95% CI); p-value
CRP score at 12 months, LS mean (SD), LOCF	37.62 (1.75)	39.33 (1.70)	−1.71 (−8.72 to 5.30); $p = 0.70$[e]
CRP without exercise, at 12 months, LS mean (SD) LOCF	30.91 (0.84)	32.50 (0.87)	−1.59 (−4.37 to 1.19); $p = 0.17$[f]
Dyspnoea score at 12 months, LS mean (SD), LOCF	8.88 (0.49)	9.20 (0.51)	−0.32 (−1.72 to 1.09); $p = 0.65$[g]

SGRQ, no data reported

Compliance

Adverse events occurring in at least 5% of participants	Acetylcysteine, $n = 80$		Placebo group, $n = 75$		
	No. of events	No. of patients	No. of events	No. of patients	p-value
All adverse events	322	72 (90)	303	67 (89)	See below
Respiratory tract infection	22	20 (25)	27	24 (32)	
Dyspnoea	16	16 (20)	21	19 (25)	
Fever	17	15 (19)	10	10 (13)	
Liver-function test abnormal	15	14 (18)	13	11 (15)	
Cough	15	13 (16)	17	16 (21)	
Abdominal pain	12	12 (15)	7	7 (9)	
URTI	11	11 (14)	15	13 (17)	
Blood glucose ↑	9	9 (11)	12	11 (15)	
C-reactive protein ↑	7	6 (8)	3	3 (4)	
Blood alkaline phosphatase ↑	6	6 (8)	1	1 (1)	
Blood lactate dehydrogenase ↑	6	6 (8)	2	2 (3)	
Back pain	6	6 (8)	6	5 (7)	
Respiratory failure	5	5 (6)	1	1 (1)	
Bone marrow toxic effects	3	3 (4)	10	10 (13)	
Oedema	3	3 (4)	5	5 (7)	
Headache	4	3 (4)	6	6 (8)	
Asthenia	3	3 (4)	5	5 (7)	
Influenza-like illness	3	3 (4)	5	5 (7)	
Muscle cramp	1	1 (1)	4	4 (5)	

States none of the differences between the study groups were significant, except for those related to bone marrow toxicity, which occurred in 4% of participants receiving acetylcysteine (3/80) and in 13% receiving placebo (10/75); $p = 0.03$

Detailed descriptions of the types of events included in these categories stated in the paper

Deaths	7 (9%)	8 (11%)	$p = 0.69$

Methodology

Allocation to treatment groups: states participants randomly assigned to treatment with study medication, with a 1 : 1 ratio of acetylcysteine to placebo. The randomisation was performed centrally with the use of a computer-generated randomisation list stratified (in blocks of four) according to country and whether the VC was ≤ 60% of the predicted value or > 60% of the predicted value. Randomisation appears to have been undertaken prior to HRCT scans assessed for eligibility

Blinding: states is double blind, and placebos were matched. No details of blinding of outcome assessors

Comparability of treatment groups: states no significant differences in baseline characteristics were found between groups (the final randomised groups)

Method of data analysis: not all randomised were treated (see above). Analyses were based on data from all who were randomised, met the inclusion criteria, received the trial medication at least once, and underwent at least one baseline observation. Missing data were replaced by the LOCF method for all who underwent at least one lung-function measurement post baseline. For the VC analysis the LOCF included 71 and 68 participants receiving acetylcysteine and placebo, respectively, with data imputed for 20 and 16 participants, respectively. For the DL_{CO} analysis the LOCF included 68 and 63 participants receiving acetylcysteine and placebo, respectively, with data imputed for 16 and 17 participants, respectively. Used an ANCOVA with country and treatment group as fixed factors, country-by-treatment as an interaction, and baseline values as covariates. Modelled all possible combinations of cofactors to obtain the best model and used a likelihood-based method to test the robustness of the LOCF-ANCOVA analysis, with sensitivity analyses using the LOCF population and the baseline population, to test the robustness of the analysis. None of the combinations of cofactors included in the fixed-effects LOCF-ANCOVA analysis were statistically significant and therefore treatment comparisons were unadjusted. As multiple testing was undertaken the nominal p-value for defining statistical significance was reduced from $p < 0.05$ to $p < 0.025$

Sample size/power calculation: calculated to provide a power of 80% ($\alpha = 0.05$ by two-sided test) to detect a treatment difference between the two groups of 15% for VC and 20% for DL_{CO} after 1 year. On the basis of previous data and with an expected withdrawal rate of 25% including deaths, a total of 150 participants were to be enrolled. Participants randomised and included $n = 155$

Attrition/drop-out: numbers and reasons provided

General comments

Generalisability: severity not reported, estimated from baseline VC that participants were predominantly mild to moderate; however, 20–30% had VC < 60%, indicating more severe disease

Outcome measures: appear appropriate

Intercentre variability: not discussed

Conflict of interests: states that the sponsor held the data but placed no limitations on study design, data analysis or the content of the manuscript. Authors' individual declarations of interests are reported

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in drop-outs between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

ANCOVA, analysis of covariance; NA, not applicable.
a Also reports the relative difference of 9%, which is calculated as being $[(LS-LOCF_{acetylcsteine} - LS-LOCF_{placebo})/LS-LOCF_{placebo}] \times 100$.
b Also reports the relative difference of 8% (calculated as stated above).
c Also reports the relative difference of 24%, calculated as stated above.
d Also reports the relative difference of 14%, calculated as stated above.
e Also reports the relative difference of 5%, calculated as stated above.
f Also reports the relative difference of 5%, calculated as stated above.
g Also reports the relative difference of 4%, calculated as stated above.
Also reports baseline characteristics for the 27 participants who were excluded from the study; as these are not presented per assigned treatment group these are not data extracted here.
Also reports details and results of bronchoalveolar lavage, not extracted here.
CRP scores range from 0 to 100; CRP scores without exercise range from 0 to 70.
Dyspnoea scores range from 0 to 20.
SGRQ: total scores from 0 to 100, with higher scores indicating a worse QoL.
States that more than 85% of participants in both groups took on average >80% of the prescribed daily dose of the drug.
Table 3 in the paper also reports the DL_{CO}: alveolar volume as mmol/minute/kPa and per cent predicted but not data extracted.

Homma *et al.*[69]

Study details	Participant details
Homma *et al.* 2012[69]	*Number of participants*: 100. NAC: 51; control: 49
Country: Japan	*Sample attrition/dropout*: states that 90 were included in the full data set for the assessment of safety. Seventy-six were included in the data set for efficacy. Numbers and reasons for exclusion were stated in the text as:
Design: RCT	
Number of centres: 27	NAC group $n = 7$ (6 = non administration of NAC; 1 = missing data); control group $n = 3$ (1 = protocol violation; 2 = missing data)
Funding: grant from Ministry of Health, Labour and Welfare	Also presents that numbers and reasons for exclusion from the full analysis set were:
	NAC group $n = 6$ (one acute exacerbation, one pneumonia, four patient request, protocol violation and other)
	Control group $n = 8$ (four acute exacerbation, one progression of disease, three patient request, protocol violation and other)
	In total: 38 (74.5%) completed the study in the NAC group and 38 (77.6%) completed the study in the placebo group
	Sample crossovers:
	Inclusion criteria: patients with well-defined IPF, diagnosed by ATS/ERS guidance and 4th version of Japanese clinical diagnostic criteria for IIP. Histological evidence of UIP was not mandatory, but HRCT evidence was required (defined as basal predominant, subpleural reticular abnormality with traction bronchiectasis and honeycomb cysts without atypical features of UIP). The presence of other typical clinical features, including bibasilar inspiratory fine crackles, abnormal PFT and increased serum levels of markers of pneumocyte injury. Aged between 50 and 79 years, firm clinical and radiological diagnoses of IPF, severity of disease classified as stage I (partial O_2 concentration ≥ 80 mmHg at rest), or stage II (partial O_2 concentration 70–80 mmHg at rest), a lowest arterial oxygen saturation of $> 90\%$ during the 6MWT
	Exclusion criteria: an improvement in symptoms during the preceding 3 months; use of NAC, immunosuppressive agents, oral prednisolone or pirfenidone; clinical suspicion of IIP other than IPF

Intervention details	Outcomes
Intervention	*Primary outcomes:* absolute change in FVC at 48 weeks
1. NAC inhaled, 352.4 mg twice daily	*Secondary outcomes:* changes in lowest arterial O_2 saturation, 6MWT distance, PFT parameters (VC, % predicted VC, TLC, % predicted TLC, DL_{CO} and predicted DL_{CO}), serum markers of pneumocyte injury (not extracted); disease progression as determined by HRCT; subjective changes in symptoms such as dyspnoea, adverse events
2. Control	
Dose details:	*Method of assessing outcome*
NAC inhalation of 352.4 mg diluted with saline to a total volume of 4 ml, twice per day, using microair nebulisers and vibration mesh technology (NE-U22, Omron Healthcare, Tokyo), which improves lung deposition compared with jet nebuliser systems	HRCT in accordance to a predetermined protocol, with two expert radiologists evaluating lung fibrosis. Progression of disease was assessed on the basis of HRCT by consensus between the site investigator and one of the radiologists. Worsening was defined as progression in the extent of fibrosis and ground-glass opacity compared with baseline. Stable disease as no change. Improvement as decrease in the extent of ground-glass opacity compared with baseline
Control: states 'no treatment or placebo'	Dyspnoea by serum KL-6 and surfactant protein D levels and Fletcher, Hugh-Jones classification scores assessed patient dyspnoea during activities of daily living. Categorised as 'improved, stable, or deteriorated' with respect to changes of 20% for KL-6 and surfactant protein D, or one grade for the Fletcher, Hugh-Jones dyspnoea classification score
Dose modifications: not reported	
Concurrent treatment: not reported	Safety assessed using grading scale of the CTC for adverse events, v3.0
Duration of treatment: 48 weeks	*Length of follow-up:* 48 weeks

Participant characteristics	NAC, $n = 38$	Control, $n = 38$	*p*-value
Sex, M/F, *n* (%)	29/9 (76/24)	29/9 (76/24)	1.00
Age, years (mean (SD))	67.6 (6.4)	68.2 (7.7)	0.78
Smoking, *n* (%)			
Current	3 (7.9)	3 (7.9)	1.00
Former	25 (65.8)	26 (68.4)	
Never	10 (26.3)	9 (23.7)	
Years since diagnosis, *n* (%)			
< 1	9 (23.7)	5 (13.2)	0.09
1–3	17 (44.7)	14 (36.8)	
> 3	12 (31.6)	19 (50)	
Mean (SD) years	3.0 (3.4)	3.2 (2.5)	
No prior treatment with oral corticosteroids and NAC, *n* (%)	38 (100)	38 (100)	1.00
Disease severity stage, *n*			
I	30	31	1.00
II	8	7	
FVC % predicted, mean (SD)	89.2 (17.8)	88.7 (15.5)	0.56
VC % predicted, mean (SD)	90.4 (18.3)	89.1 (15.0)	0.95
TLC % predicted, mean (SD)	82.5 (17.4)	81.2 (13.3)	0.84
DL_{CO} % predicted, mean (SD)	72.3 (25.3)	64.4 (20.1)	0.16
Lowest SpO_2 during 6MWT, %, mean (SD)	93.1 (2.1)	92.4 (2.0)	0.14
KL-6, U/ml, mean (SD)	995.1 (440.0)	1246.8 (114.9)	0.78
SP-D, ng/ml, mean (SD)	179.6 (102.7)	203.4 (107.4)	0.40
Results	**NAC, $n = 38$**	**Control, $n = 38$**	**Difference, *p*-value**
Mean FVC (SD), l, at 48 weeks	2.67 (0.84)	2.51 (0.68)	Not stated
Mean change (SD) in FVC, ml	–90 ml (300)	–150 ml (200)	60 ml; 0.2661
Comments: text states difference was 63 ml, but data suggest difference was 60 ml			
Decline of > 10% FVC	Data not shown	Data not shown	–36.4% in NAC group
Decline of < 10% in FVC	Data not shown	Data not shown	+14.8% in NAC group; 0.42
Comments: limited data presented and uncertainty whether these were a priori analysed			
Lowest SpO_2 during 6MWT	Data not shown	Data not shown	States NS
6MWT distance	Data not shown	Data not shown	States NS
VC, l	Data not shown	Data not shown	States NS
VC % predicted	Data not shown	Data not shown	States NS
TLC	Data not shown	Data not shown	States NS
TLC % predicted	Data not shown	Data not shown	States NS
DL_{CO}	Data not shown	Data not shown	States NS
DL_{CO} % predicted	Data not shown	Data not shown	States NS
Serum markers of pneumocyte injury	Data not shown	Data not shown	States NS

Results	NAC, $n = 38$	Control, $n = 38$	Difference, p-value
Comments			
% improved or stable (by HRCT)	88.6% (31/35)	78.1% (25/32)	0.33
	3 improved, 28 stable	0 improved, 25 stable	
% with progression (by HRCT)	11.4% (4/35)	21.9% (7/32)	NR
Comments: note proportions are calculated on $n = 35$ (three missing data) for the NAC group and $n = 32$ (six missing data) in placebo group			
NR, not reported			
Also reports data at interim time points but not data extracted			
% improved or stable dyspnoea	86.8% (33/38)	84.2% (32/38)	1.00
	2 improved, 31 stable	1 improved, 31 stable	
% deteriorated	13.2% (5/38)	15.8% (6/38)	NR
Comments: no detail of actual scores on the dyspnoea scales used is reported. Also reports data at interim time points but not data extracted			
Frequent adverse events	$N = 44$	$N = 46$	
Bacterial pneumonia			
Grade 1	2	0	
Grade 2	2	0	
Grade 3	0	0	
Grade 4	0	0	
Cough			
Grade 1	1	0	
Grade 2	1	0	
Grade 3	0	0	
Grade 4	0	0	
Sore throat			
Grade 1	2	0	
Grade 2	0	0	
Grade 3	0	0	
Grade 4	0	0	
Hypercholesteraemia			
Grade 1	2	0	
Grade 2	0	0	
Grade 3	0	0	
Grade 4	0	0	
Comments: states there were no significant differences in the number of adverse events reported for the two groups. The severity of the events was < grade 2 for the whole NAC group. Not clear what definition of 'frequent' was applied			

Methodology

Allocation to treatment groups: randomly assigned to the two groups in a 1 : 1 ratio. No further details reported

Blinding: no blinding of participants, investigators or outcome assessors

Comparability of treatment groups: states that the baseline characteristics were similar across groups. However, baseline characteristics for those in the final data analysis only were presented

Method of data analysis: changes from baseline values and continuous variables were compared with Wilcoxon test. Categorical variables and incidence rates by Fisher's exact test. The principle of LOCF was adopted. Analysis of changes in FVC and other pulmonary function tests and serum levels of markers of interstitial pneumonia used ANCOVA using the respective baseline measurements as covariates. States that post hoc subgroup analyses were performed for those with initial FVC values < 95% of predicted or initial DL_{CO} values < 55% predicted. Data for these subgroups not extracted

Sample size/power calculation: not reported

Attrition/drop-out: numbers and reasons provided. Appears to be similar rate of attrition between the two study groups but numbers were quite high (24%)

General comments

Generalisability: described as mild to moderate severity. States early stage IPF based on criteria used in Japan, which ranges from stage I to stage IV, and based on partial arterial O_2 concentrations, with stage I being less severe to stage IV more severe, and no desaturation on the 6MWT distance. States that the severity classification is well correlated with survival in IPF patients. None had prior treatments

Outcome measures: appear valid

Intercentre variability: not reported

Conflict of interests: paper acknowledges two individuals who worked for Fulcrum Pharma KK and assisted with study management and data analysis but does not explicitly declare conflicts

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Unclear
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Unclear
4. Was the care provider blinded?	No
5. Was the patient blinded?	No
6. Were outcome assessors blinded to the treatment allocation?	No
7. (i) Were there any unexpected imbalances in drop-outs between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	No

NS, not significant; PFT, pulmonary function test.

Horton *et al.*[55,75]

Study details	Participant details
Horton *et al.* 2012[55,75]	*Number of participants:* 24 randomised: 11 to thalidomide first, 13 to placebo first according to flowchart (but states 12 and 12 in text)
Country: USA	
Design: randomised crossover trial	*Sample attrition/dropout:* 23 were treated, 20 completed both treatment periods. During first period, one withdrew from placebo group due to worsening health, and one withdrew from thalidomide before receiving treatment due to lack of interest. During second period, two withdrew from thalidomide group due to worsening health
Number of centres: one	
Funding: Celgene Corporation	
	Sample crossovers: none unplanned
	Inclusion criteria: aged > 50 years, clinical history consistent with IPF (symptom duration $\geq$ 3 months and $\leq$ 5 years) and chronic cough (defined as > 8 weeks' duration, that adversely affected QoL and not due to other identifiable causes). HRCT scans consistent with IPF or surgical lung biopsy demonstrating UIP, FVC between 40% and 90% predicted, TLC between 40% and 80% predicted, and DL_{CO} between 30% and 90% predicted
	Exclusion criteria: pregnancy, female with childbearing potential, toxic or environmental exposure to respiratory irritants, collagen vascular disease, airflow obstruction, active narcotic antitussive use, peripheral vascular disease or neuropathy, inability to give informed consent, allergy or intolerance to thalidomide, life expectancy < 6 months (opinion of investigators)

Intervention details	Outcomes
Intervention	Primary outcomes: cough-specific QoL
1. Thalidomide	Secondary outcomes: cough, respiratory QoL
2. Placebo	Method of assessing outcome:
Dose details: 50 mg by mouth at bedtime; dose increased to 100 mg if no improvement in cough occurred after 2 weeks (21 of 22 receiving thalidomide and 23 of 23 receiving placebo)	Cough-specific QoL measured by the CQLQ. Cough measured by a 10-cm VAS. Respiratory QoL measured by SGRQ
Dose modifications: not reported	Length of follow-up: 12 weeks
Concurrent treatment: 100 mg sodium docusate by mouth daily to avoid thalidomide-associated constipation, daily vitamin B complex supplement. Any prescription therapy for cough was discontinued 2 weeks before study and no new therapy for cough was started. No patients began benzonate therapy during trial or reported changes in ACE inhibitors, angiotensin-receptor blocker, gastro-oesophageal reflux disease or sinus therapies	
Duration of treatment: 12 weeks each treatment with a 2-week washout period between treatments	

Participant characteristics	n = 23
Mean age (SD)	67.6 (7.8)
Male/female (%)	18/5 (78/22%)
White (%)	21 (31%)
HRCT diagnosis IPF, n (%)	23 (100)
Surgical lung biopsy, n (%)	5 (21.7)
Mean time from diagnosis (range), months	20.5 (3–59)
Mean FVC (SD), % predicted	70.4 (13.7)
Mean FEV_1–FVC ratio (SD)	0.85 (0.54)
Mean TLC (SD), % predicted	63.6 (11.4)
Mean DL_{CO} (SD), % predicted	57.4 (14.4)
Previous IPF treatment, n (%)	15 (65)
NAC	12 (52)
Oxygen	5 (22)
Prednisone	3 (13)
Previous cough treatment, n (%)	8 (35)
Benzonatate	5 (22)
Narcotic	5 (22)
GERD, n (%)	12 (52)
Therapy for GERD reported at study entry	
Proton pump inhibitor	10 (43)
High-dose proton pump inhibitor	2 (9)
ACE inhibitor/ARB use, n (%)	7 (30)
Chronic sinusitis, n (%)	8 (34)
Therapy for chronic sinusitis reported at study entry	
Antihistamine	5 (22)
Nasal steroids	5 (22)
Decongestant	4 (17)
Leuokotriene	1 (4)
CQLQ, mean (SD)	60.5 (12.0)
Cough VAS, mean (SD)	64.8 (21.4)
SGRQ total, mean (SD)	57.4 (18.8)
SGRQ symptom domain, mean (SD)	67.7 (19.7)
SGRQ impact domain, mean (SD)	48.1 (20.7)
SGRQ activity domain, mean (SD)	64.3 (22.7)

Results	After 12 weeks of thalidomide, $n = 23$	After 12 weeks of placebo, $n = 23$	Mean difference (95% CI); p-value
CQLQ, mean (SD)	47.2 (13.4)	58.7 (14.0)	–11.4 (–15.7 to –7.0); < 0.001
Cough VAS, mean (SD)	32.2 (26.1)	61.9 (26.5)	–31.2 (–45.2 to –17.2); < 0.001)
SGRQ total, mean (SD)	43.9 (16.0)	56.9 (17.1)	–11.7 (–18.6 to –4.8); 0.001
SGRQ symptom domain, mean (SD)	50.3 (20.9)	62.0 (18.3)	–12.1 (–22.2 to –2.0); 0.018
SGRQ impact domain, mean (SD)	34.3 (16.1)	49.0 (19.4)	–13.1 (–19.7 to –6.6); 0.001
SGRQ activity domain, mean (SD)	60.9 (14.2)	65.8 (18.7)	–3.3 (–9.8 to 3.2); 0.31

Comments: in the mixed-effect linear regression model, the CQLQ score was 11.4 points lower with thalidomide than with placebo (95% CI –15.7 to –7.0), complete results of mixed-effects models reported but not extracted. Sensitivity analyses to explore the effect of missing data reported but not extracted; thalidomide use resulted in a statistically significant improvement in CQLQ scores in each of these sensitivity analyses. NB: different data shown in abstract[75]

CQLQ consists of 28 questions using Likert-like 4-point scales, with lower scores indicating less effect of cough on HRQoL. The MCID of the CQLQ in IPF is unknown, but the MCID for the Leicester Cough Questionnaire, which has been shown to be similar to the CQLQ, is 1.3

SRGQ has 50 items and produces three domain scores and one overall score, measuring, symptoms, activity and impacts

Adverse events, n (%)	Thalidomide, $n = 22$	Placebo, $n = 23$	
Participants with ≥ 1 adverse event	17 (77)	5 (22)	$p = 0.001$
Participants with a serious adverse event (influenza)	0	1 (4)	
Adverse events requiring dose reduction			
Constipation	2 (9)	0	
Bradycardia	1 (5)	0	
Adverse events requiring drug discontinuation (progressive illness or inability to travel for visits)	2 (9)	1 (4)	
Gastrointestinal adverse events			
Constipation	8 (36)	1 (4)	
Change in taste	2 (9)	0	
Dry mouth	2 (9)	0	
Anorexia	1 (5)	NR	
General adverse events:			
Dizziness	6 (27)	0	
Malaise	3 (14)	0	
Oedema	2 (9)	0	
Rash	2 (9)	0	
Sleepiness	NR	1 (4)	
Respiratory adverse event: worsening dyspnoea	2 (9)	2 (9)	
Infectious adverse event: viral URTI	5 (23)	1 (4)	
Cardiac adverse event: bradycardia	1 (5)	0	

Methodology

- Allocation to treatment groups: a randomisation schedule was prepared by a third party using a manual algorithm. A random seed number was generated in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA). The pharmacist dispensing the study drug was the only person who had access to the treatment assignment
- Blinding: study investigators and participants were blinded. However, it is noted that thalidomide has characteristic effects that may have alerted participants to the identity of the drug
- Comparability of treatment groups: not applicable (crossover trial)
- Method of data analysis: the VAS of cough was not normally distributed according to the Shaprio–Wilk test. Categorical variables were compared using the McNemar's chi-squared test. The association between thalidomide and the primary and secondary outcomes was analysed using linear mixed-effect models. For each outcome measure, only the value at week 12 of thalidomide or placebo was used. The model contained covariates for treatment, period and treatment sequence as fixed effects and participants nested within sequence as a random effect. Additional analyses were performed to test the sensitivity of findings to missing data. The primary analysis was repeated by replacing CQLQ scores with scores from day 0 and with the highest CQCL score. A model was also used that included a covariate for the CQLQ at day 0. Results are mean (SD) unless stated otherwise
- Sample size/power calculation: sample size calculations were based on the CQLQ. No data on the minimum clinically important difference in CQLQ or data on the variance in IPF patients were available. Therefore, calculations were based on the ability to recruit 20 participants, which would provide 80% power to detect a difference of 4.67 units with a two-sided alpha error level of 5% in a two–treatment crossover study, based on the assumption that the within-patient SD of the response variance would be 5.0
- Attrition/drop-out: numbers and reasons provided. Data were collected for all planned study visits from 20 participants and incomplete data on three participants who received placebo first but withdrew before completion of the thalidomide period. Analysis included all available data

General comments

- Generalisability: participants generally had mild IPF, based on the per cent predicted IPF at baseline, and cough. Most participants self-referred after learning about the study online, and therefore might not be representative
- Outcome measures: validated measures of QoL and cough. No lung function tests reported
- Intercentre variability: not applicable
- Conflict of interests: Celgene Corporation provided the study drug and funding by had no role in study design, conduct, analysis or manuscript preparation. Disclosures of authors can be viewed online

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors?	NA
4. Was the care provider blinded?	Yes
5. Was the patient blinded?	Yes (but treatment may have unmasked)
6. Were outcome assessors blinded to the treatment allocation?	Yes
7. (i) Were there any unexpected imbalances in drop-outs between groups? (ii) If so, were they explained or adjusted for?	Yes Yes
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes (but not described as such) Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes Yes

ACE, angiotensin converting enzyme; ARB, Angiotensin II Receptor Blocker; GERD, gastroesophageal reflux disease.

Additional quality/methodological questions for crossover trial

1. Is a crossover trial an appropriate design for the condition of interest? Yes (symptomatic treatment of cough).
2. Was the order of treatments randomised appropriately? Yes.
3. Is there likelihood of serious carry over? No (2-week washout period between treatments).
4. What data are available? (Sometimes if carryover is seen by the triallists, they may provide data for a 'parallel' group comparison. However, this brings its own risk of bias.) End of treatment mean and SD provided.
5. Was the correct analysis applied? Yes: linear mixed-effect models contained covariates for treatment, period and treatment sequence as fixed effects and participants nested within sequence as a random effect.
6. Were the dropout rates reported and are these likely to cause a bias? Reported and unlikely to cause bias.

Idiopathic Pulmonary Fibrosis Clinical Research Network 2012[68]

Study details	Participant details
Raghu *et al.* IPFCRN 2012[68] *Country:* USA *Design:* RCT (PANTHER study) *Number of centres:* 25 *Funding:* grants from the NHLBI and the Cowlin Family fund. NAC and placebo donated by Zambon (manufacturer)	*Number of participants*: 236 randomised: NAC alone, $n = 81$ (data not presented in paper); NAC/prednisolone/azathioprine, $n = 77$; placebo $n = 78$ *Sample attrition/dropout*: NAC/prednisolone/azathioprine: six discontinued (two underwent lung transplantation, two withdrew consent, two had consent withdrawn by physician); placebo: three discontinued (two withdrew consent, one had consent withdrawn by physician) Participants discontinued treatment with one or more of the active or placebo agents ($p < 0.001$ between groups for all comparisons): NAC/prednisolone/azathioprine: 23 discontinued prednisolone, 31 azathioprine, 24 NAC (20 discontinued all three drugs) Placebo: three discontinued prednisolone placebo, four azathioprine placebo, four NAC placebo (three discontinued all three drugs) *Sample crossovers:* not reported *Inclusion criteria:* IPF aged between 35 and 85 years with mild to moderate lung function impairment (FVC $\geq 50\%$ and $DL_{CO} \geq 30\%$ of predicted values), met the modified criteria of the ATS/ERS/JRS/LATA (2011) for diagnosis of IPF, had received diagnosis on the basis of HRCT or biopsy ≤ 48 months before enrolment *Exclusion criteria:* history of clinically significant environmental exposure known to cause pulmonary fibrosis (e.g. asbestos); diagnosis of connective tissue disease; extent of emphysema greater than the extent of fibrotic change; FEV_1–FVC ratio < 0.65 at screening; PaO_2 on room air < 55 mmHg; residual volume $> 120\%$ predicted at screening; evidence of active infection; significant bronchodilator response on screening spirometry; screening and enrolment post-bronchodilator FVC measurements (in litres) differing by $> 11\%$; listed for lung transplantation; history of unstable or deteriorating cardiac disease; myocardial infarction, coronary artery bypass, or angioplasty within 6 months of screening; unstable angina or congestive heart failure requiring hospitalisation within 6 months of screening; uncontrolled arrhythmia; severe uncontrolled hypertension; known HIV or hepatitis C; known cirrhosis and chronic active hepatitis; active substance and/or alcohol abuse; pregnancy or lactation; not using means of contraception (where appropriate); any clinically relevant lab abnormalities (specified); homozygous for low thiopurine *S*-methyltransferase; overt or persistent clinical depression (defined); known hypersensitivity to study medication

Intervention details	Outcomes
Intervention	Primary outcomes: change in FVC at 60 weeks
1. NAC/placebo prednisolone/placebo azathioprine (data not presented in article as 'ongoing')	Secondary outcomes: rate of death, time until death, frequency of acute exacerbation, frequency of maintained FVC response, time to disease progression, clinical and physiological measures including: DL_{CO}, 6MWT, CPI, UCSDSBQ, SGRQ, SF-36, EQ-5D. Adverse events. Also predefined subgroups (data not presented)
2. NAC/prednisolone/azathioprine	
3. Placebo	
Dose details: prednisolone commenced at 0.5 mg/kg of ideal body weight and tapered to 0.15 mg/kg during 25 weeks. Azathioprine (maximum 150 mg/day) was based on the patient's ideal weight, concurrent use of allopurinol, and TPMT activity. NAC prescribed at 600 mg orally three times per day	Method of assessing outcome: all deaths, hospitalisations and suspected acute exacerbations were reviewed by the IPF net adjudication committee. The definition of an acute exacerbation was prespecified in accordance with previously published criteria (ref provided)
Dose modifications: detailed algorithms were provided for dose adjustments in case of potential adverse events	Length of follow-up: 60 weeks in the planned analysis. The study was stopped prior to this, although the timing of this was not reported it was stated as 'mid point'. This is likely to be different for different participants
Concurrent treatment: not reported	
Duration of treatment: up to 60 weeks	

Participant characteristics [mean (SD) unless stated]	NAC/prednisolone/ azathioprine, $n = 77$	Placebo, $n = 78$
Age, years	68.8 (7.3)	67.9 (8.1)
Male/female (%)	59/18 (77/23)	57/21 (73/27)
Race or ethnic group, n (%)[a]		
White	75 (97)	75 (96)
Black	1 (1)	0
Hispanic	1 (1)	5 (6)
History of smoking, n (%)		
Current	3 (4)	4 (5)
Former	51 (66)	54 (69)
Never	23 (30)	20 (26)
Time since diagnosis, years	0.9 (1.1)	1.1 (1.0)
Coexisting illness, n (%)		
Coronary artery disease	13 (17)	17 (22)
Diabetes	11 (14)	14 (18)
Gastro-oesophageal reflux disease	48 (62)	45 (58)
FVC, %	69.3 (15.1)	72.1 (14.1)
DL_{CO} corrected for haemoglobin, % predicted	42.1 (10.2)	45.3 (12.4)
PaO_2 while breathing air, mmHg	79.6 (9.7)	78.8 (12.6)
CPI[b]	53.7 (11.7)	49.8 (13.5)
6MWT distance, m	362.0 (113.0)	368.9 (117.3)
SBQ (UCSD)[c] score	30.1 (20.1)	29.1 (19.4)

Participant characteristics [mean (SD) unless stated]	NAC/prednisolone/ azathioprine, $n = 77$	Placebo, $n = 78$
SGRQ total score[d]	38.7 (17.4)	39.4 (17.4)
SGRQ symptoms score[d]	49.4 (21.1)	45.6 (21.8)
SGRQ activity score[d]	51.1 (19.0)	52.7 (21.0)
SGRQ impacts score[d]	27.8 (19.2)	28.8 (17.3)
SF-36 aggregate physical score[e]	40.3 (9.8)	40.6 (9.3)
SF-36 aggregate mental score[e]	53.9 (9.6)	55.7 (7.4)
EQ-5D score	0.8 (0.2)	0.8 (0.2)
EQ-5D thermometer response	76.8 (15.5)	78.1 (15.4)
HRCT definite IPF	66 (86%)	61 (78%)
Diagnosis based on surgical lung biopsy	38 (49%)	37 (47%)
CPI		

Results [assumed by reviewer as mean (95% CI)]	NAC/prednisolone/ azathioprine, $n = 77$	Placebo, $n = 78$	Difference, (95% CI); *p*-value
FVC change, l	–0.24 (–0.33 to –0.15)	–0.23 (–0.32 to –0.14)	–0.01 (–0.14 to 0.11); 0.85

Comments: FVC at 60 weeks could not be assessed because the study stopped early. Data above were at the time of the interim analysis

All-cause mortality, *n* (%)	8 (10)	1 (1)	0.01
Respiratory mortality	7 (9)	1 (1)	0.02
Estimated any-cause mortality at 60 weeks, %	19.8 (9.9 to 37.2)	2.0 (0.3 to 13.6)	HR 9.26 (1.16 to 74.1); 0.01
Estimated any-cause mortality or hospitalisation	43.6 (30.7 to 59.0)	16.9 (8.7 to 31.5)	HR 3.74 (1.68 to 8.34); < 0.001
Estimated any-cause mortality or ≥ 10% decline in FVC	36.3 (23.7 to 53.0)	32.4 (19.7 to 50.3)	HR 1.46, (0.70 to 3.05); 0.30

Comments: provides timelines until respiratory death during the study period, for each participant according to study group

All-cause hospitalisations	23 (30)	7 (9)	< 0.001

Comments

Time to death	Data in figure	Data in figure	NR
Time to death or disease progression	Data in figure	Data in figure	States NS

Comments: as defined by a composite outcome of death or a relative drop in FVC of > 10%

Time to death or hospitalisation	32%	3%	HR 12.11 (2.83 to 51.85); < 0.001

Comments

Results [assumed by reviewer as mean (95% CI)]	NAC/prednisolone/ azathioprine, $n = 77$	Placebo, $n = 78$	Difference, (95% CI); p-value
Acute exacerbation	5 (6%)	0	Not reported
Comments			
DL_{CO} corrected	–1.72 (–2.73 to –0.71)	–1.66 (–2.65 to –0.67)	–0.06 (–1.48 to 1.35); 0.93
Comments			
6MWT distance, m	–93.0 (–142.0 to –44.1)	–73.6 (–118.4 to –28.7)	–19.5 (–85.9 to 46.9), 0.56
Comments: also reports 6MWT oxygen desaturation area under the curve, 6MWT distance to saturation < 80%, 6MWT minutes walked. Not data extracted here			
CPI	6.72 (3.61 to 9.83)	5.33 (2.37 to 8.29)	1.39 (–2.90 to 5.68); 0.52
Comments			
USCDSBQ	10.6 (3.60 to 17.6)	8.01 (1.67 to 14.3)	2.57 (–6.87 to 12.0); 0.59
Comments			
SGRQ			
Total score	4.29 (–1.14 to 9.73)	7.50 (2.57 to 12.4)	–3.20 (–10.5 to 4.13); 0.39
Symptoms score	–4.42 (–11.9 to 3.1)	8.31 (1.47 to 15.2)	–12.7 (–22.9 to –2.61); 0.014
Activity score	7.33 (1.05 to 13.6)	10.3 (4.66 to 16.0)	–2.99 (–11.4 to 5.46); 0.49
Impacts score	5.23 (–0.80 to 11.3)	5.80 (0.34 to 11.27)	–0.57 (–8.71 to 7.57); 0.89
Comments			
SF-36 aggregate physical score	–4.18 (–7.40 to –0.97)	–2.96 (–5.90 to –0.02)	–1.23 (–5.58 to 3.13); 0.58
Comments: range 0–100, with higher scores indicating better health			
SF-36 aggregate mental score	0.96 (–2.51 to 4.44)	–4.35 (–7.50 to -1.20)	5.31 (0.62 to 10.00); 0.027
Comments			
EQ-5D score	–0.07 (–0.14 to 0.00)	–0.02 (–0.09 to 0.04)	–0.05 (–0.14 to 0.05); 0.31
Comments			
EQ-5D thermometer response	–6.81 (–13.0 to –0.67)	–6.66 (–12.4 to –0.94)	–0.15 (–8.54 to 8.24); 0.93
Comments			
Serious adverse event rate			
Any	24 (31)	8 (10)	0.001
Respiratory system	12 (16)	4 (5)	0.03
Infectious	5 (6)	1 (1)	0.12
Gastrointestinal	1 (1)	3 (4)	0.62
Cardiac	3 (4)	0	0.12
General disorder[f]	3 (4)	0	0.12

Results [assumed by reviewer as mean (95% CI)]	NAC/prednisolone/ azathioprine, *n* = 77	Placebo, *n* = 78	Difference, (95% CI); *p*-value
Neoplasm	2 (3)	0	0.25
Metabolism	1 (1)	0	0.50
Musculoskeletal system	0	1 (1)	1.00
Nervous system	1 (1)	0	0.50
Reproductive system	1 (1)	0	0.50
Adverse events[g]			
Any	68 (88)	61 (78)	0.09
General disorder	34 (44)	21 (27)	0.03
Skin	13 (17)	4 (5)	0.02
Renal and urinary system	10 (13)	1 (1)	0.005
Non-serious adverse events[h]			
Respiratory	35 (45.5)	34 (43.6)	0.82
Infections	37 (48.1)	28 (35.9)	0.13
Gastrointestinal	29 (37.7)	27 (34.6)	0.69
Investigations	20 (26.0)	19 (24.4)	0.81
Nervous system	18 (23.4)	12 (15.4)	0.21
Musculoskeletal	10 (13.0)	16 (20.5)	0.21
Metabolism	14 (18.2)	7 (9.0)	0.09
Psychiatric	6 (7.8)	7 (9.0)	0.79
Injury	6 (7.8)	4 (5.1)	0.53
Eye	6 (7.8)	3 (3.8)	0.33
Blood and lymphatic	5 (6.5)	2 (2.6)	0.28
Cardiac	4 (5.2)	3 (3.8)	0.72
Vascular	5 (6.5)	2 (2.6)	0.28
Ear	4 (5.2)	2 (2.6)	0.44
Immune	2 (2.6)	3 (3.8)	1.00
Neoplasms	4 (5.2)	1 (1.3)	0.21
Reproductive	2 (2.6)	3 (3.8)	1.00
Congenital or genetic	0	1 (1.3)	1.00
Endocrine	0	1 (1.3)	1.00
Social circumstances	1 (1.3)	0	0.50

Methodology

Allocation to treatment groups: A permuted-block randomisation with varying block sizes stratified according to clinical centre was used. Participants were randomised in a 1 : 1 : 1 ratio through telephone contact with a central interactive voice-response system

Blinding: states double blind and that placebo treatments were matched; no further details reported

Comparability of treatment groups: States the two groups were well matched with respect to demographic and clinical characteristics

Results [assumed by reviewer as mean (95% CI)]	NAC/prednisolone/ azathioprine, $n = 77$	Placebo, $n = 78$	Difference, (95% CI); p-value

Method of data analysis: all analyses based on ITT principle. A two-step process was used; if any difference between study groups was statistically significant at 0.05 level, then each of the three pairwise comparisons would be tested at the 0.05 level. For continuous outcomes, a mixed-model repeated-measures analysis was used to compare FVC in the three study groups. Variables included in the regression model included study group, time, time according to treatment, age, sex, race and height. Dichotomous outcomes were compared with two-sided Fisher's exact tests. Time-to-event outcomes were analysed with a Cox proportional-hazards regression model. Kaplan–Meier curves were used to display event rates. One interim analysis for efficacy was planned, where a Bonferroni approximation was applied, with the critical value set to an alpha level of 0.00001. Predefined subgroups were noted. At the planned interim analysis, the data and safety monitoring board recommended discontinuation of the three-drug regimen because of an excess number of deaths, hospitalisations and serious adverse events compared with the placebo group

Sample size/power calculation: based on previous clinical trials, an estimate of a 0.20 l decline in FVC in the placebo group, and a difference of 0.15 l from that in the placebo group as clinically meaningful, was used. After accounting for potential dropout and imperfect compliance, 130 participants per group were determined as providing a power of 90% for the first step of the testing procedure under most scenarios. The target sample size was expected to provide a power of approximately 93% to detect a significant difference at the two-sided 0.05 level

Attrition/dropout: numbers and reasons provided

General comments

Generalisability: described as mild to moderate impairment in pulmonary function

Outcome measures: appear valid

Intercentre variability: not reported

Conflict of interests: disclosure forms provided by all authors (available online)

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	Yes
	Yes
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

CPI, Composite Physiologic Index; JRS, Japanese Respiratory Society; LATA, Latin American Thoracic Association; NHLBI, National Heart, Lung and Blood Institute.
a Participants could select more than one category.
b No official upper or lower limit, higher scores indicate more severe disease.
c Scores range from 0 to 120, with higher scores indicating worse function.
d Scores range from 0 to 100, with higher scores indicating better function.
e Scores range from 0 to 100, with higher scores indicating better function.
f Includes all serious adverse events that did not fall into another body-system category, including drug reactions and drug fever.
g Specific adverse events with a significant between-group difference.
h All other adverse events recorded.

Idiopathic Pulmonary Fibrosis Clinical Research Network 2010[72]

Study details	Participant details
Zisman *et al.* IPFCRN 2010[72] *Country:* USA *Design:* RCT *Number of centres:* 14 *Funding:* NHLBI grants; the Cowlin Fund at the Chicago Community trust; Pfizer; Masimo	*Number of participants:* 180 randomised: sildenafil $n = 89$, placebo $n = 91$ *Sample attrition/dropout:* Sildenafil: eight participants dropped out of the study (four adverse events, two deaths, two lost to follow-up); placebo: six dropped out of the study (four adverse events, one death, one underwent lung transplantation) Dropout rates for open-label period also presented but not extracted here *Sample crossovers:* not stated but assume none. All placebo participants received sildenafil after 12 weeks (data for this period not extracted) *Inclusion criteria:* diagnosis of IPF as defined by the 2000 ATS/ERS consensus criteria, in an advanced stage (a DL_{CO} of < 35% predicted) *Exclusion criteria:* 6MWT distance of < 50 m (164 ft); a difference of > 15% in the 6MWT distance between two prerandomisation walks; an extent of emphysema greater than the extent of fibrotic change (by HRCT); treatment with medications containing nitrates; the presence of aortic stenosis or idiopathic hypertrophic subaortic stenosis; the initiation of pulmonary rehabilitation within 30 days after screening; the initiation or change in the dose of any investigational treatment for IPF within 30 days after screening; treatments for pulmonary hypertension with prostaglandins, endothelin-1 antagonists, or other phosphodiesterase inhibitors within 30 days after screening; a resting oxygen saturation of < 92% while breathing 6 l of supplemental oxygen; being listed on an active waiting list for lung transplantation

Intervention details	Outcomes
Intervention 1. Sildenafil 2. Placebo Dose details: sildenafil, oral, 20 mg, three times daily Dose modifications: not reported Concurrent treatment: not reported Duration of treatment: 12 weeks during a double-blind placebo controlled; 12 weeks open-label extension where all received sildenafil	Primary outcomes: presence or absence of an improvement of at least 20% in the 6MWT distance at 12 weeks Secondary outcomes: changes in the 6MWT distance, degree of dyspnoea, QoL, FVC, DL_{CO}, arterial partial pressure of oxygen and arterial oxygen saturation, and the alveolar–arterial oxygen gradient while breathing ambient air, adverse events, hospitalisations, death Method of assessing outcomes: dyspnoea with the UCSDSBQ and the Borg Dyspnoea Index. QoL was measured by the SGRQ, SF-36 and EQ-5D. Each suspected acute exacerbation was adjudicated by a central committee in a blinded fashion. 6MWT was performed with the use of a standardised protocol. Those with pulse O_2 saturation of ≥ 88% at rest were tested breathing ambient air; those below this received supplemental O_2, titrated to an O_2 saturation of at least 92% while at rest. Participants walked for 6 minutes or until their O_2 saturation fell below 80% for 6 seconds. The distance walked at that point was recorded. Subsequent tests were performed with the use of the same amount of O_2 used at screening. Those whose resting O_2 saturation on follow-up testing did not reach 88% during administration of the baseline O_2 were not retested and were recorded as having walked 0 metres Length of follow-up: 12 weeks for the randomised comparison (28 for the open-label extension study)

Participant characteristics [mean (SD) unless stated]	Sildenafil, $n = 89$	Placebo, $n = 91$
Age (years)	69.76 (8.71)	68.20 (9.25)
M/F, n (%)	75/14 (84/16)	75/16 (82/18)
Race, n (%)		
White	78 (88)	85 (93)
Black	5 (6)	1 (1)
Other	6 (7)	5 (5)
History of smoking, n (%)	68 (76)	69 (76)
Time since diagnosis, years	2.03 (1.94)	1.87 (1.93)
Supplemental O_2 during walk test, n (%)	28 (31)	24 (26)
6MWT distance, m		
First test	246.93 (99.11)	267.71 (127.75)
Second test	246.39 (103.40)	269.55 (129.83)
Borg Dyspnoea Index after walk test (range 0–10)	3.82 (1.95)	3.33 (1.73)
UCSDSBQ (range 0–120)	50.71 (22.00)	43.28 (20.18)
Total score on SGRQ (range 0–100)	54.55 (16.46)	51.72 (15.86)
Symptoms score on SGRQ (range 0–100)	58.20 (17.75)	53.99 (18.90)
Activity score on SGRQ (range 0–100)	71.20 (17.50)	68.02 (17.63)
Impacts score on SGRQ (range 0–100)	43.20 (19.26)	39.77 (18.81)
SF-36 (subset range 0–100)		
Aggregate physical score	33.17 (9.19)	34.84 (8.69)
Aggregate mental score	49.53 (9.76)	50.58 (9.52)
Bodily pain score	50.55 (10.98)	49.72 (10.44)
General health score	36.99 (9.64)	37.66 (8.73)
Mental health score	51.22 (9.07)	50.95 (8.59)
Physical functioning score	29.20 (8.49)	31.18 (8.31)
Role emotional score	45.13 (12.14)	44.0 (13.60)
Role physical score	34.67 (11.39)	36.38 (11.44)
Social functioning score	42.33 (10.88)	43.06 (10.17)
Vitality score	44.08 (9.28)	45.30 (9.88)
EQ-5D		
Self-report (range –0.59 to 100)	0.71 (0.24)	0.74 (0.19)
VAS (range 0–100)	66.49 (17.45)	67.66 (16.98)
FVC % predicted	54.89 (14.00)	58.73 (14.12)
DL_{CO} % predicted	25.81 (6.03)	26.73 (6.16)
Partial pressure O_2, mmHg	66.22 (12.22)	69.88 (12.85)
Arterial O_2 saturation, %[a]	91.24 (4.22)	92.59 (3.75)

Results	Sildenafil, *n* = 89	Placebo, *n* = 91	Absolute difference (95% CI); *p*-value
6MWT improvement in distance of ≥ 20%	9 (10%)	6 (7%)	0.39
Comments			
Borg Dyspnoea Index after walk test, mean change (95% CI)	0.04 (–0.30 to 0.37)	0.37 (0.04 to 0.70)	–0.34 (–0.81 to 0.14); 0.16

The Borg Dyspnoea Index measures perceived breathlessness on a scale of 0 (none) to 10 (maximum) and has a minimally important difference of 1 (reference provided)

| UCSDSBQ mean change (95% CI) | 0.22 (–3.10 to 3.54) | 6.81 (3.53 to 10.08) | –6.58 (–11.25 to –1.92); 0.006 |

UCSDSBQ: indicates severity of dyspnoea on a scale from 0 to 5 on 21 activities of daily living, along with three ratings on limitations caused by dyspnoea or fear of dyspnoea, for a total score ranging from 0 to 120, with higher score indicating more dyspnoea. The minimally important difference for this is reported to be 5 points (reference provided)

SGRQ mean change			–4.08 (–7.30 to –0.86); 0.01
Total score	–1.64 (–3.91 to 0.64)	2.45 (0.17 to 4.72)	
Symptoms score	–3.58 (–7.02 to –0.13)	2.15 (–1.30 to 5.61)	–5.73 (–10.61 to –0.85); 0.02
Activity score	–1.15 (–3.68 to 1.38)	2.49 (0.00 to 4.99)	–3.64 (–7.20 to –0.09); 0.04
Impacts score (social function)	–0.88 (–3.72 to 2.02)	2.82 (–0.03 to 5.67)	–3.70 (–7.76 to 0.37); 0.07

SGRQ: asks how breathing problems impair their life and is scored from 0 (no impairment) to 100 (maximum impairment). The minimally important difference for this is reported to be 5 to 8 points (reference provided)

SF-36 mean change			–0.17 (–2.06 to 1.73); 0.86
Aggregate physical score	–0.51 (–1.86 to 0.83)	–0.35 (–1.68 to 0.99)	
Aggregate mental score	1.30 (–0.59 to 3.18)	3.02 (1.15 to 4.89)	–1.72 (–4.38 to 0.93); 0.20
Bodily pain score	–0.21 (–2.13 to 1.71)	1.97 (0.08 to 3.85)	–2.17 (–4.86 to 0.52); 0.11
General health score	–1.04 (–2.52 to 0.44)	–3.89 (–5.37 to –2.42)	2.86 (0.76 to 4.95); 0.008
Mental health score	–0.16 (–1.81 to 1.49)	–1.31 (–2.93 to 0.30)	1.15 (–1.15 to 3.46); 0.32
Physical functioning score	–0.93 (–2.24 to 0.38)	–1.46 (–2.76 to –0.17)	0.53 (–1.31 to 2.37); 0.57
Role emotional score	–2.72 (–5.56 to 0.12)	–4.82 (–7.63 to –2.01)	2.10 (–1.90 to 6.10); 0.30
Role physical score	–0.87 (–2.85 to 1.10)	–2.03 (–3.98 to –0.08)	1.16 (–1.62 to 3.93); 0.41
Social functioning score	–0.72 (–3.01 to 1.57)	–2.71 (–4.97 to –0.46)	1.99 (–1.22 to 5.21); 0.22
Vitality score	0.02 (–1.70 to 1.75)	–2.01 (–3.70 to –0.31)	2.03 (–0.39 to 4.44); 0.10

The SF-36 measures functional health and well-being scores on eight scales that correlate with two aggregate scores. Each score ranges from 0 to 100, with a higher score indicating better function. Scores are normalised to a mean (SD) of 50 (10). In IPF the SF-36 has a minimally important difference of 2 to 4 points (reference provided)

EQ-5D mean change			0.02 (–0.04 to 0.08); 0.54
Self-report questionnaire	–0.01 (–0.06 to 0.03)	–0.03 (–0.08 to 0.01)	
VAS	0.48 (–3.10 to 4.06)	–1.81 (–5.34 to 1.73)	2.28 (–2.75 to 7.32); 0.37

Results	Sildenafil, $n = 89$	Placebo, $n = 91$	Absolute difference (95% CI); *p*-value
The EQ-5D measures general QoL on a self-report questionnaire on a scale of –0.59 to 1.00 (higher score indicating better QoL and a negative value indicating a health state worse than death) and on a VAS with a range of 0 to 100 (higher equals better QoL). The reported minimally important difference is approximately 0.08 for the self-report questionnaire and 7 points for the VAS (reference provided)			
FVC (% predicted), mean change	–0.97 (–2.00 to 0.06)	–1.29 (–2.30 to –0.28)	0.32 (–1.12 to 1.76); 0.66
DL_{CO} (% predicted), mean change	–0.33 (–1.36 to 0.71)	–1.87 (–2.91 to –0.83)	1.55 (0.08 to 3.01); 0.04
Partial pressure O_2 (mmHg), mean change	–0.63 (–2.41 to 1.16)	–3.64 (–5.41 to –1.87)	3.02 (0.50 to 5.53); 0.02
Partial pressure of CO_2 (mmHg), mean change	–0.01 (–0.75 to 0.73)	–0.02 (–0.75 to 0.71)	0.01 (–1.03 to 1.05); 0.98
Alveolar-arterial gradient (mmHg), mean change	0.41 (–1.54 to 2.37)	2.95 (0.99 to 4.92)	–2.54 (–5.31 to 0.23); 0.07
Arterial O_2 saturation (%), mean change	–0.17 (–1.02 to 0.69)	–1.38 (–2.23 to –0.52)	1.21 (0.00 to 2.42); 0.05
Death from any cause, *n* (%)	2 (2)	4 (4)	0.43
Acute exacerbation, *n* (%)	2 (2)	4 (4)	0.68
Serious adverse events, number (%) of participants with at least one serious adverse event involving each organ system, number of events observed (assumed by reviewer)			
Any	13 (15), 14	15 (16), 23	0.73
Respiratory, thoracic or mediastinal disorder	7 (8), 7	9 (10), 11	0.63
Worsening IPF	2 (2), 2	5 (5), 5	0.44
Worsening dyspnoea	2 (2), 2	1 (1), 1	0.62
Respiratory failure	1 (1), 1	2 (2), 2	0.99
COPD	0	1 (1), 1	0.99
Hypoxaemia	1 (1), 1	0	0.49
Pleural effusion	0	1 (1), 1	0.99
Pneumothorax	0	1 (1), 1	0.99
Pulmonary embolism	1 (1), 1	0	0.49
Infection or infestation	3 (3), 4	2 (2), 2	0.68
Pneumonia	2 (2), 2	1 (1), 1	0.62
Bronchitis	0	1 (1), 1	0.99
Influenza	1 (1), 1	0	0.49
Viral infection	1 (1), 1	0	0.49
Cardiac disorder	1 (1), 1	3 (3), 3	0.62
Atrial fibrillation	0	2 (2), 2	0.50
Congestive heart failure	1 (1), 1	0	0.49
Coronary artery disease	0	1 (1), 1	0.99
Gastrointestinal disorder	2 (2), 2	1 (1), 1	0.62
Ischaemic colitis	1 (1), 1	0	0.49
Intestinal obstruction	0	1 (1), 1	0.99
Peptic ulcer haemorrhage	1 (1), 1	0	0.49

Results	Sildenafil, $n = 89$	Placebo, $n = 91$	Absolute difference (95% CI); p-value
Hepatobiliary disorders	0	1 (1), 1	0.99
Biliary colic	0	1 (1), 1	0.99
Injury, poisoning and procedural complications	0	1 (1), 3	0.99
Fall	0	1 (1), 1	0.99
Femur fracture	0	1 (1), 1	0.99
Joint injury	0	1 (1), 1	0.99
Neoplasms benign, malignant and unspecified	0	1 (1), 1	0.99
Neoplasms malignant	0	1 (1), 1	0.99
Nervous system disorders	0	1 (1), 1	0.99
Convulsion	0	1 (1), 1	0.99
Adverse events, number (%) with at least one adverse event, number of adverse events observed[b]			
Any body system and event	80 (89.9), 442	79 (86.8), 453	0.52
Respiratory, thoracic and mediastinal disorders	46 (51.7), 85	52 (57.1), 86	0.46
Infections or infestations	42 (47.2), 58	39 (42.9), 52	0.56
Nervous system disorders	35 (39.3), 56	35 (38.5), 58	0.91
Gastrointestinal disorders	32 (36.0), 45	27 (29.7), 43	0.37
General disorders and administration site conditions	27 (30.3), 45	26 (28.6), 44	0.80
Musculoskeletal and connective tissue disorders	25 (28.1), 34	20 (22.0), 33	0.34
Vascular disorders	15 (16.9), 19	13 (14.3), 16	0.64
Skin and subcutaneous tissue disorders	12 (13.5), 21	14 (15.4), 15	0.72
Cardiac disorders	11 (12.4), 15	13 (14.3), 15	0.70
Investigations	10 (11.2), 11	13 (14.3), 18	0.54
Psychiatric disorders	6 (6.7), 7	16 (17.6), 21	0.03
Eye disorders	12 (13.5), 17	9 (9.9), 10	0.45
Metabolism and nutritional disorders	5 (5.6), 5	11 (12.1), 16	0.13
Blood and lymphatic system disorders	7 (7.9), 7	2 (2.2), 2	0.10
Ear and labyrinth disorders	5 (5.6), 5	3 (3.3), 5	0.49
Injury, poisoning and procedural complications	3 (3.4), 5	4 (4.4), 6	0.99
Renal and urinary disorders	3 (3.4), 3	4 (4.4), 5	0.99
Neoplasms benign, malignant and unspecified	2 (2.2), 2	3 (3.3), 3	0.99
Hepatobiliary disorders	0	2 (2.2), 3	0.50
Reproductive system and breast disorders	1 (1.1), 1	1 (1.1), 1	0.99

Results	Sildenafil, $n = 89$	Placebo, $n = 91$	Absolute difference (95% CI); *p*-value

Methodology

Allocation to treatment groups: states participants were randomly assigned in a 1 : 1 ratio with the use of a permuted-block design, with stratification according to clinical centre

Blinding: states double blind. For the primary outcome the tests were conducted by study personnel who were not directly involved in study co-ordination. Placebo tablets described as identical. Acute exacerbations adjudicated by a central committee in a blinded fashion. The 6MWTs were performed by personnel not directly involved in study co-ordination

Comparability of treatment groups: appear similar, authors make no reference to this, except in the table of baseline characteristics it states arterial O_2 saturation was significant (assume between groups)

Method of data analysis: on the primary outcome states participants who withdrew, died, or were unable to complete the walk test for any reason were considered to have an improvement of < 20% (an ITT analysis). The primary test statistic was based on a chi-squared test comparing the rates of improvement of ≥ 20% on testing of the 6MWT from baseline to 12 weeks in the two study groups. Analysis of continuous outcomes used a linear mixed model, with adjustment for baseline measurements for age, sex, race, height, and DL_{CO}. Survival estimates were constructed using the Kaplan–Meier method, with statistical comparison based on the log-rank statistic. No adjustment to the value (0.05) considered to indicate statistical significance was made for multiple comparisons

Sample size/power calculation: the study was powered to show an improvement of ≥ 20% on the 6MWT distance from enrolment to 12 weeks. Based on available safety and efficacy data, a response rate of 30% was expected for sildenafil. On the basis of an assumed placebo response rate of 10%, with an overall type 1 error rate of 0.05, 170 patients were needed to provide a power of 90%

Attrition/dropout: numbers and reasons provided

General comments

Generalisability: the study inclusion criteria states that participants were in an advanced stage of IPF

Intercentre variability: not reported

Conflict of interests: Pfizer donated sildenafil and placebo tablets, states had no role in the study design, accrual or analyses of data, or preparation of the manuscript. Conflicts of interests disclosed by authors

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

a   States $p < 0.05$; assume this means between the two groups.
b   Online supplement reports details of all adverse events, major category types only data extracted here.
89.8% in the sildenafil group and 85.7% in the placebo group reported that they missed no more than 1 day of medication.

Jastrezebski et al.[61]

Study details	Participant details
Jastrzebski et al. 2008[61] Note that this publication is in Polish and limited details have been translated and extracted *Country:* Poland *Design:* CCT *Number of centres:* one *Funding:* not translated	*Number of participants:* 30: inspiratory muscle training $n = 16$, control $n = 14$ *Sample attrition/dropout:* 30 completed the study; four eligible participants did not complete the study, two withdrew despite no side effects, two were excluded due to exacerbations *Sample crossovers*: not translated *Inclusion criteria:* IPF diagnosed by ATS/ERS criteria in patients aged > 50 years, and by lung biopsy in patients < 50 years. At least 2 years' duration of disease. Patients were in remission, i.e. no infection or disease exacerbation requiring increased doses of corticosteroids in past month *Exclusion criteria:* > 20 mg prednisone per day, duration of treatment > 2 years, use of home oxygen therapy

Intervention details	Outcomes
Intervention 1. Inspiratory muscle training 2. Control Description of interventions: not translated in full. Appears that both groups received pulmonary rehabilitation, described as general body conditioning exercises Duration of treatment: 12 weeks (two 6-week cycles)	Primary outcomes: not specified as primary or secondary outcomes. Dyspnoea (oxygen cost diagram, BDI). QoL (SF-36), 6MWT (distance, dyspnoea in Borg's scale), maximal inspiratory pressure, lung function tests (IC, TLC, VC, FEV_1, DL_{COSB}, DL_{CO}/VA) Method of assessing outcome: assessed before, after 6 weeks and after 12 weeks Length of follow-up: 12 weeks

Participant characteristics [mean (SD)]	Inspiratory muscle training, $n = 16$	Control, $n = 14$
Sex (M/F), %	10/6 (63/37)	9/5 (64/36)
Mean (SD) age, years	56.5 (6.5)	56.2 (7.2)
Maximal inspiratory pressure, cmH_2O	96.9 (33.5)	96.9 (25.0)
Inspiratory capacity, l	1.9 (0.7)	1.9 (0.7)
Inspiratory capacity, % predicted	67.9 (20.7)	66.1 (23.0)
TLC, l	3.76 (0.68)	3.75 (0.94)
TLC, % predicted	59.46 (9.61)	59.11 (12.87)
FVC, l	2.6 (0.6)	2.6 (0.6)
FVC, % predicted	67.3 (14.3)	69.2 (14.6)
FEV_1, l	2.7 (0.6)	2.5 (0.6)
FEV_1, % predicted	78.6 (14.3)	78.1 (14.5)
DL_{COSB} (single breath), ml/minute/mmHg	11.1 (4.9)	10.6 (4.9)
DL_{COSB},% predicted	39.5 (15.9)	38.1 (18.9)
DL_{CO}/VA, ml/minute/mmHg	3.4 (1.1)	3.3 (0.8)
DL_{CO}/VA, % predicted	73.4 (22.1)	71.6 (12.8)
6MWT, m	487.4 (100.2)	485.6 (111.9)
Saturated O_2, %, before 6MWT	95 (2)	95 (3)

Participant characteristics [mean (SD)]	Inspiratory muscle training, $n = 16$	Control, $n = 14$
Sat O_2, %, after 6MWT	87 (4)	87 (5)
Dyspnoea in Borg Dyspnoea Index before 6MWT	1.7 (1.1)	1.9 (1.0)
Dyspnoea in Borg Dyspnoea Index after 6MWT	5.3 (2.2)	5.2 (2.3)
BDI-FI	2.0 (0.7)	2.2 (1.2)
BDI-MT	2.3 (0.9)	2.29 (0.6)
BDI-ME	2.2 (1.1)	2.07 (0.6)
BDI FI+MT+ME	6.1 (2.3)	6.07 (2.3)
OCD	69.7 (11.5)	68.9 (12.1)
SF-36 PF	54.4 (23.6)	54.3 (17.4)
SF-36 RP	42.8 (32.0)	44.6 (24.4)
SF-36 BP	68.9 (27.2)	66.8 (22.2)
SF-36 GH	37.8 (17.7)	37.4 (11.1)
SF-36 VT	54.4 (18.2)	52.5 (13.3)
SF-36 SF	58.9 (23.5)	58.0 (14.4)
SF-36 RE	68.8 (39.4)	69.0 (44.3)
SF-36 MH	64.2 (17.7)	65.1 (17.9)
SF-36 PCS	38.4 (8.2)	36.1 (9.1)
SF-36 MCS	46.6 (9.9)	46.5 (10.9)

Results [mean (SD)]	Inspiratory muscle training, $n = 16$	Control, $n = 14$	Difference, p-value
After 6 weeks:			
MIP, cmH_2O	108.4 (31.4)	98.8 (24.8)	
6MWT, m	553.1 (85.3)	518.1 (101.7)	
Dyspnoea in Borg Dyspnoea Index before 6MWT	1.5 (0.7)	1.6 (0.8)	
Dyspnoea in Borg Dyspnoea Index after 6MWT	4.7 (2.4)	4.7 (2.2)	
BDI-FI	2.5 (1.2)	2.4 (1.2)	
BDI-MT	2.7 (1.2)	2.4 (0.6)	
BDI-ME	2.6 (1.0)	2.1 (0.7)	
MDI + MT + ME	7.4 (3.2)	5.7 (2.4)	
OCD	79.1 (17.4)	75.0 (14.0)	
After 12 weeks			
MIP, cmH_2O	115.4 (37.3)	98.9 (24.1)	
6MWT, m	600.8 (93.7)	544.5 (121.5)	
Dyspnoea in Borg Dyspnoea Index before 6MWT	1.2 (0.5)	1.6 (1.6)	
Dyspnoea in Borg Dyspnoea Index after 6MWT	3.8 (2.3)	4.2 (2.1)	

Results [mean (SD)]	Inspiratory muscle training, *n* = 16	Control, *n* = 14	Difference, *p*-value
BDI-FI	2.7 (1.1)	2.1 (0.8)	
BDI-MT	3.0 (0.9)	2.6 (0.8)	
BDI-ME	2.9 (0.8)	2.5 (0.8)	
MDI + MT + ME	7.9 (2.7)	6.4 (2.5)	
OCD	75.6 (16.8)	71.8 (15.4)	
After 6 weeks			
SF-36 PF	69.7 (23.8)	63.2 (12.6)	
SF-36 RP	57.8 (43.5)	53.6 (33.8)	
SF-36 BP	71.7 (23.0)	64.0 (19.2)	
SF-36 GH	48.1 (19.4)	40.2 (16.0)	
SF-36 VIT	61.6 (14.6)	59.3 (17.3)	
SF-36 SF	74.2 (33.1)	69.1 (20.8)	
SF-36 RE	77.1 (35.9)	76.2 (27.5)	
SF-36 MH	70.0 (17.2)	67.5 (20.3)	
SF-36 PCS	42.9 (9.6)	37.8 (7.4)	
SF-36 MCS	49.5 (9.4)	49.1 (12.9)	
After 12 weeks			
SF-36 PF	68.1 (22.3)	62.5 (14.5)	
SF-36 RP	65.6 (31.5)	64.3 (30.6)	
SF-36 BP	75.7 (20.7)	69.6 (17.8)	
SF-36 GH	44.2 (22.4)	42.4 (13.6)	
SF-36 VIT	60.0 (16.4)	57.1 (16.9)	
SF-36 SF	75.8 (30.8)	68.7 (18.2)	
SF-36 RE	70.8 (29.5)	78.6 (28.1)	
SF-36 MH	68.7 (16.8)	66.3 (20.2)	
SF-36 PCS	44.8 (6.0)	42.3 (5.8)	
SF-36 MCS	47.8 (11.22)	47.8 (11.6)	

Methodology

Note that this publication is in Polish and limited details have been translated and extracted

- Allocation to treatment groups: unclear, appears not to be random from the abstract
- Blinding: unclear (assume none)
- Comparability of treatment groups: no statistically significant differences between groups
- Method of data analysis: not translated. However, no obvious between-group comparisons were made. Only changes from baseline for each group were reported in the English abstract
- Sample size/power calculation: not translated
- Attrition/dropout: not translated

General comments

- Generalisability: not stated in abstract
- Outcome measures: appear valid
- Intercentre variability: not applicable
- Conflict of interests: does not appear to be reported

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	No
2. Was the allocation adequately concealed?	No
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Unclear
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	Unclear
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Unclear
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Unclear

BP, bodily pain; DLc$_{OSB}$, diffusing capacity of the lung for carbon monoxide in single breath; DL$_{CO}$/VA, diffusing capacity of the lung for carbon monoxide/alveolar volume; FI, functional impairment; FI+MT+ME, summary score; GH, general health; IC, inspiratory capacity; MCS, mental cumulative score; ME, magnitude of effort; MH, mental health; MT, magnitude of task; OCD, oxygen cost diagram; PCS, physical cumulative score; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; VIT, vitality.
p-value not stated for all baseline characteristics listed above.
p-values reported for change from baseline within groups only, not data extracted.
States that there were no significant changes in the results of lung function tests in either group after 12 weeks of pulmonary rehabilitation (translated using Google Translate); no data were presented.

Lindell *et al.*[74]

Study details	Participant details
Lindell *et al.* 2010[74] *Country:* USA *Design:* RCT *Number of centres:* one *Funding:* Fairbanks-Horix Foundation	*Number of participants*: 21 patients and 21 carers in total. Intervention $n = 10$ pairs; control $n = 11$ pairs *Sample attrition/dropout:* in the control group two patients and three care partners did not complete the data collection. Two pairs did not continue because the patient died or received a lung transplant. The third care partner was lost to follow-up *Sample crossovers:* not reported *Inclusion criteria:* aged > 21 years, be able to read and understand English, to be diagnosed with IPF, to have a FVC reflecting moderate IPF (FVC 55–70% predicted) or severe IPF (< 55% predicted). Care partners were required to be aged > 21 years, be able to read and understand English, to live with or care for the patient with IPF Diagnostic criteria for IPF not stated *Exclusion criteria:* none stated

Intervention details	Outcomes
Intervention 1. Program to Reduce IPF Symptoms and Improve Management (PRISIM) 2. Usual care Description of interventions: 6-weekly group sessions attended by patients and care partners. The content for the sessions was developed collaboratively by a pulmonary clinical nurse specialist whose practice included patients with IPF, a psychiatric clinical specialist with training as a cognitive–behavioural therapist, and an advanced care planning instructor. The sessions included: 　What is IPF and how to live with it (causes, pathophysiology, treatment) 　Gaining control of your moods and feelings: you feel the way you think (basic principles of cognitive–behavioural techniques and cognitive distortions) 　Gaining control of your moods and feelings: what can you do about depression (concepts of stress and depression and interrelationships with illness) 　Putting your life in order: what do I do now? (Planning for uncertainty, concerns related to terminal illness, communicating with clinicians, coping and planning one's affairs) 　Living with IPF (symptom management, energy conservation, oxygen therapy, the importance of exercise) 　Wrap up and review (informal discussions and review) Each session lasted approximately 2 hours	Primary outcomes: not specified as primary or secondary outcomes. Dyspnoea (UCSDSBQ); anxiety (BAI); depression (Beck Depression Inventory-II); stress (PSS); QoL (SF-36 version 2); pulmonary function tests Secondary outcomes: Method of assessing outcome: participants completed questionnaires before and after completion of the intervention. Self-report measures are reported to have established reliability and validity. The battery of tests required 1 hour to complete PFTs were obtained at 3- to 6-month intervals or more frequently, if indicated Pairs randomised to the intervention group were interviewed in their own homes at the end of the intervention. Each of the pair was interviewed separately by one researcher. The interviews were open-ended to collect data that reflected the perspectives on their experience Length of follow-up: unclear

Intervention details	Outcomes
Both groups received a copy of a book '*Feeling Good: The New Mood Therapy*'. The control groups could read at leisure, the treatment group were to use in group exercise. Usual care participants were seen by members of the clinical care team (pulmonary clinical nurse specialist, physicians) at interval of every 3–6 months. The nurse specialist was also available by telephone to answer questions, and conducted an optional monthly support group. Psychological counselling was provided if indicated but not offered on a routine basis Concurrent treatment: not reported Duration of treatment: 6 weeks	

Participant characteristics [mean (SD) unless stated]	PRISIM, $n = 10$	Usual care control, $n = 11$	*p*-value
Age, years	65.2 (10.28)	67.09 (11.90)	
Sex (M/F), %	7/3 (70/30)	9/2 (82/18)	0.6351
Caucasian, %	10 (100%)	10 (90%)	1.0
Diagnosis, %			
Biopsy	3 (30)[a]	9 (81)[a]	0.03
HRCT	7 (70)[a]	2 (19)[a]	
FVC, % predicted		$n = 10$	
> 55%	8 (80)[a]	6 (60)[a]	0.665
50–55%	1 (10)[a]	2 (20)[a]	
< 50%	1 (10)[a]	2 (20)[a]	
Prior or current depression, %	2 (20)[a]	2 (19)[a]	1.0

Comments: states the majority (58%) had scores indicating mild to severe anxiety. Four participants reported scores consistent with mild depression ($n = 2$) or moderate depression ($n = 2$)

Also reports data for the care partners, not extracted here

Results	PRISIM, $n = 10$	Usual care control, $n = 9$	Difference, *p*-value
UCSDSBQ adjusted mean (SD) score	49.51 (22.64)	49.88 (22.64)	0.972

Comments: UCSDSBQ rate severity of shortness of breath on a 6-point scale (0 = not at all, 5 = maximal or unable to do) during 21 activities of daily living associated with varying exertion. Scores were obtained by summing responses on a range of items (range 0–120)

BAI adjusted mean (SD)	15.13 (6.92)	8.56 (6.95)	0.077

Comments: BAI uses a 21-item tool that uses a four-point scale (0 = absent/not at all disturbing to 3 = I could barely stand it). Scores were obtained by summing the 21 items (range 0–63). A score of 0–7 indicated no anxiety; 8–15 mild anxiety; 16–25 moderate anxiety; >26 severe anxiety

BDI adjusted mean (SD)	9.71 (4.34)	9.44 (4.35)	0.894

Comments: the Beck Depression Inventory-II is a revised version of the 21-item instrument Beck Depression Inventory. A score of 0–13 suggests minimal depression, 14–19 mild depression, 20–28 moderate depression, 29 to 63 severe depression

PSS adjusted mean (SD)	19.32 (3.64)	18.20 (3.65)	0.531

Comments: the PSS measures the degree to which participants find their lives unpredictable, uncontrollable and overloading. Respondents are asked to indicate how they feel or thought in the last month using the option of 0 (never) to 4 (very often). Total scores range from 0 to 40, with higher scores indicating more stress

Results	PRISIM, $n = 10$	Usual care control, $n = 9$	Difference, p-value
SF-36 physical adjusted mean (SD)	31.06 (4.61)	36.04 (4.63)	0.038
SF-36 mental adjusted mean (SD)	55.98 (2.71)	55.61 (2.71)	0.772

Comments: SF-36 scores range from 0 (maximum impairment) to 100 (no impairment). The eight domains can be grouped into a physical score that includes physical functioning, role physical, bodily pain and general health, and a mental score that includes vitality, social functioning, role emotional and mental health. The physical score and mental score were normalised to responses from the general population (mean score = 50)

Methodology

Allocation to treatment groups: randomisation was undertaken after completion of the baseline questionnaires. Participants were randomised using a permuted blocked design to ensure that equal numbers of patients with moderate IPF (FVC 55–70% predicted) and severe IPF (< 55% predicted) were assigned to each group

Blinding: not reported

Comparability of treatment groups: states no statistically significant differences between groups in regard to demographic variables or FVC. Baseline scores for anxiety and physical HRQoL (SF-36) were stated to be different between groups (data not reported)

Method of data analysis: analysed those remaining in the study ($n = 37$; PRISIM $n = 20$; control $n = 17$). ANCOVA used to account for the differences identified in baseline for anxiety and physical HRQoL. Description of coding used for qualitative data provided; however, not extracted here

Sample size/power calculation: states the study was designed to obtain pilot data for a future study and therefore the sample size was not based on a power calculation

Attrition/dropout: numbers and reasons provided, unclear when dropouts occurred

General comments

Generalisability: severity of participants not stated; however, based on FVC % predicted at baseline at least 30% of participants could be classed as severe IPF (FVC ≤ 50%)

Outcome measures: all self-report measures are validated measures. No validation in this participant group reported

Intercentre variability: not applicable

Conflict of interests: authors declare no conflict of interests

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	No
4. Was the care provider blinded?	No
5. Was the patient blinded?	No
6. Were outcome assessors blinded to the treatment allocation?	No
7. (i) Were there any unexpected imbalances in dropouts between groups?	Yes
(ii) If so, were they explained or adjusted for?	Yes
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	No

PFT, pulmonary function test.
a Numbers and percentages calculated by reviewer based on the proportions presented for the total patient group.

Nishiyama et al.[73]

Study details	Participant details
Nishiyama et al. 2008[73]	*Number of participants:* 30 participants: pulmonary rehabilitation $n = 15$; control $n = 15$
Country: Japan	*Sample attrition/dropout:* two in the pulmonary rehabilitation group refused to undergo treatment and did not have initial evaluations undertaken
Design: RCT	
Number of centres: one	*Sample crossovers*: not reported
Funding: Japanese Ministry of Health, Labour and Welfare	*Inclusion criteria:* age < 75 years, diagnosis of IPF, shortness of breath on effort, stable condition with no infection or exacerbations in the previous 3 months. The diagnosis of IPF was made in accordance with the ATS/ERS statement 2002 using the major criteria: exclusion of other known causes of ILD; abnormal pulmonary function with restriction and impaired gas exchange; bibasilar reticular abnormalities on HRCT; transbronchial lung biopsy or BAL showing no features to support and alternative diagnosis. Minor criteria included age > 50 years, insidious onset of otherwise unexplained dyspnoea; duration of illness > 3 months; bibasilar inspiratory crackles. All major and at least three minor criteria had to be satisfied. For those with a surgical lung biopsy showing UIP, only the major criteria were required
	Exclusion criteria: severe comorbid illnesses, collagen vascular diseases, the need for long-term oxygen therapy and previous treatment with corticosteroids or immunosuppressives

Intervention details	Outcomes
Intervention	Primary outcomes: not specified as primary or secondary
1. PRP	Pulmonary function tests (FVC, FEV$_1$, TLC, PaO_2, $PaCO_2$,); DL$_{CO}$; 6MWT; BDI; SGRQ
2. Control	
Description of intervention: the PRP was a general programme, not specific to IPF. It involved a twice-weekly outpatient programme of exercise training integrated with peripheral muscle training. Week 1 was the baseline measurements. From weeks 2–9, exercise training was performed on a treadmill at 80% of the patient's maximal walking speed assessed at the baseline 6MWT. In those who underwent baseline cycle ergometer test, exercise intensity was also targeted at 80% of the initial maximum workload. Supplemental oxygen was given to maintain oxygen saturation above 90% if desaturation occurred. Strength training for the limbs was conducted using elastic bands; exercises included arm raising and knee extensions for about 20 minutes. Some educational lectures were also held (no details reported)	Method of assessing outcome: PFTs according to the method described by the ATS; 6MWT according to ATS statement. The test was undertaken twice at each evaluation and the longer distance was used. BDI and SGRQ have been validated in Japanese participants (reference provided)

Length of follow-up: 10 weeks after the start of the programme |
No detail of the control group reported	
Concurrent treatment: no treatments with corticosteroids or immunosuppressive agents during the study	
Duration of treatment: 10-week programme	

Participant characteristics [all mean (SD) unless stated]	Pulmonary rehabilitation, $n = 13$	Control, $n = 15$	
Sex (M/F), n	12/1	9/6	
Age (years)	68.1 (8.9)	64.5 (9.1)	
BMI (kg/m^2)	23.0 (3.8)	22.9 (2.8)	
FVC, l	2.1 (0.4)	2.0 (0.8)	
FVC % predicted	66.1 (13.2)	68.7 (19.5)	
FEV_1, l	1.6 (0.2)	1.7 (0.6)	
FEV_1, % predicted	73.3 (15.0)	78.3 (19.4)	
FEV_1/FVC (%)	78.8 (8.2)	85.2 (6.1)	
TLC, l	3.2 (0.7)	3.1 (1.0)	
TLC, % predicted	64.1 (13.1)	66.6 (16.1)	
DL_{CO}%	59.4 (16.7)	48.6 (16.7)	
PaO_2 mmHg	79.8 (11.5)	83.0 (12.3)	
$PaCO_2$ mmHg	33.6 (6.5)[a]	39.5 (6.0)	
6MWT distance, m	385 (116)	476 (128)	
BDI score	6.7 (1.4)[b]	8.4 (1.5)	
SGRQ score			
Symptoms	56.4 (22.3)	38.0 (25.8)	
Activity	64.7 (17.1)	50.4 (26.2)	
Impacts	39.7 (17.6)	29.9 (23.7)	
Total	50.2 (16.3)	37.8 (22.7)	
Results (absolute values after programme)	**Pulmonary rehabilitation, $n = 13$**	**Control, $n = 15$**	**Difference in change from baseline (95% CI); p-value**
FVC, l	2.1 (0.4)	2.0 (0.8)	0.03 (–0.13 to 0.19); NS
Comments			
FEV_1, l	1.6 (0.2)	1.7 (0.6)	0.04 (–0.17 to 0.08); NS
Comments			
TLC, l	3.3 (0.6)	3.3 (1.0)	0.03 (–0.18 to 0.24); NS
Comments			
PaO_2, mmHg	79.5 (9.7)	75.2 (5.4)	5.5 (–5.0 to 16.0); NS
Comments			
$PaCO_2$, mmHg	35.4 (5.6)	42.3 (2.9)	–1.0 (–5.8 to 3.9); NS
Comments			
6MWT distance, m	427 (84)	472 (130)	46.3 (8.3 to 84.4); < 0.01
Comments			
BDI score	6.7 (1.3)	8.0 (2.2)	0.4 (–0.6 to 1.4); NS

Comments: BDI comprises three categories: functional impairment, magnitude of task and magnitude of effort. Each category recognises five grades (0–4), and the total BDI score ranges from 0 to 12, with higher score denoting milder dyspnoea in daily living

Results (absolute values after programme)	Pulmonary rehabilitation, $n = 13$	Control, $n = 15$	Difference in change from baseline (95% CI); p-value
SGRQ score			
Symptoms	53.4 (25.8)	40.6 (21.2)	–5.7 (–18.7 to 7.2); NS
Activity	62.5 (16.9)	54.0 (22.6)	–5.8 (–14.7 to 3.1); NS
Impacts	36.5 (17.5)	32.9 (23.5)	–6.2 (–12.8 to 0.3); NS
Total	47.3 (17.4)	40.9 (20.7)	–6.1 (–11.7 to 0.5); $p < 0.05$

Comments: SGRQ comprises three component scores (symptoms, activity, impacts), which sum to a total score. Each component can range from 0 to 100, with a lower score denoting better HRQoL

Methodology

Allocation to treatment groups: participants were randomly assigned to groups using sealed envelopes that had been prepared prior to the study

Blinding: not reported whether or not outcome assessors were blind to treatment allocation. Blinding of participants would have been inappropriate

Comparability of treatment groups: The $PaCO_2$ and BDI scores were lower in the PRP group. No other differences were observed

Method of data analysis: paired t-tests were used to test differences in the values for each participant before and after treatment. Comparisons between groups were tested with ANCOVA to account for differences in baseline characteristics

Sample size/power calculation: not stated

Attrition/dropout: numbers and reasons provided

General comments

Generalisability: severity of participants not stated, baseline FVC suggests participants might have moderate IPF

Outcome measures: appear valid

Intercentre variability: not applicable

Conflict of interests: not stated

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Unclear
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	No
4. Was the care provider blinded?	No
5. Was the patient blinded?	No
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	Yes Yes
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	No

BAL, bronchoalveolar lavage; NS, not stated; $PaCO_2$, partial pressure of carbon dioxide; PFT, pulmonary function test; PRP, pulmonary rehabilitation programme.
a $p < 0.05$.
b $p < 0.01$.

Noble et al.[62] (Capacity study 004)

Study details	Participant details
Noble et al. 2011[62] Capacity study 004 *Country:* Australia, Belgium, Canada, France, Germany, Ireland, Italy, Mexico, Poland, Spain, Switzerland, UK, USA *Design:* RCT *Number of centres:* 110 *Funding:* InterMune	*Number of participants:* 435 randomised. (1) Pirfenidone 2403 mg/day: $n = 174$; (2) pirfenidone 1197 mg/day: $n = 87$; (3) placebo: $n = 174$ *Sample attrition/dropout:* study discontinuations: (1) Pirfenidone 2403 mg/day: $n = 13$ [four withdrew consent, eight adverse events, one other reasons (deportation)] (2) Pirfenidone 1197 mg/day: $n = 5$ (two withdrew consent, three adverse events) (3) Placebo: $n = 8$ (five withdrew consent, three adverse events) Discontinued treatment: (1) Pirfenidone 2403 mg/day: $n = 38$ [21 adverse events, five patient's decision, three lung transplant, five deaths, four other (unknown reaction to chemotherapy, deportation, non-adherence, spontaneous discontinuation of study drug)] (2) Pirfenidone 1197 mg/day: $n = 17$ (11 adverse events, two patient's decision, four deaths) (3) Placebo: $n = 31$ (14 adverse events, four patient's decision, four lung transplant, nine deaths) *Sample crossovers:* not reported *Inclusion criteria:* aged 40–80 years with a diagnosis of IPF in the previous 48 months and no evidence of improvement in measures of disease severity over the preceding year. Predicted FVC of at least 50%, predicted DL_{CO} of $\geq 35\%$, either predicted FVC or $DL_{CO} \leq 90\%$ and 6MWT distance of at least 150 m. Those younger than 50 years and those not meeting the protocol criteria for definite IPF by HRCT were required to have a lung biopsy sample showing UIP *Exclusion criteria:* obstructive airway disease, connective tissue disease, alternative explanation for ILD, and being on a waiting list for a lung transplant

Intervention details	Outcomes
Intervention 1. Pirfenidone 2403 mg/day (801 mg × three daily doses) 2. Pirfenidone 1197 mg/day (399 mg × three daily doses) 3. Placebo Dose details: 2403 mg was derived from normalisation of the dose used in previous Japanese trials (1800 mg) to account for different predicted body weights. The dose was increased to the full dose over 2 weeks Dose modifications: a protocol for dose modifications was provided for expected adverse events Concurrent treatment: concomitant treatments for IPF were prohibited, with exceptions for short courses of azathioprine, cyclophosphamide, corticosteroids, or acetylcysteine for protocol defined acute exacerbations, acute respiratory decompensation, or disease progression Duration of treatment: 72 weeks	*Primary outcomes:* change in per cent predicted FVC *Secondary outcomes:* categorical FVC (5-point scale), PFS, worsening IPF, dyspnoea, 6MWT distance, worst peripheral oxygen saturation (SpO_2) during the 6MWT, per cent predicted DL_{co}, mortality *Method of assessing outcomes:* PFS defined as time to confirmed $\geq 10\%$ decline in per cent predicted FVC, $\geq 15\%$ decline in per cent predicted DL_{CO}, or death. Worsening IPF defined as time to acute exacerbation, death, lung transplantation, or admission to hospital for respiratory problems. Dyspnoea (UCSDSBQ) *Length of follow-up:* 72 weeks from the date the last patient was enrolled

Participant characteristics	Pirfenidone 1197 mg/day, $n = 87$	Pirfenidone 2403 mg/day, $n = 174$	Placebo, $n = 174$
Mean age (SD), years	68.0 (7.6)	65.7 (8.2)	66.3 (7.5)
Male/female, %	65/22 (75/25%)	118/56 (68/32%)	128/46 (74/26%)
White, %	83 (95%)	168 (97%)	168 (97%)
Mean weight (SD), kg			
Male	88.4 (13.5)	91.3 (15.9)	88.9 (16.1)
Female	72.8 (13.0)	77.0 (13.2)	77.0 (13.6)
Smoking status, *n* (%)			
Never	27 (31)	56 (32)	51 (29)
Former	57 (66)	110 (63)	114 (66)
Current	3 (3)	8 (5)	9 (5)
HRCT diagnosis IPF, *n* (%)	83 (95)	159 (91)	164 (94)
Surgical lung biopsy, *n* (%)	32 (37)	86 (49)	85 (49)
Diagnosis IPF ≤ 1 year, *n* (%)	46 (53)	83 (48)	81 (47)
Per cent predicted FVC (SD)	76.4 (14.4)	74.5 (14.5)	76.2 (15.5)
Mean (SD) DL_{CO} (per cent predicted)	47.2 (8.2)	46.4 (9.5)	46.1 (10.2)
A-a gradient (mmHg)	15.5 (10.4)	17.7 (10.6)	18.9 (14.7)
Mean (SD) 6MWT distance, m	417.5 (112.8)	411.1 (91.8)	410.0 (90.9)
Use of supplemental O_2 (%)	15 (17)	29 (17)	25 (14)
Comments:			

Results [pirfenidone 1197 mg/day results were not compared (see below) and hence data not extracted here]	Pirfenidone 2403 mg/day, n = 174	Placebo, n = 174	Difference (95% CI); p-value
Mean change per cent predicted FVC at 72 weeks	–8.0% (SD 16.5)	–12.4% (SD 18.5)	4.4% (0.7 to 9.1); 0.001
Comments: Figure 2a shows the mean change from baseline for weeks 12, 24, 26, 48, 60 and 72			
Proportion with a decline in FVC ≥ 10%	35 (20%)	60 (35%)	14.4 (95% CI 7.4 to 21.3); 0.01[a]
Comments: subgroup analysis, unclear if stated a priori or powered			
PFS	138/174	116/174	HR 0.64 (0.44 to 0.95); 0.023
Comments: data for PFS presented graphically			
Mean change in 6MWT distance, m	–60.4	–76.8	16.4 (–10.9 to 43.7); 0.171
Comments: presents a post hoc subgroup analysis of proportion of participant with a ≥ 50 m decrement in the 6MWT. Data not extracted			
Mean change in DL_{CO},% predicted	–7.9	–9.9	2.0 (–0.4 to 4.4); 0.145
Comments			
Mean change in dyspnoea score (UCSDSBQ)	12.1	15.2	–3.1 (–8.5 to 2.3); 0.509
Comments: UCSDSBQ total score ranges from 0 to 120, with larger scores indicating greater shortness of breath			
Mean change in worst SpO_2 during 6MWT, %	–1.5	–2.3	0.8 (–0.2 to 1.8); 0.087
Comments			
Time to worsening in IPF			HR 0.84 (0.50 to 1.42); 0.515
Comments			
Adverse events			
Any adverse event	171 (98%)	169 (97%)	
Nausea	60 (35%)	32 (18%)	
Fatigue	48 (28%)	36 (21%)	
Diarrhoea	43 (25%)	30 (17%)	
Rash	53 (31%)	18 (10%)	
Dizziness	33 (19%)	17 (10%)	
Dyspepsia	30 (17%)	16 (9%)	
Gastro-oesophageal reflux	26 (15%)	14 (8%)	
Vomiting	24 (14%)	7 (4%)	
Insomnia	22 (13%)	12 (7%)	
Arthralgia	20 (12%)	13 (8%)	
Anorexia	19 (11%)	7 (4%)	
Abdominal distension	15 (9%)	12 (7%)	
Photosensitivity reaction	25 (14%)	2 (1%)	

Results [pirfenidone 1197 mg/day results were not compared (see below) and hence data not extracted here]	Pirfenidone 2403 mg/day, n = 174	Placebo, n = 174	Difference (95% CI); p-value
Urinary tract infection	19 (11%)	9 (5%)	
Stevens–Johnson syndrome	0	0	
Toxic epidermal necrosis	0	0	

Comments: based on adverse events occurring in ≥ 10% of participants receiving pirfenidone and an incidence of ≥ 1.5 times placebo. Events were generally mild or moderate, without clinical consequences

Serious adverse events

IPF	13 (7.5%)	14 (8.0%)	
Pneumonia	4 (2.3%)	6 (3.4%)	
Respiratory failure	2 (1.1%)	2 (1.1%)	
Bronchitis	2 (1.1%)	2 (1.1%)	
Lobar pneumonia	2 (1.1%)	2 (1.1%)	
Myocardial infarction	0	4 (2.3%)	
Acute respiratory failure	2 (1.1%)	3 (1.7%)	
Angina pectoris	2 (1.1%)	1 (0.6%)	
Atrial fibrillation	1 (0.6%)	1 (0.6%)	
Coronary artery disease	0	2 (1.1%)	
Pneumothorax	3 (1.7%)	0	
Pulmonary embolism	1 (0.6%)	1 (0.6%)	
Syncope	3 (1.7%)	1 (0.6%)	
Dyspnoea	0	3 (1.7%)	
Non-cardiac chest pain	2 (1.1%)	2 (1.1%)	
Prostate cancer[b]	0	2 (1.6%)	
Aortic aneurysm	2 (1.1%)	0	
Chest pain	3 (1.7%)	0	
Hypoxia	1 (0.6%)	2 (1.1%)	
Acute renal failure	1 (0.6%)	0	
Bladder cancer	2 (1.1%)	0	
Gastro-oesophageal reflux disease	2 (1.1%)	0	

Comments: based on serious adverse events occurring in ≥ 2 participants in any treatment group

Methodology

Allocation to treatment groups: participants were randomly assigned in a 2 : 1 : 2 ratio to pirfenidone 2403 mg/day, pirfenidone 1197 mg/day or placebo. The randomisation code (permuted block design with five participants per block) was computer generated, stratified by region, by an independent statistician. Study centres used an interactive voice response system to assign study bottles to participants

Blinding: all personnel involved in the study were masked to treatment group assignment until after final database lock

Comparability of treatment groups: states that there were no pronounced baseline imbalances between groups; not tested statistically

Method of data analysis: states analyses were ITT. A rank analysis of covariance model, stratified by region, with standardised rank change in FVC as the outcome and standardised rank baseline per cent predicted FVC as a covariate, evaluated against a final adjusted two-tailed p-value of 0.0498, was used. Data analyses compared pirfenidone 2403 mg/day with placebo. The group assigned to pirfenidone 1197 mg/day were summarised descriptively (and hence not reported here)

Sample size/power calculation: not reported

Attrition/dropout: numbers and reasons provided

General comments

Generalisability: participants from predominantly European populations; majority were men. Severity of IPF stated as being mild to moderate IPF in the discussion and, based on per cent predicted FVC at baseline, this appears correct. Diagnosis was within 2 years

Outcome measures: mostly surrogate end points. Unclear if the definition of PFS is a standard one. No mortality data presented

Intercentre variability: not discussed

Conflict of interests: states the sponsor participated in the study design, data collection, data analysis and writing the report. After study completion, the sponsor analysed and maintained the data. Authors participated in design, conduct, analysis and reporting; had full access to data; and no limits were placed on the content of the report

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Yes
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Yes
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

A-a gradient, alveolar–arterial oxygen gradient.
a　p-value from an analysis of five categories: severe decline ($\geq 20\%$), moderate decline ($< 20\%$ but $\geq 10\%$), mild decline ($< 10\%$ but ≥ 0), mild improvement (> 0 but $< 10\%$), moderate improvement ($\geq 10\%$).
b　Percentage based on number of males in each treatment group.

Noble *et al.*[62] (Capacity study 006)

Study details	Participant details
Noble *et al.* 2011[62] Capacity study 006 *Country:* Australia, Belgium, Canada, France, Germany, Ireland, Italy, Mexico, Poland, Spain, Switzerland, UK, USA *Design:* RCT *Number of centres:* 110 *Funding:* InterMune	*Number of participants:* 344 randomised: (1) pirfenidone: $n = 171$; (2) placebo: $n = 173$ *Sample attrition/dropout*: Study discontinuations: (1) Pirfenidone: $n = 13$ [six withdrew consent, five adverse events, one sponsor's decision, one other reasons (placement lung transplant list)] (2) Placebo: $n = 9$ (five withdrew consent, four adverse events) Discontinued treatment: (1) Pirfenidone: $n = 34$ [24 adverse events, three patient's decision, two lung transplant, one sponsor's decision, one death, three other (placement on lung transplant list, prolonged QTc interval, unknown)] (2) Placebo: $n = 31$ (14 adverse events, three patient's decision, three lung transplant, 11 deaths) *Sample crossovers:* not reported *Inclusion criteria:* aged 40–80 years with a diagnosis of IPF in the previous 48 months and no evidence of improvement in measures of disease severity over the preceding year. Predicted FVC of $\geq 50\%$, predicted DL_{CO} of $\geq 35\%$, either predicted FVC or $DL_{CO} \geq 90\%$, 6MWT distance of ≥ 150 m. Those younger than 50 years and those not meeting the protocol criteria for definite IPF by HRCT were required to have a lung biopsy sample showing UIP *Exclusion criteria:* obstructive airway disease, connective tissue disease, alternative explanation for ILD, and being on a waiting list for a lung transplant
Intervention details	**Outcomes**
Intervention 1. Pirfenidone 2403 mg/day (801 mg × 3 daily) 2. Placebo Dose details: 2403 mg was derived from normalisation of the dose used in previous Japanese trials (1800 mg) to account for different predicted body weights The dose was increased to the full dose over 2 weeks Dose modifications: a protocol for dose modifications was provided for expected adverse events Concurrent treatment: concomitant treatments for IPF were prohibited, with exceptions for short courses of azathioprine, cyclophosphamide, corticosteroids, or acetylcysteine for protocol defined acute exacerbations, acute respiratory decompensation, or disease progression Duration of treatment: 72 weeks	Primary outcomes: change in per cent predicted FVC Secondary outcomes: categorical FVC (5-point scale), PFS, worsening IPF, dyspnoea, 6MWT distance, worst SpO_2 during the 6MWT, per cent predicted DL_{CO}, fibrosis, mortality Method of assessing outcomes: PFS defined as time to confirmed $\geq 10\%$ decline in per cent predicted FVC, $\geq 15\%$ decline in per cent predicted DL_{CO}, or death. Worsening IPF defined as time to acute exacerbation, death, lung transplantation, or admission to hospital for respiratory problems. Dyspnoea (UCSDSBQ). Fibrosis by the HRCT Length of follow-up: 72 weeks from the date the last patient was enrolled

Participant characteristics	Pirfenidone, n = 171	Placebo, n = 173
Mean age (SD), years	66.8 (7.9)	67.0 (7.8)
Male/female, %	123/48 (72/28)	124/49 (72/28)
White, %	169 (99%)	171 (99%)
Mean weight (SD), kg		
Male	95.4 (17.4)	93.2 (15.1)
Female	76.6 (14.0)	77.5 (14.8)
Smoking status		
Never	59 (35%)	64 (37%)
Former	112 (65%)	101 (58%)
Current	0	8 (5)
HRCT diagnosis IPF	149 (87%)	158 (91%)
Surgical lung biopsy	94 (55%)	94 (54%)
Diagnosis IPF ≤ 1 year	100 (58%)	107 (62%)
Per cent predicted FVC (SD)	74.9 (13.2)	73.1 (14.2)
Mean DL_{CO} (per cent predicted)	47.8 (9.8)	47.4 (9.2)
A-a gradient (mmHg)	18.3 (11.1)	17.0 (10.4)
Mean 6MWT distance, m	378.0 (82.2)	399.1 (89.7)
Use of supplemental O_2	48 (28%)	49 (28%)

Comments:

Results	Pirfenidone, n = 171	Placebo, n = 173	Difference (95% CI); p-value
Mean change per cent predicted FVC at 72 weeks	–9.0% (SD 19.6)	–9.6% (SD 19.1)	0.6% (–3.5% to 4.7%); 0.501

Comments: Figure 2b shows the mean change from baseline for weeks 12, 24, 26, 48, 60 and 72

Proportion with a decline in FVC ≥ 10%	39 (23%)	46 (27%)	3.8 (95% CI –2.7 to 10.2); 0.440[a]

Comments: subgroup analysis, unclear if stated a priori or powered

Results	Pirfenidone, *n* = 171	Placebo, *n* = 173	Difference (95% CI); *p*-value
PFS	126/171	123/171	HR 0.84 (0.58 to 1.22); 0.355
Comments: data for PFS time presented graphically			
Mean change in 6MWT distance, m	–45.1	–76.9	31.8 (3.2 to 60.4); 0.0009
Comments: presents a post hoc subgroup analysis of proportion of participant with a ≥ 50 m decrement in the 6MWT. Data not extracted			
Mean change in DL_{CO} (% predicted)	–9.8	–9.2	–0.5 (–3.2 to 2.2); 0.996
Comments			
Mean change in dyspnoea score (UCSDSBQ)	11.9	13.9	–2.0 (–7.6 to 3.6), 0.604
Comments: UCSDSBQ total score ranges from 0 to 120, with larger scores indicating greater shortness of breath			
Mean change in worst *S*pO₂ during 6MWT, %	–1.9	–1.3	–0.5 (–1.7 to 0.7); 0.893
Comments			
Time to worsening in IPF			HR 0.73 (0.43 to 1.24); 0.248
Comments			
Adverse events, *n* (%)			
Any adverse event	169 (99)	170 (98)	
Nausea	65 (38)	28 (16)	
Fatigue	56 (33)	35 (20)	
Diarrhoea	56 (33)	37 (21)	
Rash	58 (34)	22 (13)	
Dizziness	30 (18)	18 (10)	
Dyspepsia	36 (21)	10 (6)	
Gastro-oesophageal reflux	10 (6)	12 (7)	
Vomiting	23 (14)	8 (5)	
Insomnia	12 (7)	11 (6)	
Arthralgia	16 (9)	11 (6)	
Anorexia	18 (11)	6 (4)	
Abdominal distension	18 (11)	8 (5)	
Photosensitivity reaction	17 (10)	4 (2)	
Urinary tract infection	16 (9)	20 (12)	

Comments: based on adverse events occurring in ≥ 10% of participants receiving pirfenidone and an incidence of ≥ 1.5 times placebo. States these events were generally mild or moderate in severity and without clinical consequences

Results	Pirfenidone, *n* = 171	Placebo, *n* = 173	Difference (95% CI); *p*-value
Serious adverse events, *n* (%)			
IPF	13 (7.6)	17 (9.8)	
Pneumonia	7 (4.1)	7 (4.0%)	
Respiratory failure	4 (2.3)	6 (3.5)	
Coronary artery disease	6 (3.5)	0	
Acute respiratory failure	2 (1.2)	3 (1.7)	
Bronchitis	0	5 (2.9)	
Atrial fibrillation	2 (1.2)	1 (0.6)	
Renal failure	2 (1.2)	2 (1.2)	
Fall	2 (1.2)	1 (0.6)	
Hypotension	2 (1.2)	1 (0.6)	
Prostate cancer[b]	2 (1.6)	0	
Colitis	2 (1.2)	0	
Hip fracture	2 (1.2)	0	
Hypertension	0	2 (1.2)	
Hypoxia	0	2 (1.2)	
Intervertebral disc protrusion	2 (1.6)	0	
Liver function test abnormal	2 (1.2)	0	
Nephrolithiasis	2 (1.2)	0	
Sick sinus syndrome	2 (1.2)	0	
Transitional cell carcinoma	0	2 (1.2)	

Comments: based on serious adverse events occurring in ≥ 2 participants in any treatment group

Methodology

Allocation to treatment groups: participants were randomly assigned to oral pirfenidone or placebo in a 1 : 1 ratio. The randomisation code (permuted block design with four participants per block) was computer generated, stratified by region, by an independent statistician. Study centres used an interactive voice response system to assign study bottles to participants

Blinding: all personnel involved in the study were masked to treatment group assignment until after final database lock

Comparability of treatment groups: states that there were no pronounced baseline imbalances between groups, not tested statistically

Method of data analysis: states that analyses were intention to treat. A rank analysis of covariance model, stratified by region, with standardised rank change in FVC as the outcome and standardised rank baseline per cent predicted FVC as a covariate, evaluated against a final adjusted two-tailed *p*-value of 0.0498 was used

Sample size/power calculation: not reported

Attrition/dropout: numbers and reasons provided

General comments

Generalisability: participants from predominantly European populations; majority were men. Severity of IPF stated as being mild to moderate IPF in the discussion and based on per cent predicted FVC at baseline this appears correct. Diagnosis was within 2 years

Outcome measures: mostly surrogate end points. Unclear if the definition of PFS is a standard one. No mortality data presented

Intercentre variability: not discussed

Conflict of interests: states that the sponsor participated in the study design, data collection, data analysis and writing the report. After study completion the sponsor analysed and maintained the data. Authors participated in design, conduct, analysis, and reporting; had full access to data; and no limits were placed on the content of the report

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Yes
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Yes
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

A-a gradient, alveolar–arterial oxygen gradient.
a *p*-value from an analysis of five categories: severe decline ($\geq 20\%$), moderate decline ($< 20\%$ but $\geq 10\%$), mild decline ($< 10\%$ but $\geq 0\%$), mild improvement ($> 0\%$ but $< 10\%$) and moderate improvement ($\geq 10\%$).
b Percentage based on number of males in each treatment group.

Raghu et al.[70]

Study details	Participant details
Raghu et al. 1991[70] Country: USA Design: RCT Number of centres: two Funding: grant from Virginia Mason Research Centre, Seattle, WA, USA	Number of participants: 27: prednisone and placebo $n = 13$; prednisolone and azathioprine $n = 14$ Sample attrition/dropout: four participants from each group died before completion Sample crossovers: three participants crossed over because of clinical deterioration: two from placebo group (at 4 and 3 months) and one in prednisolone/azathioprine group (at 6 months) Inclusion criteria: a diagnosis of IPF was suspected in patients with diffuse reticulonodular infiltrates on chest roentgenogram, bibasilar crackles, finger clubbing, absence of fever, decrease in FVC and DL_{CO} and no evidence of extrathoracic disease. Only those who were previously untreated and available for routine follow-up were considered candidates. The diagnosis was supported by lung biopsy in all patients with the histological diagnosis based on typical microscopic findings excluding infection, granuloma, vasculitis, or malignancy, and examined by a pathologist who was blinded to clinical details. A transbronchial lung biopsy was accepted in patients with typical clinical, roentgenographic and physiologic features of IPF who either refused to have open lung biopsy or had documented family history of familial IPF. All fulfilled criteria for progressive clinical disease (one or more of (1) progressive dyspnoea from day of onset, (2) progressive roentgenographic parenchymal abnormality, (3) $\geq 10\%$ decrease in FVC or TLC compared with previous values and (4) $\geq 20\%$ reduction in DL_{CO} compared with previous values Exclusion criteria: collagen vascular disease, hypersensitivity pneumonitis, pneumoconiosis, drug-induced diffuse pulmonary injury, or irradiation fibrosis
Intervention details	**Outcomes**
Intervention 1. Prednisone and placebo 2. Prednisone and azathioprine Dose details: prednisolone to both groups according to an identical protocol: initially 1.5 mg/kg/day (not to exceed 100 mg/day) for 2 weeks, followed by a fortnightly decrease of 20 mg/day until a dose of 40 mg/day was reached. Was further decreased in 5- or 10-mg/day decrements every 2 weeks according to patient tolerance in order to maintain a dose of ≤ 20 mg/day Azathioprine administered at 3 mg/kg/day (not to exceed 200 mg/day) to the nearest 25-mg dose increment for the duration of the study Similar numbers of placebo tablets were dispensed to those in group 1 Dose modifications: participants crossed over to the other treatment arm if any of the following occurred: (1) nausea, vomiting, diarrhoea unresponsive to symptomatic treatment, (2) white blood count < 3500/ml, (3) platelet count < 80,000/ml, (4) respiratory failure requiring mechanical ventilation, (5) coma, (6) abnormal liver function tests, (7) rapid disease progression or (8) patient's request	Primary outcomes: not stated as primary or secondary: measurable change in lung function at 12 months; survival Secondary outcomes: see above Method of assessing outcome: improvement in lung function defined as any one of (1) $\geq 10\%$ improvement in FVC, (2) $\geq 20\%$ improvement in DL_{CO} or (3) $\geq 10\%$ improvement in resting $PA - aO_2$. Patients who 'worsened' had decrements of similar magnitude in one or more of these tests. Lesser degrees of change were categorised as unchanged. These were analysed separately to assess the outcome as independent variables rather than overall patient improvement Length of follow-up: at least 12 months

Intervention details	Outcomes

Concurrent treatment: no use of cyclophosphamide, d-penicillamine, non-steroid anti-inflammatory agents, colchicine or any other agents that could potentially influence the course of IPF

Duration of treatment: 12 months

Participant characteristics [mean (SE) unless stated]	Prednisone and placebo, $n = 13$	Prednisolone and azathioprine, $n = 14$	p-value
Sex, M/F	7/6	5/9	
Age, years	54 (3)	58 (2)	0.41
Duration of illness, months	23 (6)	26 (6)	0.87
FVC, % predicted	65 (4)	70 (4)	0.45
DL_{CO} ml/minute/mmHg haemoglobin corrected, % predicted	40 (4)	48 (5)	0.32
$PA - aO_2$, mmHg, at rest	35 (4)	36 (3)	0.83
Severity of fibrosis on open lung biopsy	2.5 (0.3)	2.8 (0.4)	0.50
Comments			

Results	Prednisone and placebo, $n = 9$	Prednisolone and azathioprine, $n = 10$	Difference, p-value
Change in FVC, % predicted	1.7 (7.4)	6.5 (5.3)	0.87
Comments			
Change in DL_{CO}, % predicted	0.9 (5.7)	7.3 (5.3)	0.70
Comments			
Change in rest $PA - aO_2$, mmHg	−1.0 (2.0)	−6.0 (4.0)	0.12
Comments			

	Prednisone and placebo, $n = 13$	Prednisolone and azathioprine, $n = 14$	p-value
FVC, % predicted (%)			
Improved	3 (23.1)	5 (35.7)	% improved, 0.68
Unchanged	3 (23.1)	3 (21.4)	
Deteriorated[a]	7 (53.8)	6 (42.9)	
DL_{CO},% predicted (%)			
Improved	2 (15.4)	3 (21.4)	% improved, 1.00
Unchanged	5 (38.5)	6 (42.9)	
Deteriorated[a]	6 (46.2)	5 (35.7)	
$PA - aO_2$, mmHg (%)			
Improved	3 (23.1)	7 (50)	Improved, 0.24
Unchanged	3 (23.1)	1 (7.1)	Assume others not significant as not stated
Deteriorated[a]	7 (53.8)	6 (42.9)	
Comments			

			HR (95% CI); p-value[b] [p-value[c]]
Survival, all patients	NA	NA	0.48 (0.17 to 1.38); 0.16 [0.15]
Survival, all patients adjusted for age	NA	NA	0.26 (0.08 to 0.88); 0.02 [0.05]
Deaths during observation period	4 (3 respiratory failure, 1 myocardial infarction)	4 (3 respiratory failure, 1 myocardial infarction)	
Alive at 12 months	9	10	
Alive at last follow-up	3	8	
Adverse events	**Prednisone and placebo, n = 13**	**Prednisolone and azathioprine, n = 14**	**p-value**
Any	25	28	Not reported
Subjective			
Gastrointestinal	6	3	Not reported
Neuropsychiatric	4	1	
Objective			
Elevated liver enzymes	0	1	Not reported
Vertebral fractures	0	3	
Acne	1	0	
Cushingoid features	4	5	
Hypertension	1	2	
Diabetes treatment (oral)	3	2	
Diabetes treatment (insulin)	0	2	
Congestive heart failure	2	0	
Myocardial infarction	1	1	
Bacterial pneumonia	1	1	
Herpes zoster	0	2	
Urosepsis	0	1	
Cataracts	1	1	
Myopathy	1	2	
Peptic ulcer disease	1	1	
Pancytopenia	0	0	

Comments

Methodology

Allocation to treatment groups: participants were randomised by block randomisation (blocks of 10) by a research pharmacist. No other details reported

Blinding: states double blinded; all medications were dispensed by the research pharmacist

Comparability of treatment groups: the two groups were comparable in age, clinical duration of disease, pulmonary function testing and amount of fibrosis on open lung biopsy

Method of data analysis: ordinal and interval variables compared with Wilcoxon rank-sum test, proportions compared by Fisher's exact test. Those who died before 1 year were included in the analysis of change in lung function at 1 year by assigning the worst possible change. States all analyses were done on an ITT basis. Kaplan–Meier survival plots and Cox proportional hazards regression model were used to analyse survival time data

Sample size/power calculation: not reported

Attrition/dropout: states no loss to follow-up. Crossovers and deaths reported

General comments

Generalisability: patients were newly diagnosed, had no previous treatments, FVCs in the region of mild to moderate disease

Outcome measures: appear valid

Intercentre variability: not reported

Conflict of interests: not reported

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Unclear

F, female; M, male; NA, not applicable; $PA – aO_2$, alveolar – arterial oxygen partial pressure; SE, standard error.
a Includes those who died in the first year of treatment.
b *p*-value based on large sample approximation.
c *p*-value based on randomisation test.
Also presents data for survival based on subgroups according to three age categories (< 50 years, 50–59 years, > 60 years) – not data extracted as not a predefined subgroup analysis.
Survival calculated based on observation period; duration not stated but assumed 9 years (reported in abstract and in the figure and no discussion of any estimation of data).

Richeldi *et al.*[71]

Study details	Participant details
Richeldi *et al.* 2011[71] *Country:* 25 countries including Italy, Mexico, Germany, USA, Republic of Korea, UK and France *Design:* RCT (dose finding phase II study) *Number of centres:* 92 *Funding:* supported by Boehringer Ingelheim (sponsor)	*Number of participants:* 432 randomised, 428 treated: BIBF 1120 50 mg/day *n* = 86; BIBF 1120 100 mg/day *n* = 86; BIBF 1120 200 mg/day *n* = 86; BIBF 1120 300 mg/day *n* = 85; placebo *n* = 85 *Sample attrition/dropout:* four were randomised but did not participate 112 discontinued study medication: BIBF 1120 50 mg/day *n* = 24 (27.9%); BIBF 1120 100 mg/day *n* = 18 (20.9%); BIBF 1120 200 mg/day *n* = 14 (16.3%); BIBF 1120 300 mg/day *n* = 32 (37.6%); placebo *n* = 24 (28.2%). 96 of the total 112 discontinuations (85.7%) were due to adverse events Of the 32 in the BIBF 1120 300 mg/day group, 11 had previously had their dose reduced The dose was reduced by one dose level in 5, 7, 11, 20 and 7 participants in the five group, respectively *Sample crossovers:* none *Inclusion criteria:* patients ≥ 40 years of age who had IPF that was consistent with the ATS/ERS 2000 criteria and who had received the diagnosis < 5 years before screening. Eligible patients had undergone HRCT < 1 year before randomisation and had a FVC that was ≥ 50% of their predicted value, a DL_{CO} that was 30–79% of predicted value, and a PaO_2 when breathing ambient air that was ≥ 55 mmHg at altitudes up to 1500 m, or ≥ 50 mmHg at altitudes above 1500 m. Diagnosis was confirmed by independent review of HRCT scans by an expert chest radiologist and assessment of surgical lung-biopsy specimens (if available) by an expert lung pathologist *Exclusion criteria:* medical conditions or concomitant medications that might interfere with the performance of the study, other diseases that might interfere with testing procedures (e.g. myocardial infarction or unstable angina), continues (> 15 hours per day) oxygen supplementation at randomisation, known predisposition to bleeding or thrombosis, concomitant anticoagulation medication, elevated liver enzymes, likelihood of lung transplantation during the study (investigator's opinion), or life expectancy < 2.5 years for a disease other than IPF (investigator's opinion)
Intervention details	**Outcomes**
Intervention 1. BIBF 1120 50 mg/day 2. BIBF 1120 50 mg twice per day (100 mg/day) 3. BIBF 1120 100 mg twice per day (200 mg/day) 4. BIBF 1120 150 mg twice per day (300 mg/day) 5. Placebo Dose details: a group-wise dose escalation was used initially, beginning with the lowest dose or placebo, with stepwise increases in dose for serial cohorts (states concealment of allocation maintained). Each step to the next dose cohort was reviewed by a data monitoring committee before proceeding	Primary outcomes: annual rate of decline in FVC Secondary outcomes: changes from baseline in % predicted FVC; DL_{CO}; change in oxygen saturation (measured by SpO_2); TLC; distance achieved on 6MWT, total score on SGRQ; decrease in FVC of > 10% or > 200 ml; SpO_2 decrease of more than four percentage points; incidence of acute exacerbations, survival (at 52 weeks); death from a respiratory cause; adverse events Method of assessing outcome: spirometry results were centrally reviewed by an independent third party to meet ATS/ERS criteria Overall survival referred to all randomised participants and on-treatment survival based on the number of fatal adverse events that began during the treatment or up to 14 days after treatment ended Death from a respiratory cause included all randomised participants. All deaths were adjudicated for cause of death by an independent committee unaware of treatment assignments

Intervention details	Outcomes
Dose modifications: for individual participants, the investigator was permitted to interrupt treatment and restart with the next lower dose during the trial in the event of unacceptable side effects	Safety data collected up to 14 days after administration of the last dose included in the analysis. Serious adverse events defined as fatal, life-threatening, disabling, incapacitating, requiring hospitalisation, or medically significant
Concurrent treatment: ≤ 15 mg of predisone per day or equivalent was permitted if the treatment dose had been stable for at least 8 weeks before screening	Length of follow-up: 54 weeks
Duration of treatment: 52 weeks	

Participant characteristics [mean (SD) unless stated]	50 mg/day, n = 86	100 mg/day, n = 86	200 mg/day, n = 86	300 mg/day, n = 85	Placebo, n = 85
Sex (M/F), n (%)	65/21 (75.6/24.4)	62/24 (72.1/27.9)	65/21 (75.6/24.4)	65/20 (76.5/23.5)	63/22 (74.1/25.9)
Age, years	65.3 (9.4)	64.9 (8.5)	65.1 (8.6)	65.4 (7.8)	64.8 (8.6)
Race, n (%) (self-reported)					
White	68 (79.1)	72 (83.7)	72 (83.7)	61 (71.8)	65 (76.5)
Asian	18 (20.9)	14 (16.3)	14 (16.3)	24 (28.2)	20 (23.5)
Weight, kg	78.8 (13.4)	79.0 (16.2)	76.0 (14.5)	74.9 (14.6)	77.3 (13.3)
Time since diagnosis, years	1.4 (1.3)	1.1 (1.2)	1.2 (1.2)	1.0 (1.2)	1.4 (1.5)
Surgical lung biopsy, n (%)	25 (29.1)	27 (31.4)	20 (23.3)	29 (34.1)	19 (22.4)
IPF diagnosis, n (%)					
Definite	27 (31.4)	26 (30.2)	31 (36.0)	33 (38.8)	24 (28.2)
Probable	49 (57.0)	53 (61.6)	54 (62.8)	52 (61.2)	57 (67.1)
Possible[a]	9 (10.5)	7 (8.1)	1 (1.2)	0	4 (4.7)
Definitely not[a]	1 (1.2)	0	0	0	0
FVC % predicted					
Mean (SD)	80.4 (17.8)	79.8 (15.8)	85.5 (19.2)	79.1 (18.5)	81.7 (17.6)
Median	79.8	80.4	83.0	78.1	77.6
FVC, l					
Mean (SD)	2.8 (0.8)	2.7 (0.7)	2.9 (0.8)	2.7 (0.8)	2.8 (0.8)
Median	2.8	2.7	2.8	2.7	2.7
SpO_2, %					
Mean (SD)	95.0 (2.7)	95.4 (2.2)	95.3 (2.0)	95.6 (1.7)	95.3 (2.2)
Median	96.0	96.0	96.0	96.0	96.0
DL_{CO}, mmol/minute/kPa					
Mean (SD)	3.8 (1.3)	3.9 (1.4)	3.8 (1.2)	3.7 (1.0)	3.8 (1.1)
Median	3.5	3.6	3.7	3.5	3.7
PaO_2, mmHg					
Mean (SD)	76.5 (10.7)	80.4 (20.5)	80.6 (14.4)	79.6 (13.3)	76.5 (14.1)
Median	75.8	78.4	80.0	78.3	75.0

Participant characteristics [mean (SD) unless stated]	50 mg/day, $n = 86$	100 mg/day, $n = 86$	200 mg/day, $n = 86$	300 mg/day, $n = 85$	Placebo, $n = 85$
Concomitant therapy, n (%)[b]					
Any glucocorticoid	47 (54.7)	42 (48.8)	45 (52.3)	33 (38.8)	43 (50.6)
Prednisone	19 (22.1)	14 (16.3)	17 (19.8)	18 (21.2)	21 (24.7)
SGRQ, total score	43.7 (17.5)	42.5 (17.0)	43.7 (16.6)	40.1 (18.3)	41.2 (17.9)
Symptoms	44.4 (21.5)	42.3 (21.2)	45.4 (20.0)	43.1 (25.2)	42.2 (21.6)
Activity	55.8 (19.0)	56.1 (18.4)	58.7 (18.4)	53.9 (21.6)	54.2 (22.2)
Impacts	36.2 (19.1)	34.3 (18.7)	34.1 (18.7)	30.8 (19.0)	33.1 (19.7)
Comments					

Results [mean (SE) (95% CI) unless stated]	50 mg/day, $n = 86$	100 mg/day, $n = 86$	200 mg/day, $n = 86$	300 mg/day, $n = 85$	Placebo, $n = 85$
Annual rate of decline in FVC, l	−0.17 (0.04) (−0.25 to −0.10)	−0.21 (0.04) (−0.28 to −0.14)	−0.16 (0.04) (−0.23 to −0.09)	0.06 (0.04) (−0.14 to 0.02)[c]	0.19 (0.04) (−0.26 to −0.12)
Absolute change in FVC, l	−0.18 (0.04) (−0.25 to −0.11)	−0.19 (0.04) (−0.26 to −0.12)	−0.13 (0.04)[d] (−0.20 to −0.06)	−0.06 (0.04) (−0.13 to 0.01)[e]	−0.23 (0.04) (−0.30 to −0.16)
Comments: includes those who took the therapy at least once during the treatment period					
Absolute change in FVC, % predicted	−4.58 (1.03) (−6.60 to −2.55)	−4.90 (0.98) (−6.84 to −2.97)	−3.15 (1.00) (−5.12 to −1.17)[f]	−1.04 (0.99) (−2.98 to 0.91)[g]	−6.00 (1.02) (−8.01 to −4.00)
Participants with a reduction in mean FVC of > 10% or 200 ml, n (%)	35 (41.2)	41 (47.7)	30 (35.3)	20 (23.8)[h]	37 (44.0)
Comments: unclear if this was an a priori, powered, subgroup analysis					
Absolute change in SpO_2, %	−0.86 (0.38) (−1.60 to −0.11)	−0.97 (0.36) (−1.67 to −0.27)	0.06 (0.36) (−0.65 to 0.78)[j]	−0.18 (0.36) (−0.89 to 0.53)[j]	−1.29 (0.37) (−2.03 to −0.56)
Proportion of patients with ≥ 4% decrease in SpO_2	Not reported	Not reported	Not reported	3.6%[k]	11.0%
Absolute change in TLC, l	−0.22 (0.08) (−0.38 to −0.06)	−0.10 (0.07) (−0.24 to 0.04)	−0.08 (0.07) (−0.22 to 0.06)	0.12 (0.08) (−0.03 to 0.27)[g]	−0.24 (0.08) (−0.39 to −0.09)
Change in DL_{CO}	NR	NR	NR	NR	NR
Change in 6MWT	NR	NR	NR	NR	NR
Comments: data not reported (NR) in the text or supplement report. States no significant differences between any of the groups and placebo					
Numbers reported in table, possible reporting error	$n = 87$	$n = 86$	$n = 86$	$n = 85$	$n = 87$
SGRQ total score (change from baseline)	4.67 (1.78) (1.17 to 8.16)	2.18 (1.65) (−1.07 to 5.43)	1.48 (1.66) (−1.78 to 4.75)	−0.66 (1.71) (−4.02 to 2.71)[l]	5.46 (1.73) (2.06 to 8.86)
SGRQ symptoms domain (change from baseline)	3.39 (2.51) (−1.55 to 8.34)	2.11 (2.34) (−2.48 to 6.71)	2.33 (2.35) (−2.28 to 6.94)	−3.14 (2.40) (−7.86 to 1.58)[m]	6.45 (2.45) (1.65 to 11.26)
SGRQ activity domain (change from baseline)	7.39 (1.96) (3.53 to 11.25)	3.54 (1.82) (−0.05 to 7.13)	3.00 (1.83) (−0.60 to 6.60)	0.32 (1.89) (−3.39 to 4.03)[h]	7.48 (1.91) (3.73 to 11.24)
SGRQ impacts domain (change from baseline)	3.71 (2.04) (−0.30 to 7.72)	1.73 (1.90) (−2.00 to 5.46)	0.79 (1.91) (−2.96 to 4.54)	−0.14 (1.97) (−4.00 to 3.73)	4.21 (1.99) (0.31 to 8.12)
SGRQ, % improving ≥ 4 points	*23.5%*	*27%*	32.6%[l]	29.1%[k]	16.1%
Comments: score range 0 to 100, with lower scores indicating better QoL; a minimally clinical important difference in total score = 4 points (later estimated as 5–8 points)					

Results [mean (SE) (95% CI) unless stated]	50 mg/day, n = 86	100 mg/day, n = 86	200 mg/day, n = 86	300 mg/day, n = 85	Placebo, n = 85
Incidence of acute exacerbations, n per 100 patient-years	*13*	*12.5*	*7.5*	2.4[n]	15.7
Incidence of acute exacerbations, RR (95% CI), compared with placebo	0.83 (0.36 to 1.93)	0.80 (0.34 to 1.84)	0.48 (0.18 to 1.27)	0.16 (0.03 to 0.70)	NA
Deaths from respiratory causes	9	3	2[o]	2[p]	8
Deaths from any cause	11	3	4	7	9
Comments: no statistically significant differences observed between any group and placebo					
Adverse events (any), n (%)	78 (90.6)	78 (90.7)	82 (95.3)	80 (94.1)	77 (90.6)
Adverse events occurring in > 10% in any study arm, n (%)					
Diarrhoea	9 (10.5)	17 (19.8)	32 (37.2)	47 (55.3)	13 (15.3)
Cough	11 (12.8)	17 (19.8)	20 (23.3)	8 (9.4)	17 (20.0)
Nausea	9 (10.5)	8 (9.3)	17 (19.8)	20 (23.5)	8 (9.4)
Bronchitis	11 (12.8)	16 (18.6)	7 (8.1)	9 (10.6)	11 (12.9)
Dyspnoea	7 (8.1)	14 (16.3)	13 (15.1)	6 (7.1)	11 (12.9)
Progression of IPF	11 (12.8)	7 (8.1)	9 (10.5)	4 (4.7)	11 (12.9)
Vomiting	1 (1.2)	6 (7.0)	11 (12.8)	11 (12.9)	4 (4.7)
Upper abdominal pain	6 (7.0)	10 (11.6)	2 (2.3)	10 (11.8)	3 (3.5)
Nasopharyngitis	11 (12.8)	8 (9.3)	15 (17.4)	6 (7.1)	11 (12.9)
URTI	7 (8.1)	10 (11.6)	13 (15.1)	7 (8.2)	13 (15.3)
Headache	7 (8.1)	9 (10.5)	8 (9.3)	11 (12.9)	5 (5.9)
Fatigue	4 (4.7)	5 (5.8)	8 (9.3)	9 (10.6)	7 (8.2)
Decreased appetite	3 (3.5)	4 (4.7)	4 (4.7)	13 (15.3)	0
Severe adverse events	21 (24.4)	17 (19.8)	19 (22.1)	19 (22.4)	20 (23.5)
Serious adverse events	26 (30.2)	23 (26.7)	18 (20.9)	23 (27.1)	26 (30.6)
Fatal adverse events	10 (11.6)	4 (4.7)	5 (5.8)	1 (1.2)	12 (14.1)
Adverse events requiring hospitalisation	22 (25.6)	18 (20.9)	15 (17.4)	23 (27.1)	22 (25.9)
Drug-related adverse event	24 (27.9)	30 (34.9)	41 (47.7)	55 (64.7)	25 (29.4)
Adverse events leading to discontinuation	20 (23.3)	14 (16.3)	12 (14.0)	26 (30.6)	22 (25.9)
Respiratory, thoracic and mediastinal disorders	8 (9.3)	2 (2.3)	3 (3.5)	4 (4.7)	10 (11.8)
Gastrointestinal disorders (see also below)	2 (2.3)	2 (2.3)	2 (2.3)	14 (16.5)	2 (2.4)
Infections and infestations	2 (2.3)	3 (3.5)	2 (2.3)	0	6 (7.1)
Cardiac disorders	2 (2.3)	2 (2.3)	1 (1.2)	0	6 (7.1)
Gastrointestinal adverse events, n (%)					
Any	33 (38.4)	31 (36.0)	49 (57.0)	63 (74.1)	27 (31.8)
Severe diarrhoea	1 (1.2)	1 (1.2)	2 (2.3)	4 (4.7)	0

Results [mean (SE) (95% CI) unless stated]	50 mg/day, n = 86	100 mg/day, n = 86	200 mg/day, n = 86	300 mg/day, n = 85	Placebo, n = 85
Serious diarrhoea	0	0	0	3 (3.5)	0
Serious gastrointestinal	Not reported	Not reported	Not reported	4.7%	0
Severe gastrointestinal	Not reported	Not reported	Not reported	5.9%	0
Related to treatment (any gastrointestinal)	16 (18.6)	19 (22.1)	29 (33.7)	48 (56.5)	11 (12.9)
Related to treatment (diarrhoea)	5 (5.8)	8 (9.3)		36 (42.4)	5 (5.9)
Reduced dose due to gastrointestinal adverse event	2 (2.3)	5 (5.8)	4 (4.7)	9 (10.6)	0
Discontinuation due to any gastrointestinal adverse event	2 (2.3)	2 (2.3)	2 (2.3)	14 (16.5)	2 (2.4)
Discontinuation due to diarrhoea	1 (1.2)	1 (1.2)	0	10 (11.8)	0
Mean number days with diarrhoea	70.2	39.3	138.6	85.4	6.5
Discontinuation due to nausea	Not reported	Not reported	Not reported	4.7%	0%
Discontinuation due to vomiting	Not reported	Not reported	Not reported	2.4%	1.2%

Comments: also reports rates of mild and moderate diarrhoea, not extracted here

Also reports elevations in liver enzyme levels but not extracted here

Methodology

Allocation to treatment groups: an interactive voice-response system was used to perform randomisation. Patients, investigators and the team from Boehringer Ingelheim were unaware of the treatment assignments throughout the study. When the next dose was introduced to a cohort, additional patients underwent randomisation to the previous dose group to maintain blinding between the groups

Blinding: see above regarding allocation concealment

Comparability of treatment groups: no significant differences between groups were observed

Method of data analysis: states that all efficacy analyses were based on an ITT basis. Participants were assessed within the dose group to which they were randomly assigned at the start of the study. For secondary end points the LOCF approach was used when data were not available for the entire 52-week period of assessment. A random coefficient mixed-model repeated measure was used to calculate the annual decline in FVC, with all data taken into consideration. Analysis of covariance was performed for other continuous end points, the log-rank test and Cox regression model for time to event end points, ordinal logistic regression for ordered categorical end points, Cochran–Mantel–Haenszel test for responder end points and negative binomial model for incidence rates. Safety by descriptive statistics only

Sample size/power calculation: sample size was calculated to achieve 80% power to detect a difference of 0.1 in the annual decrease in FVC

Attrition/dropout: reports numbers, reasons only reported for the total cohort

General comments

Generalisability: not stated; participants within 5 years of diagnosis and baseline FVC are suggested likely to be in mild IPF. States that population reflected the range of disease seen in clinical practice

Outcome measures: appear valid

Intercentre variability: not reported

Results [mean (SE) (95% CI) unless stated]	50 mg/day, $n = 86$	100 mg/day, $n = 86$	200 mg/day, $n = 86$	300 mg/day, $n = 85$	Placebo, $n = 85$

Conflict of interests: states that all authors designed the study and had access to the data, which were analysed by statisticians at Boehringer Ingelheim and checked by an independent consultant. The manuscript was written by medical writers, funded by the sponsor, and was reviewed, amended and edited by all authors

Quality assessment/risk of bias	Yes/no/not reported/unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Yes
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups?	Yes
(ii) If so, were they explained or adjusted for?	Unclear
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Yes
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	Yes
	Yes
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

F, female; M, male; SE, standard error; NA, not applicable.
a 21 patients with possible IPF underwent randomisation early in the study period; none was included in the later stages of the trial. One patient who did not have IPF was erroneously enrolled in the study.
b Administered at least once during the treatment period.
c $p = 0.06$ between 300 mg/day group and placebo using the closed-testing procedure for multiplicity, $p = 0.01$ with hierarchical testing. States that this difference corresponds to a reduction of 68.4% in the annual rate of decline in FVC.
   Also presents results of a sensitivity analysis, undertaken to account for discontinuations, which included all FVC assessments (baseline and all follow-up visits, including visits after discontinuation). Results are reported to be similar despite the 300 mg/day group having a higher number of discontinuations. Data not extracted here.
   Data also analysed according to the final dose (rather than randomised dose) to account for dose reductions during the trial. Results were also reported to be similar despite more participants in the 300 mg/day group undergoing dose reductions. Data not extracted here.
d $p < 0.01$ vs. placebo.
e $p = 0.001$ vs. placebo.
f $p < 0.05$ vs. placebo.
g $p < 0.001$ vs. placebo.
h $p = 0.004$ vs. placebo.
i $p = 0.005$ vs. placebo.
j $p = 0.02$ vs. placebo.
k $p = 0.03$ vs. placebo.
l $p = 0.007$ vs. placebo.
m $p = 0.003$ vs. placebo.
n $p = 0.02$.
o $p = 0.04$.
p $p = 0.06$.
Italics indicate estimated by reviewer from a figure.

Taniguchi *et al.*[65]

Study details	Participant details
Taniguchi *et al.* 2010[65]	*Number of participants:* 275 randomised: high-dose group, 110; low-dose group, 56; placebo group, 109
Country: Japan	
	267 in full analysis set: high-dose group, 108; low-dose group, 55; placebo group, 104
Design: RCT	
Number of centres: 73	*Sample attrition/dropout:* eight participants were deemed 'ineligible' after randomisation because no post-baseline data were available. Of the 'full analysis set': high-dose group: 40 (37%) withdrew (15 adverse events, eight disease progression, four acute exacerbation, 13 'other'); low-dose group: 15 (27%) withdrew (nine adverse events, two acute exacerbation, four 'other'); placebo group 31 (30%) withdrew (15 disease progression, seven adverse events, four acute exacerbation, five 'other')
Funding: supported by grants from the Japanese Ministry of Health, Labour and Welfare, and the Japanese Respiratory Society's committee for diffuse lung disease. Drug and placebo were provided by Shionogi & Co. (manufacturer in Japan)	
	Sample crossovers: not reported
	Inclusion criteria: adults 20–75 years with IPF diagnosed using ATS/ERS 2002 consensus statement and the fourth version of the clinical diagnostic criteria guidelines for IIP in Japan. HRCT scans were independently evaluated by two independent radiologists, with disagreement being resolved with a third. Probably UIP confirmed by surgical lung biopsy. Arterial oxygen saturation criteria (by pulse oximetry, SpO_2) of (1) oxygen desaturation of $\geq 5\%$ difference between resting SpO_2 and the lowest SpO_2 during a 6MET; and (2) the lowest SpO_2 during the 6MET of $\geq 85\%$ while breathing air
	Exclusion criteria: a decrease in symptoms during the preceding 6 months; use of immune-suppressants and/or oral corticosteroids at a dose of > 10 mg/day-[1] during the preceding 3 months; clinical features of IIP other than IPF; evidence of known coexisting pulmonary hypertension, asthma, tuberculosis, bronchiectasis, aspergillosis or severe respiratory infection

Intervention details	Outcomes
Intervention	Primary outcomes: change in VC to week 52
1. Pirfenidone high dose (1800 mg/day)	Secondary outcomes: PFS time, change in lowest SpO_2 during the 6MET. Also pulmonary function tests (PFTs, PaO_2, $PA – aO_2$, TLC and DL_{CO}), acute exacerbation, serum levels of markers of interstitial pneumonias, subjective/objective symptoms (cough, presence/absence of sputum and dyspnoea in daily living assessed with Hugh-Jones classification)
2. Pirfenidone low dose (1200 mg/day)	
3. Placebo	
Dose details: oral doses were increased in a stepwise manner from one tablet per dose three times daily for the first 2 weeks, then two tablets per dose for 2 weeks, then three tablets three times a day for the remaining 48 weeks. Therefore: high-dose group 600 mg/day weeks 1–2, 1200 mg/day weeks 3–4, 1800 mg/day until week 52; low-dose group 600 mg/day weeks 1–2, 600 mg/day weeks 3–4, 1200 mg/day until week 52 Dose modifications: not reported	Method of assessing outcome: disease progression was defined as death and/or $\geq 10\%$ decline in VC from baseline, or, when VC data not obtainable due to worsening respiratory symptoms, including acute exacerbation
	Acute exacerbation defined according to previous reports and revised criteria from Japan (references provided) Length of follow-up: 52 weeks
Concurrent treatment: concomitant use of corticosteroid ≤ 10 mg/day (prednisolone equivalent) was permitted. Concomitant use of immunosuppressants and other experimental agents under investigation not allowed	
Duration of treatment: 52 weeks	

Participant characteristics [mean (SD) unless stated]	High-dose (1800) pirfenidone, $n = 108$	Low-dose (1200) pirfenidone, $n = 55$	Placebo, $n = 104$	*p*-value
Sex (M/F), *n* (%)	85/23 (78.7/21.3)	47/8 (85.5/14.5)	81/23 (77.9/22.1)	0.53
Age, years	65.4 (6.2)	63.9 (7.5)	64.7 (7.3)	0.44
Smoking status (%)				
Smoker	5 (4.6)	10 (18.2)	13 (12.5)	0.07[a]
Former	81 (75.0)	33 (60.0)	70 (67.3)	
Never smoked	22 (20.4)	12 (21.8)	21 (20.2)	
Time since diagnosis, years (%)				
< 1	38 (35.2)	20 (36.4)	41 (39.4)	0.86
1–3	29 (26.9)	13 (23.6)	25 (24.0)	
≥ 3	41 (38.0)	22 (40.0)	38 (36.5)	
Prior treatment with steroids				
No (%)	99 (91.7)	49 (89.1)	98 (94.2)	0.49
Yes (%)	9 (8.3)	6 (10.9)	6 (5.8)	
Current steroid use (%)	8 (7.4)	6 (10.9)	5 (4.8)	
Surgical lung biopsy (%)	26 (24.1)	16 (29.1)	28 (26.9)	0.78
VC, ml	2400.8 (638.4)	2437.8 (684.8)	2472.3 (698.9)	0.74
VC, % predicted	77.3 (16.8)	76.2 (18.7)	79.1 (17.4)	0.57
TLC, % predicted	73.2 (16.5)	72.4 (15.6)	75.2 (15.7)	0.50
DL_{CO},% predicted	52.1 (16.8)	53.6 (19.1)	55.2 (18.2)	0.44
PaO_2 at rest, mmHg	79.8 (10.2)	81.6 (8.4)	81.0 (9.5)	0.48
$PA - aO_2$, mmHg	18.4 (11.3)	16.9 (9.6)	17.4 (9.7)	0.64
Lowest SpO_2, %	89.0 (2.3)	88.8 (2.4)	89.0 (2.2)	0.86
< 88% saturation 6MET (%)	34 (31.5)	19 (34.5)	24 (23.1)	

Comments: states that VC, TLC, DL_{CO}, PaO_2 and $PA - aO_2$ were measured for 106 participants in the high-dose group, and TLC and DL_{CO} for 103 participants in the placebo group

Results	High-dose pirfenidone, $n = 108$	Low-dose pirfenidone, $n = 55$	Placebo, $n = 104$	Difference, *p*-value high dose/low dose vs. placebo
Adjusted (LOCF) mean change in VC, l (SE)	$n = 104$ –0.09 (0.02)	$n = 54$ –0.08 (0.03)	$n = 103$ –0.16 (0.02)	High dose: 0.07 (0.03); $p = 0.0416$ Low dose: 0.09 (0.04); $p = 0.0394$
Comments: states no significant differences were shown between high-dose and low-dose groups				
Unadjusted mean change in VC, l (SE)	$n = 67$ 2.36 (0.73)	$n = 38$ 2.34 (0.71)	$n = 72$ 2.42 (0.75)	Not tested
Comments				
PFS	45/106	26/55	40/104	High dose: $p = 0.0280$ Low dose: $p = 0.0655$
Comments: data reported in figure high dose vs. low dose, $p = 0.9106$				
Change in lowest SpO_2 during the 6MET (SE)	$n = 99$ –1.70 (0.35)	$n = 53$ –0.84 (0.48)	$n = 100$ –1.53 (0.35)	High dose: –0.17 (0.50); $p = 0.7393$ Low dose: 0.69 (0.59); $p = 0.2485$
Comments				
Acute exacerbation rate	6 (5.6%)	3 (5.5%)	5 (4.8%)	NS
Comments: calculated as during the study or within 28 days of the termination of the study				
Change in TLC	$n = 99$ –0.16	$n = 52$ –0.06	$n = 99$ –0.20	High dose: $p = 0.5344$ Low dose: $p = 0.0408$
Comments: high dose vs. low dose, $p = 0.1250$				
Change in DL_{CO}	$n = 96$ –0.88	$n = 51$ –0.51	$n = 98$ –1.36	High dose: $p = 0.2317$ Low dose: $p = 0.0768$
Comments: high dose vs. low dose, $p = 0.4379$				
Change in PaO_2	$n = 98$ –2.09	$n = 54$ –3.39	$n = 103$ –3.85	High dose: $p = 0.2433$ Low dose: $p = 0.7996$
Comments: high dose vs. low dose, $p = 0.4710$				
Change in $AaDO_2$[b]	$n = 98$ 2.14	$n = 54$ 3.16	$n = 103$ 3.59	High dose: $p = 0.3325$ Low dose: $p = 0.8081$
Comments: b is $PA – aO_2 \geq$ (alveolar – arterial oxygen tension difference); high dose vs. low dose, $p = 0.5709$				

Adverse events ≥ 5% n (%)	High-dose pirfenidone n = 109	Low-dose pirfenidone n = 55	Placebo, n = 107	Difference, p-value high dose/low dose vs. placebo
Any	109 (100)	54 (98.2)	106 (99.1)	High dose: p = 0.50
				Low dose: p = 1.00
High dose vs. low dose, p = 0.34				
Photosensitivity	56 (51.4)	29 (52.7)	24 (22.4)	High dose: p < 0.01
				Low dose: p < 0.01
High dose vs. low dose, p = 1.00				
Eczema asteatotic	0	3 (5.5)	0	Low dose: p = 0.04
High dose vs. low dose, p = 0.04				
Anorexia	18 (16.5)	6 (10.9)	3 (2.8)	High dose: p < 0.01
				Low dose: p = 0.06
High dose vs. low dose, p = 0.48				
Abdominal discomfort	3 (2.8)	4 (7.3)	0	High dose: p = 0.25
				Low dose: p = 0.01
High dose vs. low dose, p = 0.23				
Dizziness	8 (7.3)	0	1 (0.9)	High dose: p = 0.04
				Low dose: p = 1.00
High dose vs. low dose, p = 0.05				
Nasopharyngitis	54 (49.5)	30 (54.5)	70 (65.4)	High dose: p = 0.02
				Low dose: p = 0.23
High dose vs. low dose, p = 0.62				
URTI	1 (0.9)	3 (5.5)	9 (8.4)	High dose: p < 0.01
				Low dose: p = 0.75
High dose vs. low dose, p = 0.11				
γ-GTP elevation	25 (22.9)	12 (21.8)	10 (9.3)	High dose: p < 0.01
				Low dose: p = 0.05
High dose vs. low dose, p = 1.00				
WBC decrease	4 (3.7)	3 (5.5)	0	High dose: p = 0.12
				Low dose: p = 0.04

High dose vs. low dose, p = 0.69

Comments: specific adverse events leading to discontinuations for each treatment group also reported but not data extracted. Total numbers are as extracted above under 'sample attrition/dropout'. No one particular adverse event appears to have been more common reason for discontinuation

Methodology

Allocation to treatment groups: states that eligible patients were allocated to the three groups in a ratio of 2 : 1 : 2 with a modified minimisation method, including some random allocation based on biased coin design to balance baseline $S\text{pO}_2$

Blinding: states placebo was matched, and described as double blind. No further details

Comparability of treatment groups: no significant differences seen in the distribution of the demographic and baseline characteristics except for smoking history (lower proportion of smokers in the high-dose group)

Adverse events ≥ 5% *n* (%)	High-dose pirfenidone *n* = 109	Low-dose pirfenidone *n* = 55	Placebo, *n* = 107	Difference, *p*-value high dose/low dose vs. placebo

Method of data analysis: the main analysis was between the high-dose and placebo groups. The low dose was intended to assess benefit-risk profiles of pirfenidone at a tapered dose. For the change in VC and lowest SpO_2 analyses an ANCOVA was used, with the respective baseline measurements as covariates. Other outcomes analysed with the least significant difference method based on one-way ANOVA. The LOCF was adopted to impute missing values if patient data were available for ≥ 4 weeks after the baseline. A mixed-model approach using available repeated measures of changes in VC was performed using a sensitivity analysis. Cumulative PFS rates were estimated using the Kaplan–Meier method and compared using a log-rank test. Incidences were compared with Fisher's exact test

Sample size/power calculation: initially the primary outcome was the lowest SpO_2 during the 6MET test; however, this was changed by the independent Data and Safety Monitoring Board before the randomisation code was broken, based on evolved knowledge of assessment with objective measurements in IPF, as well as the lack of validation in the 6MET study and difficulty in reproducibility of the SpO_2 measurement during the 6MWT. The planned sample size was 250, with 100, 50, and 100 patients in the high-dose, low-dose and placebo groups, respectively. The sizes for high dose and placebo were determined on the basis of simulations that would provide a statistical power of 0.8 to detect assumed difference of the mean changes in the lowest SpO_2 from baseline to week 52 between the two groups at a significance level of 0.1. States that the power calculated on the basis of change in VC was the same

Attrition/dropout: numbers and reasons provided for those withdrawing during treatment; however, eight were excluded after randomisation to groups (two high dose; one low dose, five placebo) because no post-baseline data were available

General comments

Generalisability: paper states that participants were assumed to have relatively mild functional impairment based on pulmonary function tests

Outcome measures: appear appropriate. Primary outcome initially intended to be the lowest SpO_2 during the 6MET but was changed following recommendation by the independent Data and Safety Monitoring Board and before the (randomisation) code was broken

Intercentre variability: not reported

Conflict of interests: statements of interests for each author are declared online (link not working)

Quality assessment/risk of bias	Yes/no/not reported/ unclear
1. Was the method used to generate random allocations adequate?	Yes
2. Was the allocation adequately concealed?	Unclear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
4. Was the care provider blinded?	Unclear
5. Was the patient blinded?	Unclear
6. Were outcome assessors blinded to the treatment allocation?	Unclear
7. (i) Were there any unexpected imbalances in dropouts between groups? (ii) If so, were they explained or adjusted for?	No
8. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Unclear
9. (i) Did the analysis include an ITT analysis? (ii) If so, was this defined?	No
10. (i) Did the analysis account for missing data? (ii) If so, were the methods appropriate?	Yes
	Yes

γ-GTP, glutamyl-transpeptidase; 6MET, 6-minute exercise test; ANCOVA, analysis of covarience; ANOVA, analysis of vanerice; $PA - aO_2$, alveolar – arterial oxygen partial pressure; PFT, pulmonary function test; SE, standard error; WBC, white blood cell.
a States $p < 0.15$; unclear what this refers to.

Appendix 6 Model code network meta-analysis

WinBUGs code (adapted from Decision Support Unit report 2) (Dias *et al.*[83]).

\# Fixed effects model for two-arm trials

```
model{ # *** PROGRAM STARTS

for(ii in 1:ns) { # LOOP THROUGH 2-ARM STUDIES

y[ii,2] ~ dnorm(md[ii,2],prec[ii,2]) # normal likelihood for 2-arm trials


var[ii,2] <- pow(se[ii,2],2) # calculate variances

prec[ii,2] <- 1/var[ii,2] # set precisions

dev[ii,2] <- (y[ii,2]-md[i,2])*(y[ii,2]-md[ii,2])*prec[ii,2] #Deviance contribution


md[ii,2] <- d[t[ii,1]] - d[t[ii,2]] # mean of treat effects distributions

}

totresdev <- sum(dev[,2]) #Total Residual Deviance

d[1]<-0 # treatment effect is zero for reference treatment

for (kk in 2:nt){ d[kk] ~ dnorm(0,.0001) } # vague priors for treatment effects


} # *** PROGRAM ENDS
```

Data (IPF – trial-level data: standardised treatment differences)

list(ns=10, nt=8)

```
t[,1]     t[,2]     y[,2]       se[,2]

2         1         0.234793937   0.461455736   #Raghu 1991

3         1         0.457405244   0.155432826   #Richeldi 2011

4         5         0.343971125   0.170963162   #Demedts 2005

4         1        −0.02487725    0.160653585   #IPFCRN 2012

6         1         0.232946081   0.230234875   #Homma 2012

7         1         0.471090122   0.208655219   #Azuma 2005

7         1         0.250474568   0.107635613   #Capacity 004

7         1         0.030939458   0.107841122   #Capacity 006

7         1         0.348717949   0.140084088   #Taniguchi 2010

8         1         0.065705107   0.149121516   #IPFCRN 2010

END
```

1=placebo

2= Azathioprine

3=BIBF 1120

4=NAC triple therapy

5=Azathioprine+placebo

6=Inhaled NAC

7=Pirfenidone

8=Sildenafil

Initial Values

#chain 1

list(d=c(NA, 0,0,0,0, 0,0,0))

#chain 2

list(d=c(NA, 1,1,1,1, 1,1,1))

Appendix 7 Validation code for the network meta-analysis

The NMA OpenBUGS code was validated using an alternative software package, SAS 9.3 (SAS Institute Inc., Cary, NC, USA). The standardised mean difference data were recreated at the arm level, and then fixed- and random-effects models were conducted using proc genmod and proc mcmc, respectively. Both SAS models produced comparable estimates to the OpenBUGS code: within ± 0.01 for the fixed effects and ± 0.012 for the random effects.

TABLE 89 Validated results using SAS: standardised mean differences

Treatment	Comparator	Fixed effects		Random effects	
		Estimate	SE	Estimate	SE
Azathioprine	Placebo	0.2449	0.2091	0.1534	0.3619
Inhaled NAC	Placebo	−0.2330	0.2302	−0.5213	0.3605
NAC triple	Placebo	−0.03326	0.1527	−0.0810	0.3062
Nintedanib	Placebo	−0.5338	0.1562	−0.2278	0.4132
Pirfenidone	Placebo	−0.2151	0.06379	−0.2430	0.1858
Sildenafil	Placebo	−0.06571	0.1491	−0.0765	0.3638
reSD				0.2433	0.2286

reSD, random-effects standard deviation; SE, standard error.

SAS code

```
/* load data */

data ipf;

    input study   trt   y   SE;

    datalines;

1      1       0    0.317595619

1      2      -0.23479   0.334775177

2      1       0    0.110424628

2      3      -0.53375   0.110424628

3      2       0    0.119577535

3      4      -0.34397   0.122186807

4      1       0    0.113965099

4      4       0.02488   0.113232197

5      1       0    0.162800641

5      5      -0.23295   0.162800641

6      1       0    0.119335975

6      6      -0.47109   0.171160526

7      1       0    0.076109872

7      6      -0.25047   0.076109872

8      1       0    0.076476539

8      6      -0.03094   0.076476539

9      1       0    0.098814857

9      6      -0.34872   0.099293382

10       1        0    0.106029022

10       7       -0.06571   0.104857394

;
```

```
/* data manipulation */

data ipf;

set ipf;

length Treatment $ 18;

if trt=1 then treatment="Placebo";

if trt=2 then treatment="Azathioprine";

if trt=3 then treatment="Nintedanib";

if trt=4 then treatment="NAC triple";

if trt=5 then treatment="Inhaled NAC";

if trt=6 then treatment="Pirfenidone";

if trt=7 then treatment="Sildenafil";

Var=SE**2;

Weight=1/Var;

run;

* Fixed effects;

Title1 "Fixed effects model";

procgenmod data=ipf ;

class Study Treatment;

model Y= Study Treatment /dist=normal noscale;

weight Weight;

lsmeans Treatment / diff=control("Placebo");

run;
```

* Random effects, Bayesian;

title1 "Random effects model";

proc mcmc data=ipf nmc=**1000000** nthin=**20** seed=**246810**

monitor=(mysd);

random Studyeffect ~general(**0**) subject=Study init=(**0**) ;

random Treat ~general(**0**) subject=Treatment init=(**0**) zero=first

monitor=(Treat);

parms logsd **0**;

prior logsd ~ general(logsd, upper=log(**5**));

mysd=exp(logsd);

random RE ~normal(**0**,sd=mysd/sqrt(**2**)) subject=_OBS_ init=(**0**);

Mu= Studyeffect + Treat +RE;

model Y ~ normal(mean=Mu, sd=SE);

run;

Appendix 8 Excluded studies systematic review of cost-effectiveness

Elfferich MD, De Vries J, Drent M. Type D or 'distressed' personality in sarcoidosis and idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2011;**28**:65–71.

Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, *et al.* The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012;**106**:1447–55.

De Vries J, Kessels BL, Drent M. Quality of life of idiopathic pulmonary fibrosis patients. *Eur Respir J* 2001;**17**:954–61.

Clark M, Cooper B, Singh S, Cooper M, Carr A, Hubbard R. A survey of nocturnal hypoxaemia and health related quality of life in patients with cryptogenic fibrosing alveolitis. *Thorax* 2001;**56**:482–6.

Swigris JJ, Stewart AL, Gould MK, Wilson SR. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;**3**:61.

Ohno S, Nakazawa S, Kobayashi A, Bando M, Sugiyama Y. Reassessment of the classification of the severity in idiopathic pulmonary fibrosis using SF-36 questionnaire. *Int Med* 2005;**44**:196–9.

Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest* 2005;**127**:284–94.

Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;**60**:588–94.

Feltrim MIZ, Rozanski A, Borges AC, Cardoso CA, Caramori ML, Pego-Fernandes P. The quality of life of patients on the lung transplantation waiting list. *Transplan Proc* 2008;**40**:819–21.

Eisenstein EL, Yow E, Zhou J, Davidson-Ray L, Anstrom KJ. Validating the use of generic quality of life instruments to estimate disease-specific measures in patients with idiopathic pulmonary fibrosis: for the IPFNET investigators. *Chest* 2009;**136**:55S.

Swigris JJ, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;**8**:77.

Lindell KO, Olshansky E, Song MK, Zullo TG, Gibson KF, Kaminski N, *et al.* Impact of a disease-management program on symptom burden and health-related quality of life in patients with idiopathic pulmonary fibrosis and their care partners. *Heart Lung* 2010;**39**:304–13.

Han MK, Swigris J, Liu L, Bartholmai B, Murray S, Giardino N, *et al.* Gender influences health-related quality of life in IPF. *Respir Med* 2010;**104**:724–30.

Lutogniewska W, Jastrzebski D, Wyrwol J, Ksiazek B, Ochman M, Kowalski K, *et al.* Dyspnea and quality of life in patients referred for lung transplantation. *Eur J Med Res* 2010;**15** (Suppl. 2):76–8.

Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010;**65**:921–6.

Coelho AC, Knorst VM, Gazzana MB, Barreto SSV. Predictors of physical and mental health-related quality of life in patients with interstitial lung disease: a multifactorial analysis. *J Bras Pneumol* 2010;**36**:562–70.

Patel AS, Siegert R, Brignall K, Keir G, Bajwah S, Desai SR, *et al.* The assessment of health related quality of life in interstitial lung disease with the King's Brief Interstitial Lung Disease questionnaire (K-BILD). *Thorax* 2011;**66**:A61.

Clark M, Cooper B, Carr A, Singh S, Cooper M, Hubbard R. Health related quality of life (HRQoL) and lung function in patients with cryptogenic fibrosing alveolitis (CFA). *Thorax* 1999;**54**:A15.

Raghu G, Brown KK, Behr J, du Bois RM, King TE. Bosentan in idiopathic pulmonary fibrosis (IPF) quality of life (QOL) results of the BUILD 1 study. *Eur Respir J* 2006;**28**:571s.

Raghu G, McDermott L, Khandker R. St George's Respiratory Questionnaire (SGRQ) for assessing quality of life (QOL) in patients with idiopathic pulmonary fibrosis (IPF). American Thoracic Society International Conference, San Francisco, CA, USA, 18–23 May 2007.

Krishnan V, McCormack MC, Mathai SC, Agarwal S, Richardson B, Horton MR, *et al.* Sleep quality and health-related quality of life in idiopathic pulmonary fibrosis. *Chest* 2008;**134**:693–8.

Santana M-J, Feeny D, Jackson K, Weinkauf J, Lien D. Improvement in health-related quality of life after lung transplantation. *Can Respir J* 2009;**16**:153–8.

Richeldi L, Brown KK, Costabel U, Flaherty KR, Kim DS, Noble PW. Treatment with BIBF 1120 reduces acute exacerbations and improves quality of life in patients with IPF: results from the Tomorrow Study. *Am J Respir Crit Care Med* 2011;**183**:A3809.

Appendix 9 Data extraction from systematic review of cost-effectiveness

Study characteristics

Reference

Hagaman 2010.[99]

Health technology

TPMT testing genotypic assay to identify IPF patients at risk for developing leukopenia with azathioprine.

Interventions and comparators

What interventions/strategies were included?

Azathioprine, NAC and steroids (prednisolone) with TPMT testing.

Azathioprine, NAC and steroids (prednisolone) without TPMT testing.

Conservative therapy.

Was a no treatment/supportive care strategy included?

Yes – conservative therapy which consisted of no specific pharmacologic treatment for IPF and utilised only supportive measures.

Describe interventions/strategies

In the testing strategy, TPMT activity was assessed before initiating therapy and defined as normal, intermediate or low 'according to convention' (references provided in paper). Patients with normal activity were started on azathioprine, prednisolone and NAC at standard doses. Those with intermediate activity were started on azathioprine at a reduced dose along with prednisolone and NAC, consistent with previous recommendations (references provided in paper). In patients with absent TPMT activity, patients were started on conservative therapy. Actual dosages are not stated.

Research question

What are the stated objectives of the evaluation?

To determine whether or not TPMT testing before initiation of azathioprine, NAC and steroids is a cost-effective treatment in IPF when compared with no TMPT testing or conservative therapy.

Study type

Cost–utility.

Study population

What definition was used for [condition]? What are the characteristics of the baseline cohort for the evaluation?

No definitions of IPF or leukopenia are given. References given for classification of TPMT activity into normal, intermediate and absent, although actual basis is not outline in the report itself. Baseline cohort includes patients with IPF who are considered appropriate for treatment with azathioprine, NAC and prednisolone or conservative therapy. No other details are provided.

Institutional setting

Where is/are the intervention(s) being evaluated usually provided?

Not stated; however, TPMT and treatment regimens will usually be provided in secondary care.

Country/currency

Has a country setting been provided for the evaluation? What currency are costs expressed in and does the publication give the base year to which those costs relate?

Costs expressed in 2007 US$.

Funding source

None stated.

Analytical perspective

What is the perspective adopted for the evaluation [health service, health and personal social services, third party payer, societal (i.e. including costs borne by individuals and lost productivity)]?

Not stated; however, costs included focus on health service costs.

Effectiveness

Were the effectiveness data derived from a single study, a review/synthesis of previous studies or expert opinion? Give the definition of treatment effect used in the evaluation. Give the size of the treatment effect used in the evaluation.

No systematic review was undertaken. Two studies on patients with inflammatory bowel disease were used to provide the risk of developing leukopenia by TPMT status and their estimates were combined by weighted average. These estimates showed that 5% of patients who had normal TMPT activity, 21.4% intermediate activity and 100% no activity developed leukopenia. Risk of developing leukopenia for intermediate TPMT status assumed to be halved at reduced dose of azathioprine (10%), based on two studies. Efficacy of interventions were obtained from the Infigenia RCT, with 37% of those receiving azathioprine, NAC and prednisolone incurring disease progression at 12 months compared with 51% in those receiving azathioprine and prednisolone.

Intervention costs

Were the cost data derived from a single (observational) study, a review/synthesis of previous studies or expert opinion? Were the methods for deriving these data adequately described (give sources if using data from other published studies)? List the direct intervention costs and other direct costs used in the evaluation – include resource estimates (and sources for these estimates, if appropriate) as well as sources for unit costs used.

No systematic review was undertaken. Costs tended to originate from routinely available sources. Drug costs obtained from an online resource (www.drugstore.com). Average Medicare reimbursement for Current Procedural Terminology or Diagnosis-Related-Group was used as a proxy for the cost of procedures, office visits and hospitalizations. Source of cost of TPMT assay was not made clear. The basis for resources use was not stated.

Direct intervention costs included:

- cost of TMPT assay: US$300
- cost of azathioprine, NAC and steroids (per month): US$297
- cost of azathioprine at reduced dose, NAC and steroids (per month): US$240
- cost of conservative therapy (per month): US$66
- complicated leukopenia: US$10,136
- complicated leukopenia leading to death: US$14,666
- uncomplicated leukopenia: US$423
- cost of IPF disease progression: US$14,773.

Detailed description of microcosting is provided.

Indicate the source for individual cost values (if appropriate).

Indirect costs (costs due to lost productivity, unpaid inputs to patient care)
Were indirect costs included?

Indirect costs were not included.

Health state valuations/utilities (if study uses quality of life adjustments to outcomes)
Were the utility data derived from a single (observational) study, a review/synthesis of previous studies or expert opinion. Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

QoL for 'progression of IPF' state derived from single study. QoL in leukopenia and complicated leukopenia states derived from three studies; derivation not discussed in text.

List the utility values used in the evaluation.

Utility values used for the base case (range for sensitivity analysis):

- progression of IPF: 0.63 (0.5 to 0.9)
- leukopenia: 0.95 (0.92 to 0.98)
- complicated leukopenia: 0.76 (0.66 to 0.86).

Sources of the data are provided, although there is no discussion as to the nature of the data (e.g. patient group).

Indicate the source for individual cost values (if appropriate).

Modelling

If a model was used, describe the type of model used (e.g. Markov state transition model, discrete event simulation). Was this a newly developed model or was it adapted from a previously reported model? If an adaptation, give the source of the original. What was the purpose of the model (i.e. why was a model required in this evaluation)? What are the main components of the model (e.g. health states within a Markov model)? Are sources for assumptions over model structure (e.g. allowable transitions) reported – list them if reported.

A newly developed decision tree model was used to assess cost–utility of the addition of TPMT screening prior to the treatment of IPF patients with azathioprine, NAC and prednisolone compared with not screening and a conservative therapy.

Decision node:

- whether to screen, not screen, or opt for conservative therapy.

Chance nodes

- TPMT status (normal, intermediate or absent activity)
- leukopenia or no leukopenia
- complicated leukopenia or no complications form leukopenia
- death from leukopenia or survive leukopenia and change to conservative therapy
- progression of IPF to either death or alive with either progression or no progression.

The rationale for the model and its assumptions are not discussed or referenced.

Extract transition probabilities for [natural history/disease progression] model and show sources (or refer to table in text).

Prevalence of TPMT activity [base case (range for sensitivity analysis)]:

- normal (high) 87.6% (85.6% to 90%)
- intermediate 11.9% (7.8% to 13.5%)
- low (absent) 0.5% (0 to 3%).

1-year cumulative probabilities of developing leukopenia with standard azathioprine dosing:

- normal TPMT activity 1% (0.5% to 1.5%)
- intermediate TMPT activity 21.4% (16.8% to 26%)
- absent TMPT activity 100% (not tested).

1-year cumulative probabilities of developing leukopenia with reduced azathioprine dosing:

- normal TPMT activity 1% (0.5% to 5%)
- intermediate TMPT activity 10% (5.4% to 14.6%)
- absent TMPT activity – not applicable.

1-year cumulative probabilities of miscellaneous complications on azathioprine: 2.5% (1% to 4%) in patients with leukopenia:

- probability of complicated leukopenia 16% (10% to 22%)
- probability of death 8% (2% to 16%).

1-year cumulative probability of disease progression

- azathoprine, NAC, steroids 37% (31.6% to 42.4%)
- conservative therapy 51% (30% to 65%).

Excess mortality due to IPF per year: 9% (5% to 20%).

Life expectancy with IPF after diagnosis: 3 years (3 to 10 years).

Sources of the data are provided, with some limited discussion as to the rationale for their use.

What is the model time horizon?

1 year, with all events modelled occurring at an average of 0.5 years (including treatment-related adverse events, disease progression and death).

What, if any, discount rates have been applied in the model? Same rate for costs and outcomes?

Not stated.

Results/analysis

What measure(s) of benefit were reported in the evaluation?

> Marginal QALYs.

Provide a summary of the clinical outcome/benefits estimated for each intervention/strategy assessed in the evaluation.

> Conservative therapy – 2.50 QALYs.
>
> Azathioprine without TPMT testing – 2.61 QALYs.
>
> Azathioprine with TPMT testing – 2.62 QALYs.

Provide a summary of the costs estimated for each intervention/strategy assessed in the evaluation.

> Conservative therapy $9691.
>
> Azathioprine without testing $15,802.
>
> Azathioprine with testing $15,818.

Synthesis of costs and benefits – are the costs and outcomes reported together (e.g. as cost-effectiveness ratios)? If so, provide a summary of the results.

> Outcomes were reported as ICERs.
>
> Marginal cost-effectiveness ratio with extended dominance ($/QALY gained)
>
> * azathioprine, NAC and steroids with TPMT testing vs. conservative therapy: $49,245
> * azathioprine, NAC and steroids without TPMT testing is extendedly dominated by azathioprine, NAC and steroids with TPMT testing.
>
> Marginal cost-effectiveness ratio without extended dominance ($/QALY gained)
>
> * azathioprine, NAC and steroids without TPMT testing vs. conservative therapy: $49,245.
> * azathioprine, NAC and steroids with TPMT testing vs. azathioprine, NAC and steroids without TPMT testing: $29,663.

Give results of any statistical analysis of the results of the evaluation.

Not applicable.

Was any sensitivity analysis performed – if yes, what type(s) (i.e. deterministic (one-way, two-way, etc.) or probabilistic)?

Sensitivity analyses were performed on all parameters. Although it is not specified whether analyses were deterministic or probabilistic, selective results of one-way and two-way analyses are presented.

What scenarios were tested in the sensitivity analysis? How do these relate to structural uncertainty (testing assumptions over model structure such as relationships between health states), methodological uncertainty (such as choices of discount rate or inclusion of indirect costs) or parameter uncertainty (assumptions over values of parameters in the model, such as costs, QoL or disease progression rates)?

There is limited discussion concerning the sensitivity analyses, with only three scenarios assessing parameter uncertainty that had the greatest impact (not defined) discussed, including:

- frequency of abnormal TPMT alleles
- probability of developing leukopenia as a consequence of treatment with reduced dose of azathioprine, patients with intermediate TPMT status
- treatment efficacy.

Give a summary of the results of the sensitivity analysis – did they differ substantially from the base-case analysis? If so, what were the suggested causes?

ICER of testing vs. no testing increases dramatically when the frequency of abnormal alleles is < 12%.

As the probability of leukopenia increases above the base case value of 10%, TPMT testing becomes increasingly expensive. Above a probability of 12% testing is no longer 'cost-effective'.

As the rate of disease progression in 1 year on conservative therapy increases, TPMT testing is favoured at increasingly higher rates of disease progression on azathioprine, NAC and steroids.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> TPMT testing before starting a regimen of azathioprine in combination with NAC and steroids is cost-effective in patients with IPF.

What are the implications of the evaluation for practice?

> Little implication for practice, as TPMT screening before initiating azathioprine is already common.

Southampton Health Technology Assessments Centre commentary

Selection of comparators:

Appropriate.

Validity of estimate of measure of benefit:

Given that the focus of the evaluation is the benefit of the addition of TMPT testing prior to treatment with azathioprine in combination with NAC and steroids (rather than the effectiveness of the treatment itself), QoL appears appropriate. Utilities associated with various disease states are presented but not discussed in the paper. It is not clear how valid the specific utilities used are.

Validity of estimate of costs:

Although the estimates for costs appear reasonable and detailed descriptions of microcosts have been provided, there is limited discussion regarding the resources used.

Appendix 10 Excluded studies: systematic review of health-related quality of life

Efferich MD, De Vries J, Drent M. Type D or 'distressed' personality in sarcoidosis and idiopathic pulmonary fibrosis. *Sarcoidosis Vasculitis Diffuse Lung Dis* 2011;**28**:65–71.

Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, *et al.* The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012.

De Vries J, Kessels BL, Drent M. Quality of life of idiopathic pulmonary fibrosis patients. *Eur Respir J* 2001;**17**:954–61.

Clark M, Cooper B, Singh S, Cooper M, Carr A, Hubbard R. A survey of nocturnal hypoxaemia and health related quality of life in patients with cryptogenic fibrosing alveolitis. *Thorax* 2001;**56**:482–6.

Swigris JJ, Stewart AL, Gould MK, Wilson SR. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. *Health Qual Life Outcomes* 2005;**3**:61.

Ohno S, Nakazawa S, Kobayashi A, Bando M, Sugiyama Y. Reassessment of the classification of the severity in idiopathic pulmonary fibrosis using SF-36 questionnaire. *Intern Med* 2005;**44**:196–9.

Swigris JJ, Gould MK, Wilson SR. Health-related quality of life among patients with idiopathic pulmonary fibrosis. *Chest* 2005;**127**:284–94.

Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;**60**:588–94.

Feltrim MIZ, Rozanski A, Borges AC, Cardoso CA, Caramori ML, Pego-Fernandes P. The quality of life of patients on the lung transplantation waiting list. *Transplant Proc* 2008;**40**:819–21.

Eisenstein EL, Yow E, Zhou J, Davidson-Ray L, Anstrom KJ. Validating the use of generic quality of life instruments to estimate disease-specific measures in patients with idiopathic pulmonary fibrosis: for the IPFNET investigators. *Chest* 2009;**36**:55S.

Swigris JJ, Wilson SR, Green KE, Sprunger DB, Brown KK, Wamboldt FS. Development of the ATAQ-IPF: a tool to assess quality of life in IPF. *Health Qual Life Outcomes* 2010;**8**:77.

Lindell KO, Olshansky E, Song MK, Zullo TG, Gibson KF, Kaminski N, *et al.* Impact of a disease-management program on symptom burden and health-related quality of life in patients with idiopathic pulmonary fibrosis and their care partners. *Heart Lung* 2010;**39**:304–13.

Han MK, Swigris J, Liu L, Bartholmai B, Murray S, Giardino N, *et al.* Gender influences health-related quality of life in IPF. *Respir Med* 2010;**104**:724–30.

Lutogniewska W, Jastrzebski D, Wyrwol J, Ksiazek B, Ochman M, Kowalski K, *et al.* Dyspnea and quality of life in patients referred for lung transplantation. *Eur J Med Res* 2010;**15**(Suppl. 2):76–8.

Yorke J, Jones PW, Swigris JJ. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010;**65**:921–6.

Coelho AC, Knorst VM, Gazzana MB, Barreto SSV. Predictors of physical and mental health-related quality of life in patients with interstitial lung disease: a multifactorial analysis. *J Bras Pneumol* 2010;**36**:562–70.

Patel AS, Siegert R, Brignall K, Keir G, Bajwah S, Desai SR, *et al.* The assessment of health related quality of life in interstitial lung disease with the King's Brief Interstitial Lung Disease questionnaire (K-BILD). *Thorax* 2011;**66**:A61.

Clark M, Cooper B, Carr A, Singh S, Cooper M, Hubbard R. Health related quality of life (HRQoL) and lung function in patients with cryptogenic fibrosing alveolitis (CFA). *Thorax* 1999;**54**:A15.

Raghu G, Brown KK, Behr J, du Bois RM, King TE. Bosentan in idiopathic pulmonary fibrosis (IPF) quality of life (QOL) results of the BUILD 1 study. *Eur Respir J* 2006;**28**:571s.

Raghu G, McDermott L, Khandker R. St George's Respiratory Questionnaire (SGRQ) for assessing quality of life (QOL) in patients with idiopathic pulmonary fibrosis (IPF). American Thoracic Society International Conference, 18–23 May 2007, San Francisco, CA, USA.

Krishnan V, McCormack MC, Mathai SC, Agarwal S, Richardson B, Horton MR, *et al.* Sleep quality and health-related quality of life in idiopathic pulmonary fibrosis. *Chest* 2008;**134**:693–8.

Santana M-J, Feeny D, Jackson K, Weinkauf J, Lien D. Improvement in health-related quality of life after lung transplantation. *Can Respir J* 2009;**16**:153–8.

Richeldi L, Brown KK, Costabel U, Flaherty KR, Kim DS, Noble PW. Treatment with BIBF 1120 reduces acute exacerbations and improves quality of life in patients with IPF: results from the Tomorrow Study. *Am J Respir Crit Care Med* 2011;**183**:A3809.

Berry CE, Drummond MB, Han MK, Li D, Fuller C, Limper AH, *et al.* Relationship between lung function impairment and health-related quality of life in COPD and interstitial lung disease. *Chest* 2012;**142**:704–11.

Huppmann P, Sczepanski B, Boensch M, Winterkamp S, Schonheit-Kenn U, Neurohr C, *et al.* Effects of in-patient pulmonary rehabilitation in patients with interstitial lung disease. *Eur Respir J* 2012.

Shulgina L, Cahn AP, Chilvers ER, Parfrey H, Clark AB, Wilson EC, *et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: a randomised controlled trial. *Thorax* 2013;**68**:155–62.

Han MK, Bach DS, Hagan PG, Yow E, Flaherty KR, Toews GB, *et al.* Sildenafil preserves exercise capacity in patients with idiopathic pulmonary fibrosis and right-sided ventricular dysfunction. *Chest* 2013;**143**:1699–708.

Johnson-Warrington V, Williams J, Bankart J, Steiner M, Morgan M, Singh S. Pulmonary rehabilitation and interstitial lung disease: aiding the referral decision. *J Cardiopulm Rehabil Prev* 2013;**33**:189–95.

Appendix 11 Data extractions from systematic review of health-related quality of life

Reference

Baddini Martinez 2002.[111]

Study characteristics

Research question
What are the stated objectives of the study?

To investigate how some commonly used dyspnoea scales correlate with measurements of HRQoL in IPF patients.

Describe the type of study and study design.

Cross-sectional study/single-cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosed by surgical lung biopsy or HRCT.

What are the characteristics of the baseline cohort for the evaluation?

Age	58.6
Sex, M/F	18/12
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 30$
	FVC per cent predicted 61.9% (standard error of the mean 3.24)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Brazil, outpatient clinic.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL data obtained in this single study.

Results
Summarise the results.

All dyspnoea scales showed significant correlations with both physical and mental HRQoL domains. The highest Spearman correlation coefficients were obtained using the BDI

SF-36 scores	Mean	SEM
Physical function	40.83	4.41
Role physical	44.17	8.86
Pain index	78.57	4.40
General health	53.07	4.33
Vitality	50.50	5.10
Social function	60.83	7.26
Role emotional	64.44	8.74
Mental health	59.47	4.15
SEM, standard error of the mean.		

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> Dyspnoea ratings correlate with HRQoL in IPF patients and can be used as a measure of HRQoL in special circumstances.

What are the implications of the study for the model?

> Model might possibly use dyspnoea measurements as a proxy for HRQoL in patients with IPF. Baseline SF-36 may be useful in the absence of utility data as may be able to map this to the EQ-5D. Limited by the small sample size.

Reference

De Vries 2000.[117]

Study characteristics

Research question
What are the stated objectives of the study?

To identify aspects of QoL that are relevant to this population and to establish which measure is preferable to assess these aspects.

Describe the type of study and study design.

Three focus groups/cross-sectional survey.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) From three hospitals, uncertain diagnostic criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	61.1
Sex, M/F	4/6
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 10$
	FVC % predicted not reported
QoL instrument	SGRQ and WHOQoL-100 (translated to Dutch)
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting

What is the country and setting for the evaluation?

> The Netherlands; university hospital setting assumed by reviewers.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, expert opinion?

> QoL derived in this single observational study, prior to focus groups.

Results

Summarise the results

> Results indicate aspects relevant to IPF patients QoL – mobility, working capacity, energy, etc.
>
> | SGRQ scores | Mean | SD |
> | --- | --- | --- |
> | Activity | 56.0 | 20.2 |
> | Impact | 38.7 | 22.7 |
> | Symptoms | 46.1 | 20.8 |
> | Total | 44.5 | 17.8 |
> | WHOQOL-100 overall | 12.5 | 2.3 |
>
> Also reports data from six domains and 24 facets of the WHOQOL but not data extracted as not per inclusion criteria

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

WHOQOL-100, with additional social support questionnaire, was preferred to the SGRQ, as it mentioned all aspects of QoL mentioned by the IPF patients.

What are the implications of the study for the model?

SGRQ is not a detailed instrument for capturing IPF QoL. May provide useful data if no utility data identified, although small numbers limit the reliability and uncertainty over the diagnosis of IPF as not reported.

Reference

Horton and colleagues 2012,[55] and Lechtzin and colleagues 2013.[121]

Study characteristics

Research question
What are the stated objectives of the study?

To assess the effect of thalidomide on cough in patients with IPF/to validate the CQLQ.

Describe the type of study and study design.

RCT, crossover design.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Patients with IPF diagnosed with IPF by HRCT or lung biopsy, and a symptom duration of between 3 months and 5 years and chronic cough (more than 8 weeks' duration) and aged over 50 years.

What are the characteristics of the baseline cohort for the evaluation?

Age	67.6 years
Sex, M/F	18/5
Race (if appropriate)	White: 91.2%
	Black: 4.4%
	Hispanic: 4.4%
Indication/disease	IPF
Other characteristics (sample size)	Number randomised: 23
	FVC per cent predicted: 70.4% (SD 13.7)
QoL instrument	SGRQ
Utility values, Y/N	N
Treatment effect, if reported	CQLQ

Country/setting
What is the country and setting for the evaluation?

USA; university setting.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

Single study.

Results
Summarise the results.

Results from the combined thalidomide and placebo group at baseline only data extracted as no baseline data for the placebo group alone were presented:

SGRQ	Mean	SD
Total	57.4	18.8
Symptom	67.7	19.7
Impact	48.1	20.7
Activity	64.3	22.7

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> Overall QoL measured by the SGRQ is significantly improved with thalidomide.

What are the implications of the study for the model?

> SGRQ data could be mapped to EQ-5D if no relevant data were available.

Reference

Jastrzebski 2005.[112]

Study characteristics

Research question
What are the stated objectives of the study?

To compare changes in the QoL in patients with IPF at point of meeting qualification criteria for lung transplant and 12 months later.

Describe the type of study and study design.

Cohort study with reference group (emphysema) controls.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Patients with disease of interest who also received lung transplant, diagnosed by ATS/ERS criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	48.3
Sex, M/F	11/5
Race (if appropriate)	Not reported
Indication/disease	IPF, eligible for lung transplantation (reference group patients with emphysema)
Other characteristics (sample size)	$n = 16$
	FVC per cent predicted reported (44.0%, SD 15.3)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting

What is the country and setting for the evaluation?

Poland; hospital.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

Single study.

Results

Summarise the results.

A year after the qualification for lung transplantation the results of the SF-36 questionnaire were not different in the IPF patients, compared with their baseline scores.

SF-36 mean	Baseline	12 months
Physical function	45	47
Role physical	43	25
Bodily pain	62	64
General health	28	28
Vitality	38	42
Social function	58	63
Role emotional	65	68
Mental health	50	56
Physical overall	35	33
Mental overall	42	45

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

SF-36 is a sensitive tool to assess the QoL in IPF patients qualified for lung transplantation.

What are the implications of the study for the model?

Data may be useful if no utility data identified as may be able to be mapped onto the EQ-5D, population with severe IPF based on FVC per cent predicted. Limited by small numbers and no measure of variance provided.

Reference

Jastrzebski 2006.[62]

Study characteristics

Research question
What are the stated objectives of the study?

To examine the effects of pulmonary rehabilitation in patients suffering from IPF on QoL and dyspnoea.

Describe the type of study and study design.

Cohort study, pre and post design.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(iv) Patients with the disease of interest and other patients treated for ILD; diagnostic criteria for IPF not provided.

What are the characteristics of the baseline cohort for the evaluation?

Age	48.7
Sex, M/F	19/12
Race (if appropriate)	
Indication/disease	13 patients had IPF
Other characteristics (sample size)	31 patients finished rehabilitation programme ($n = 38$ qualified)
	FVC % predicted not reported
QoL instrument	SF-36 and SGRQ (Polish versions)
Utility values, Y/N	N
Treatment effect, if reported	Change in dyspnoea severity before and after rehabilitation

Country/setting
What is the country and setting for the evaluation?

Poland; hospital outpatients.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL obtained from this study, before and after the 6-week pulmonary rehabilitation program.

Results
Summarise the results.

Rehabilitation caused the scores of all SF-36 components to increase, most significantly ($p < 0.05$). Two SGRQ domains showed significant improvement. Data reported in figures only and not extracted given other ILD participants constitute the majority of cases in this study.

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis

> The rehabilitation programme improved the QoL of included patients.

What are the implications of the study for the model?

> Not entirely relevant to model as it appears that very few patients had IPF and data in figures only.

Reference

Kozu 2011.[114]

Study characteristics

Research question

What are the stated objectives of the study?

1. To evaluate the effects of pulmonary rehabilitation on dyspnoea, exercise capacity and health status in subjects with IPF with health limitations.
2. To compare the responses in subjects with IPF with the changes seen in a control group of COPD subjects.

Describe the type of study and study design.

Prospective non-randomised pre- and post-cohort study with COPD controls.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosis based on ATS/ERS criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	67.5
Sex, M/F	37/8
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF
Other characteristics (sample size)	$n = 45$
	FVC per cent predicted reported (68.6%, SD 16)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Change in SF-36 score, 6 MWD, muscle force, etc.

Country/setting

What is the country and setting for the evaluation?

Japan/outpatient programme.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL obtained from this study – taken at baseline, immediately following rehabilitation programme (≈8 weeks) and at 6 months thereafter

Mean (SD) scores SF-36 at baseline

	Mean	SD
Physical function	35.7	18.7
Role physical	33.9	21.7
Bodily pain	62.4	30.3
General health	34.7	19.9
Vitality	38.3	21.3
Social function	48.3	23.7
Role emotional	36.5	30.1
Mental health	50.0	18.7

Also reports SF-36 for participants who completed the programme at baseline, at 8 weeks ($n = 36$) and 6 months ($n = 30$) following pulmonary rehabilitation. Not extracted here

Results

Summarise the results.

No changes in SF-36 scores in IPF patients although other improvements seen (muscle force, dyspnoea, etc.).

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

Pulmonary rehab in IPF does not improve health status as measured by SF-36; a pulmonary rehab program specifically tailored to the needs of IPF should be investigated in future.

What are the implications of the study for the model?

The baseline measurements of the SF-36 may be useful if no utility data are identified as these can be mapped to the EQ-5D if required. The population were approximately moderate disease severity. Small sample size limit the reliability of the data.

Reference

Kozu 2011.[113]

Study characteristics

Research question
What are the stated objectives of the study?

> To investigate differences in response to pulmonary rehabilitation in subjects with IPF who were grouped according to their level of disability, as categorised by MRC dyspnoea scale.

Describe the type of study and study design.

> Prospective non-randomised and uncontrolled, pre- and post-cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

> (ii) Patients with a diagnosis of IPF (based on the International Consensus statement) who were referred for pulmonary rehabilitation.

What are the characteristics of the baseline cohort for the evaluation?

Age	67.5
Sex M/F	46/19
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF and referred for pulmonary rehabilitation
Other characteristics (sample size)	$n = 65$
	FVC per cent predicted reported by dyspnoea grade, ranged 51–83%
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Change in 6 MWD, %
	Change in SF-36

Country/setting
What is the country and setting for the evaluation?

Japan/hospital outpatients.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

SF-36 obtained from this study – completed at baseline and immediately following the 8-week rehabilitation programme. Baseline scores reported by dyspnoea grade.

Results
Summarise the results.

Baseline characteristics

MRC dyspnoea grade	2, $n = 16$	3, $n = 17$	4, $n = 17$	5, $n = 15$
FVC per cent predicted, mean (SD)	83 (11)	67 (13)	60 (16)	51 (11)
SF-36, mean (SD)				
Physical function	55.3 (7.2)	34.1 (18.4)	20.3 (7.0)	16.0 (9.1)
Role physical	55.9 (15.9)	22.4 (17.3)	23.2 (13.4)	19.6 (10.3)
Bodily pain	66.5 (25.1)	57.2 (29.0)	65.6 (29.1)	65.6 (28.4)
General health	50.9 (11.0)	35.8 (21.5)	24.1 (16.8)	19.1 (10.7)
Vitality	54.7 (11.7)	37.9 (21.5)	26.5 (18.0)	19.6 (15.3)
Social function	62.5 (18.8)	42.6 (27.6)	36.0 (15.2)	30.0 (14.8)
Role emotional	66.7 (15.2)	47.1 (28.2)	30.9 (21.4)	19.4 (15.3)
Mental health	61.6 (14.3)	42.9 (20.8)	41.8 (17.2)	35.0 (12.0)

Differential improvement in SF-36 by disability level in response to rehabilitation programme (data not extracted)

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> Disability level has an effect on response to pulmonary rehabilitation in patients with IPF, when response is measured in terms of 6MWD and SF-36.

What are the implications of the study for the model?

> Overall response to treatment will vary by mix of disability levels in IPF population. Scores may be of value to the model if no utility data are identified.

Reference

Lutogniewska *et al.* 2010.[127]

Study characteristics

Research question
What are the stated objectives of the study?

To determine the correlation between QoL and dyspnoea in patients awaiting lung transplantation.

Describe the type of study and study design.

Single-cohort study with subgroup data.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Included patients fulfilling the criteria for lung transplant with IPF, COPD or IIP. Subgroup data for IPF patients reported here.

What are the characteristics of the baseline cohort for the evaluation?

Age	Not reported
Sex, M/F	Not reported
Race (if appropriate)	Not reported
Indication/disease	IPF, no definition of the diagnostic criteria used provided
Other characteristics (sample size)	$n = 30$
	FVC per cent predicted not reported
QoL instrument	SGRQ, SF-36
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Poland.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

Single study.

Results
Summarise the results.

	Mean	SD
SGRQ		
Total score	64.0	19.0
Symptoms	58.7	20.3
Activity	70.4	20.1
Impact	65.4	24.8
SF-36		
Physical function	19.1	17.0
Role physical	20.7	23.2
Bodily pain	47.3	25.9
General health	28.5	10.7
Vitality	38.7	19.8
Social function	33.4	26.0
Role emotional	40.3	36.9
Mental health	49.2	22.8
Physical component score	25.9	7.8
Mental component score	42.4	14.2

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Provides a reference for the methodological details (is an included study).

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

The SF-36 and SGRQ are sensitive tools in the assessment of the QoL of patients referred for lung transplant.

What are the implications of the study for the model?

If no utility data available these results could be mapped.

Reference

Martinez 2000.[109]

Study characteristics

Research question

What are the stated objectives of the study?

To validate the use of the SF-36 questionnaire in patients with IPF and no comorbidity.

Describe the type of study and study design.

Cross sectional with matched healthy controls, matched for sex and age.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) With no significant comorbidity, diagnosis by lung biopsy or HRCT, and healthy controls.

What are the characteristics of the baseline cohort for the evaluation?

Age	58.29
Sex M/F	20/14
Race (if appropriate)	Not reported
Indication/disease	IPF no comorbidity
Other characteristics (sample size)	$n = 34$
	FVC per cent predicted reported (62.41%, SD 2.96)
QoL instrument	SF-36 validated for Portuguese language
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Brazil; outpatient clinic.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL obtained from this study only.

Results
Summarise the results.

Patients with IPF scored significantly worse than control subjects for most of the SF-36 components apart from pain index

SF-36	Mean	SEM
Physical function	42.79	4.40
Role physical	44.12	8.11
Bodily pain	76.91	4.16
General health	53.50	3.90
Vitality	50.44	4.88
Social function	60.29	6.69
Role emotional	60.78	8.26
Mental health	57.33	4.07

SEM, standard error of the mean.

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

Patients with IPF have significant impairment of HRQoL in both physical and psychological functioning. Dyspnoea is the most important factor in influencing QoL.

What are the implications of the study for the model?

SF-36 is a valid instrument to measure HRQoL in IPF patients, limited by small numbers; however, may be useful if no utility data are identified. Moderate severity cases.

Reference

Mermigkis *et al.* 2013.[116]

Study characteristics

Research question

What are the stated objectives of the study?

To assess the results of effective continuous positive airway pressure (CPAP) therapy in terms of sleep quality and overall QoL in IPF patients with moderate to severe obstructive sleep apnoea.

Describe the type of study and study design.

Single cohort study, before and after.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Newly diagnosed, histologically proven IPF or fulfilled consensus criteria for IPF and had features compatible with moderate to severe obstructive sleep apnoea.

What are the characteristics of the baseline cohort for the evaluation?

Age	67.1 (SD 7.2) years
Sex, M/F	10/2
Race (if appropriate)	Not reported
Indication/disease	IPF and moderate to severe obstructive sleep apnoea
Other characteristics (sample size)	$n = 12$ (also present characteristics of 11 others who had IPF but did not fulfil the criteria for obstructive sleep apnoea)
	FVC per cent predicted: 77.1% (SD 18.1)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Various sleep rating scales used to assess the effectiveness of CPAP

Country/setting
What is the country and setting for the evaluation?

Greece.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

A single study.

Results
Summarise the results.

SF-36 overall score only presented. At baseline this was 63.2 (SD 13.9).

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications
Give a brief summary of the authors' conclusions from their analysis.

CPAP can lead to improvements in sleep and overall QoL.

What are the implications of the study for the model?

Minimal SF-36 data presented.

Reference

Nishiyama *et al.* 2012.[122]

Study characteristics

Research question
What are the stated objectives of the study?

To determine the prognostic significance of HRQL (HRQoL) scores in IPF assessed with the SGRQ.

Describe the type of study and study design.

Retrospective cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) With diagnosis by ATS/ERS criteria, and newly diagnosed having not received any other treatments.

What are the characteristics of the baseline cohort for the evaluation?

Age	66.3 years
Sex, M/F	77/10
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	*n* = 87
	FVC per cent predicted: 75.0 (SD 19.2)
QoL instrument	SGRQ
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting

What is the country and setting for the evaluation?

Japan.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

A single study.

Results

Summarise the results.

SGRQ at baseline

	Mean	SD
Symptoms	45.0	23.3
Activity	48.0	24.7
Impact	31.6	20.7
Total	39.0	20.2

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> No details of SGRQ; states that it is standard practice for this to be completed by IPF patients at diagnosis.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications
Give a brief summary of the authors' conclusions from their analysis.

> No significant relationship between HRQoL evaluated by the SGRQ and subsequent mortality in IPF was demonstrated.

What are the implications of the study for the model?

> Data could be mapped if no EQ-5D data available.

Reference

Nishiyama 2005.[119]

Study characteristics

Research question

What are the stated objectives of the study?

To identify factors significantly contributing to HRQoL in IPF patients using the SGRQ, and to examine hypotheses that dyspnoea and exercise capacity affect HRQoL.

Describe the type of study and study design.

Cross-sectional/single-cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosed by surgical lung biopsy or HRCT in accordance with ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	64
Sex, M/F	35/6
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF
Other characteristics (sample size)	$n = 41$
	FVC per cent predicted not reported, only VC (76.6%, SD 16.8)
QoL instrument	SGRQ (Japanese version)
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Japan; hospital outpatients.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies or expert opinion?

QoL data obtained in this single observational study.

Results
Summarise the results.

The BDI score showed the most significant correlation with each SGRQ score

SGRQ score	Mean	SD
Symptoms	40.1	24.6
Activity	44.5	26.7
Impacts	28.9	19.8
Total	35.7	20.6

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> Dyspnoea was most important factor determining HRQoL in IPF. The types of other variables that correlated with HRQoL were different from those in COPD.

What are the implications of the study for the model?

> SGRQ would be reasonably valid instrument to use for utility scores based on correlations with physiological variables and may be useful if no utility data are identified.

Reference

Nishiyama 2008.[73]

Study characteristics

Research question
What are the stated objectives of the study?

To evaluate effects of pulmonary rehabilitation in patients with IPF, compared with control.

Describe the type of study and study design.

RCT (included in clinical effectiveness systematic review).

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Consecutive patients referred to outpatient clinic with IPF, diagnosed using ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	68.1 for treated group, 66.3 for total group
Sex, M/F	21/7
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF
Other characteristics (sample size)	$n = 28$
	FVC per cent predicted reported (67.4%)
QoL instrument	SGRQ
Utility values, Y/N	N
Treatment effect, if reported	FVC

Country/setting
What is the country and setting for the evaluation?

> Japan; outpatient clinic.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

> Single study (RCT).

Results
Summarise the results.

Significant improvements were seen in 6MWD and HRQoL as a result of rehabilitation programme

SGRQ scores baseline	Control, mean (SD)	Treated, mean (SD)
Symptoms	38.0 (25.8)	56.4 (22.3)
Activity	50.4 (26.2)	64.7 (17.1)
Impacts	29.9 (23.7)	39.7 (17.6)
Total	37.8 (22.7)	50.2 (16.3)

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis

Pulmonary rehab improves both exercise capacity and HRQoL in patients with IPF.

What are the implications of the study for the model?

May be useful if no valid utility data identified. Limited by small numbers. Participants mostly moderate severity IPF.

Reference

Ozalevli 2009.[110]

Study characteristics

Research question
What are the stated objectives of the study?

To investigate the effects of a home-based pulmonary rehab programme on functional outcome parameters in patients with IPF.

Describe the type of study and study design.

Prospective cohort pre and post study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosed by ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	62.8
Sex, M/F	10/5
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 15$
	FVC per cent predicted reported (71.6%, SD 8.2)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Pulmonary function tests

Country/setting
What is the country and setting for the evaluation?

Turkey; home.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

Single study.

Results
Summarise the results.

Baseline results:

SF-36 scores baseline	Mean	SD
Physical function	56.0	5.7
Role physical	25.0	1.7
Bodily pain	67.3	2.6
General health	57.0	4.6
Vitality	52.0	4.9
Social function	75.8	2.7
Role emotional	29.0	1.3
Mental health	49.9	6.7

A significant increase in general HRQoL was seen after the programme

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

Home-based pulmonary rehab may improve HRQoL in patients with IPF.

What are the implications of the study for the model?

Data may be useful if no utility data are identified as may be mapped onto EQ-5D. Limited by small sample size.

Reference

Peng 2008.[118]

Study characteristics

Research question
What are the stated objectives of the study?

To confirm the cross-sectional and longitudinal construct validity of the SGRQ for the measurement of HRQoL in patients with IPF.

Describe the type of study and study design.

Observational, cross-sectional and longitudinal.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosed by surgical lung biopsy or HRCT.

What are the characteristics of the baseline cohort for the evaluation?

Age	64
Sex, M/F	54/14
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF
Other characteristics (sample size)	$n = 68$
	FVC per cent predicted reported (66%)
QoL instrument	SGRQ (Chinese version)
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting

What is the country and setting for the evaluation?

China; outpatient clinic.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

From this single observational study.

Results

Summarise the results.

SGRQ score at baseline	Mean	SD
Symptoms	65	16
Activity	56	15
Impact	49	19
Total	54	15

HRQoL substantially impaired in IPF patients, especially in symptom and activity domains. SGRQ was correlated with lung function and exercise function. Over time, $n = 45$, SGRQ scores reduced (not significantly) in two domains (symptoms/impact) and improved on one (activity) and total score significantly

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> SGRQ has good longitudinal and construct validity in patients with IPF; however, further assessment of reliability and responsiveness is needed.

What are the implications of the study for the model?

> SGRQ would be a reasonable instrument to map to utility if no data available. Limitation using longer-term data owing to dropout rates. Population mostly moderate in severity.

Reference

Raghu 2010.[125]

Study characteristics

Research question
What are the stated objectives of the study?

To examine longitudinal changes in HRQoL and dyspnoea in IPF patients.

Describe the type of study and study design.

Randomised, placebo-controlled trial of bosentan.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Well-defined diagnosis of IPF based on ATS/ERS consensus criteria (BUILD-1) (excluded from the clinical effectiveness systematic review).

What are the characteristics of the baseline cohort for the evaluation?

Age	65.2
Sex, M/F	112/42
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	158 randomised (154 analysed)
	FVC per cent predicted not reported in this publication, was 67.7% in another publication[128]
QoL instrument	SGRQ and SF-36
Utility values, Y/N	N
Treatment effect, if reported	Mitigation of impairment of HRQoL

Country/setting
What is the country and setting for the evaluation?

29 centres in Europe (Germany, France, Italy, Switzerland and the UK), the USA, Canada and Israel.[128] Outpatient setting.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL from this study.

Results
Summarise the results.

No differences in HRQoL between treatment groups at baseline or at 12 months

	Bosentan	Placebo
Baseline SGRQ, median (95% CI)		
Symptoms	51.2 (44.7 to 56.0)	54.8 (47.9 to 59.1)
Activity	59.5 (53.6 to 66.1)	59.5 (56.0 to 65.6)
Impacts	32.9 (25.0 to 40.3)	30.5 (25.7 to 38.4)
Total	43.9 (36.6 to 52.1)	44.3 (40.2 to 50.0)
Baseline SF-36, median (95% CI)		
Physical function	55.0 (40.0 to 65.0)	47.5 (35.0 to 55.0)
Role physical	25.0 (25.0 to 50.0)	25.0 (25.0 to 50.0)
Bodily pain	74.0 (64.0 to 100.0)	64.0 (61.0 to 74.0)
General health	47.0 (37.0 to 57.0)	46.0 (40.0 to 52.0)
Vitality	50.0 (40.0 to 55.0)	50.0 (40.0 to 55.0)
Social function	75.0 (75.0 to 87.5)	81.3 (62.5 to 87.5)
Role emotional	100.0 (66.7 to 100.0)	100.0 (66.7 to 100.0)
Mental health	80.0 (68.0 to 84.0)	76.0 (68.0 to 80.0)

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> HRQoL changed minimally as a result of treatment with bosentan, although exploratory analysis suggested a QoL benefit among those diagnosed by surgical lung biopsy.

What are the implications of the study for the model

> Bosentan treatment does not appear to give rise to changes in QoL for IPF patients. Baseline data may be of use to the model if no utility data are identified, although median not mean scores reported.

Reference

Swigris 2010.[126]

Study characteristics

Research question
What are the stated objectives of the study?

To examine the validity of the SF-36 and SGRQ in IPF and to determine scores from each that would constitute a minimum important difference.

Describe the type of study and study design.

Retrospective analysis of BUILD-l trial data (BUILD-1 was a randomised placebo-controlled trial of bosentan).

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) With IPF defined according to consensus guidelines (ATS/ERS).

What are the characteristics of the baseline cohort for the evaluation?

Age	65.12
Sex, M/F	73/27
Race (if appropriate)	92% Caucasian
Indication/disease	IPF
Other characteristics (sample size)	$n = 100$. 158 in original sample and for baseline measurement
	FVC per cent predicted reported (66.97%)
QoL instrument	SF-36 and SGRQ
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting

What is the country and setting for the evaluation?

> BUILD-1 trial was performed in 29 centres in Europe (Germany, France, Italy, Switzerland and the UK), the USA, Canada and Israel.

Data sources

Effectiveness

Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

> SF-36 and SGRQ data obtained from BUILD-1 trial for two time points – baseline and 6 months.

Results

Summarise the results.

SF-36 and SGRQ are valid instruments in longitudinal studies of IPF

Baseline	Mean	SD
SRGQ		
Symptoms	50.1	21.9
Activity	60.6	22.8
Impact	33.7	20.6
Total	44.8	19.5
SF-36		
Physical function	35.4	10.3
Role physical	37.8	11.6
Bodily pain	47.6	10.7
General health	37.8	9.4
Vitality	43.1	9.2
Social function	44.6	12.3
Role emotional	42.7	14.2
Mental health	48.2	10.1
Physical summary	37	10
Mental summary	44.2	10.84

NB: 95% CI data available but not extracted

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> For SF-36 used scoring algorithms to generate linear *t*-score transformations.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

> SF-36 and SGRQ possess reasonable validity for differentiating subjects whose disease severity changes over time.

What are the implications of the study for the model?

> Either SF-36 or SGRQ would be OK to use for utilities if no data available. Patients were mostly moderate severity.

Reference

Tomioka 2007.[115]

Study characteristics

Research question

What are the stated objectives of the study?

To validate the cross-sectional and longitudinal use of SF-36 for measuring HRQoL in patients with IPF.

Describe the type of study and study design.

Cross-sectional and longitudinal study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Diagnosed on basis of lung biopsy or HRCT as defined by ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	69.9
Sex, M/F	32/14
Race (if appropriate)	Asian (assumed)
Indication/disease	IPF
Other characteristics (sample size)	$n = 46$
	FVC per cent predicted not reported, VC only (71.0%, SD 17.5)
QoL instrument	SF-36
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

> Japan/outpatient clinic.

Data sources

Effectiveness
Were the QoL data derived from: a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

> QoL from this study (baseline and > 1 year later).

Results
Summarise the results.

> Patients with IPF had significantly lower scores across all eight domains of the SF-36 than the general population. Presented as deviation values against national reference values; no actual data presented in the publication. Also showed decline over time from 32 participants but this was no statistically significant in six of the eight domains. Data in a figure only.

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

IPF patients have significantly impaired HRQoL and this is reflected in their SF-36 scores. SF-36 scores also measure dimensions not fully estimated by clinical assessment.

What are the implications of the study for the model?

SF-36 scores would be a reasonable choice for model utilities in the absence of data. However, data are not presented.

Reference

Tzanakis 2005.[120]

Study characteristics

Research question
What are the stated objectives of the study?

To test whether existing instruments for obstructive airway disease could be applied to patients with IPF.

Describe the type of study and study design.

Cross-sectional/matched control study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Plus healthy control group. Diagnosis clinically and histologically consistent with IPF by ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	66
Sex, M/F	21/4
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 25$
	FVC per cent predicted reported (68.8%, SD 16)
QoL instrument	SGRQ
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Greece; hospital (outpatients).

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

Single study.

Results
Summarise the results.

Lung volumes (FVC and TLC) correlated significantly with the SGRQ. IPF patient scores were significantly different from scores in the control group

SGRQ domain	Mean	SD
Symptoms	55.9	25.3
Activity	36.2	21.4
Impact	29.6	21
Total	37.7	18.9

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping
If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

SGRQ is sensitive to HRQoL in IPF patients.

What are the implications of the study for the model?

May be useful if no utility data are identified, limited by the small sample size. Participants were approximately of moderate IPF severity.

Conclusions/implications

Reference

Verma 2011.[123]

Study characteristics

Research question

What are the stated objectives of the study?

To determine whether or not an association between SF-36 and SGRQ scores and other markers of disease severity exists, and to identify which physiological and functional variables are independently associated with HRQoL.

Describe the type of study and study design.

Cross-sectional study/single-cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) With diagnosis based on ATS/ERS criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	59.4
Sex, M/F	90/47
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 137$
	FVC per cent predicted reported (61.7%, SD 19.8)
QoL instrument	SF-36 and SGRQ
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Canada. Web-based versions of SGRQ and SF-36 were administered. Clinical assessment at hospital clinic.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL data from this study

	Mean	95% CI
Score SF-36		
Physical function	25.6	21.9 to 29.4
Role physical	31.0	26.9 to 35.1
Bodily pain	68.0	63.4 to 72.7
General health	35.5	31.9 to 39.0
Vitality	39.3	35.8 to 42.9
Social function	59.2	54.05 to 64.38
Role emotional	74.5	69.6 to 79.3
Mental health	71.2	67.9 to 74.6
Physical summary	29.4	27.9 to 30.9
Mental summary	49.7	47.6 to 51.7
Score SGRQ		
Symptoms	59.8	56.2 to 63.4
Activity	81.6	78.7 to 84.4
Impacts	54.1	50.6 to 57.6
Total	63.4	60.4 to 66.3

Results
Summarise the results.

6MWD was the only functional measure of disease severity significantly associated with all domain scores of the SGRQ and SF-36.

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

> Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

> SF-36 component summary scores were normalised.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis

> 6MWD is an important predictor of HQoL in patients with IPF.

What are the implications of the study for the model?

> 6MWD might be used as proxy for HRQoL. Scores of SGRQ or SF-36 may be useful if no utility data are identified and if they can be mapped to the EQ-5D. Population mostly moderate in severity.

Reference

Zimmermann 2007.[124]

Study characteristics

Research question
What are the stated objectives of the study?

To determine if there is a HRQoL instrument that best represents functional capacity in IPF patients.

Describe the type of study and study design.

Cross-sectional/single-cohort study.

Was the sample from (i) the general population, (ii) patients with the disease of interest, (iii) individuals with knowledge of the disease or (iv) other?

(ii) Identified through clinical records, diagnosis based on ATS/ERS consensus criteria.

What are the characteristics of the baseline cohort for the evaluation?

Age	61.4
Sex, M/F	12/8
Race (if appropriate)	Not reported
Indication/disease	IPF
Other characteristics (sample size)	$n = 20$
	FVC per cent predicted reported (70.4%, SD 19.4)
QoL instrument	SF-36 and SGRQ
Utility values, Y/N	N
Treatment effect, if reported	Not applicable

Country/setting
What is the country and setting for the evaluation?

Brazil; outpatient clinic.

Data sources

Effectiveness
Were the QoL data derived from a single (observational) study, a review/synthesis or combination of previous studies, or expert opinion?

QoL obtained in this study.

Results
Summarise the results.

Almost all of the SGRQ domains were strongly correlated with functional status, but only physical function and vitality from SF-36 had a strong correlation with functional status

	Mean	SD
SGRQ scores		
Symptoms	46.4	20.3
Activity	62.4	19
Impact	43.6	20.9
Total	48.4	17.9
SF-36		
Physical function	46	18.3
Role physical	57.5	39.8
Bodily pain	60.6	31.9
General health	53.7	24.1
Vitality	49.2	24.3
Social function	56.9	32.2
Role emotional	46.6	39.5
Mental health	66.8	17

Were the methods for deriving these data adequately described (give sources if using data from other published studies)?

Yes.

Mapping

If a model was used, describe the type of model (e.g. regression) or other conversion algorithm.

Not applicable.

Conclusions/implications

Give a brief summary of the authors' conclusions from their analysis.

A specific rather than a generic instrument is more appropriate for HRQoL evaluation in IPF.

What are the implications of the study for the model?

SGRQ is to be preferred to SF-36 for providing utilities (if correlation to functional status does indeed suggest that it is a more 'valid' instrument). May be useful if no utility data identified, limited by small sample size.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library