National Clinical Guideline Centre

Diagnosis and management of suspected idiopathic pulmonary fibrosis

Idiopathic pulmonary fibrosis

National Clinical Guideline Centre

Methods, evidence and recommendations

June 2013

Commissioned by the National Institute for Health and Care Excellence

1

Disclaimer
Healthcare professionals are expected to take NICE clinical guidelines fully into account when exercising their clinical judgement. However, the guidance does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of each patient, in consultation with the patient and/or their guardian or carer.

Funding
National Institute for Health and Care Excellence

Update information
May 2017: Recommendation 27 was amended to add a link to the NICE technology appraisal on nintedanib for the treatment of idiopathic pulmonary fibrosis. Two outdated research recommendations have been stood down and removed from the short version. In this version they have been greyed out in the list of Key Research Recommendations.

Contents

1

1 Guideline development group members

Name	Role
Guideline development group (GDG)	
Nik Hirani (Chair)	Senior Clinical Lecturer and Honorary Consultant in Respiratory Medicine
Geraldine Burge	Interstitial Lung Disease Specialist Nurse (with expertise in palliative care)
Sue Copley	Consultant Radiologist and Reader in Thoracic Imaging
Annette Duck	Interstitial Lung Disease Specialist Nurse (with expertise in palliative care)
Nicholas Kim Harrison	Senior Clinical Lecturer and Honorary Consultant Respiratory Physician
Melissa Hippard	Patient member
Richard Hubbard	British Lung Foundation Professor of Respiratory Epidemiology
Angela Key	Chief Respiratory Physiologist
Tessa Lewis	General Practitioner
Anne Millar	Professor of Respiratory Medicine
Nick Screaton	Consultant Cardiothoracic Radiologist
Malcolm Weallans	Patient member
Patrick Wilson	Senior Respiratory Pharmacist
Co-opted External Advisors	
Stephen Clarke	Consultant Cardiothoracic and Cardiopulmonary Transplant Surgeon
Andrew Nicholson	Consultant Histopathologist
Sally Singh	Head of Pulmonary and Cardiac Rehabilitation
National Clinical Guideline Centre (NCGC) technical team	
Vanessa Nunes	Guideline lead
Nina Balachander	Senior research fellow and project manager
Vicki Pollit	Health economist
Izaba Younis	Research fellow
Zahra Naqvi	Research fellow

Acknowledgments

The development of this guideline was greatly assisted by the following people:

NCGC: Hati Zorba, Tamara Diaz, Emmert Roberts, Maggie Westby and Bernard Higgins.

National Institute for Health Research, Health Technology Appraisal Programme (NIHR HTA) Peer Reviewers: Emma Loveman and Jeremy Jones

2 Introduction

Why the guideline is needed

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive fibrotic interstitial lung disease (ILD) of unknown origin. It is a difficult disease to diagnose and often requires the collaborative expertise of a chest physician, radiologist and histopathologist to reach a consensus diagnosis. Most people with idiopathic pulmonary fibrosis experience symptoms of breathlessness, which may initially be only on exertion. Cough, with or without sputum is a common symptom. Over time, these symptoms are associated with a decline in lung function, reduced quality of life and ultimately death. Specific pharmacological therapies for IPF are limited but the last decade has seen more trials of new drugs which have had a variable impact on clinical practice. A number of difficulties arise when undertaking clinical trials in IPF in terms of defining precise, diagnostic inclusion criteria and clinically meaningful end-points. However, such trials are the only way by which promising new treatments will come to benefit patients. Furthermore, it is only by performing rigorous clinical trials, we have learned that drugs once widely used to treat IPF may in fact have been harmful. The limitations of current pharmacological therapies for IPF highlight the importance of other forms of treatment including lung transplantation and best supportive care such as oxygen therapy, pulmonary rehabilitation and palliation of symptoms. These are interventions which justifiably require scrutiny in the context of healthcare delivery by the modern NHS. Despite the significant burden of disease caused by IPF, there is currently no established framework within the NHS for its diagnosis and management thus creating an environment in which significant variations in clinical care may occur. In recognition of this, the Department of Health commissioned the National Institute of Health and Care Excellence (NICE) to produce a guideline aimed at improving the care of people with IPF.

Terminology and definitions

Idiopathic pulmonary fibrosis is the commonest of many interstitial lung diseases and it must be distinguished from the ILDs which have known causes or associations such as asbestosis, lung disease associated with connective-tissue disease, hypersensitivity pneumonitis and drug-induced lung disease. Historically the term 'cryptogenic fibrosing alveolitis (CFA)' has been considered synonymous with IPF. It is now recognised that CFA is a syndrome that encompasses a group of distinct interstitial lung diseases of unknown cause that often present with clinical features which resemble IPF. These include several idiopathic interstitial pneumonias (IIP), including fibrotic non-specific interstitial pneumonia (NSIP). The term 'CFA' should not be used as a diagnostic label in individuals. Instead, the diagnostic pathway in suspected IPF should aim to lead to a more precise diagnosis such as 'IPF', 'idiopathic NSIP' etc. This attempt at precision is important because of the implications a diagnosis of 'IPF' has on all aspects of management. If there is diagnostic uncertainty, this should be conveyed to the patient and their carers[59,78]. Recognising that terminology has evolved and 'case-definition' of IPF has become more refined over time, the evidence on which this guideline is founded derives almost entirely from studies performed after 1998 when the histopathological features of IPF were re-defined [60]. However, in some pre-defined circumstances, specifically where the evidence-base is sparse, older studies have also been included.

Epidemiology of IPF

The incidence of IPF is approximately 8 to 9 per 100,000 person years, which means more than 5000 new cases occur in the UK each year. It is rare in people younger than 45 and the median age of presentation is 70 years. The prevalence is around 15 to 25 per 100,000 and increases with age. The average hospital with a catchment of 500,000 will have 40 to 45 new cases a year and the average GP

surgery of 10,000 patients will have 2 to 3 new cases every three years [85]. Around two-thirds of people with IPF are smokers and IPF often co-exists with chronic obstructive pulmonary disease (COPD).

The natural history of IPF

The median survival for people with IPF in the UK is approximately 3 years from the time of diagnosis, but it is recognised that there is a very wide spectrum associated with survival. However, approximately 20% of patients survive for greater than 5 years. This observation emphasises how the rate of disease progression varies between individuals and an individual patient's prognosis is difficult to estimate at the time of diagnosis and may only become apparent after a period of careful follow up. Such uncertainty is disconcerting for patients and their carers and further emphasises the importance of establishing a confident diagnosis before imparting its implications.

What is in this guideline?

The guideline offers recommendations on the diagnosis and delivery of care to people with IPF, from initial suspicion of the disease, usually in primary care, through referral to a chest specialist, the role of multidisciplinary diagnostic and management teams and specific therapeutic interventions. The guideline addresses the timing, frequency and nature of tests that inform diagnosis and prognosis. It addresses the value of drugs aimed primary at modifying disease progression, and interventions which largely provide symptom relief including oxygen therapy, pulmonary rehabilitation and palliation of breathlessness and cough. The timing of referral for lung transplantation and the value of mechanical respiratory support in IPF is also included. Recommendations have been made on the clinical benefits and cost-effectiveness of interventions and these are founded on a rigorously reviewed evidence-base wherever possible. However, there are some questions of importance to patients and healthcare professionals for which, as yet, there is a paucity of evidence. In these areas, recommendations are based on expert consensus opinion.

What is not in this guideline?

The guideline does not include recommendations on interstitial lung disease other than IPF. It does not include guidance on secondary pulmonary hypertension or lung cancer, which are both recognised common complications of IPF. Lung transplantation, other than the timing of referral, is not included. Wherever relevant, cross-reference is made to other NICE guidelines and every effort has been made to achieve consistency across such guidelines that include recommendations of relevance to people with IPF.

3 Development of the guideline

3.1 What is a NICE clinical guideline?

NICE clinical guidelines are recommendations for the care of individuals in specific clinical conditions or circumstances within the NHS – from prevention and self-care through primary and secondary care to more specialised services. We base our clinical guidelines on the best available research evidence, with the aim of improving the quality of healthcare. We use predetermined and systematic methods to identify and evaluate the evidence relating to specific review questions.

NICE clinical guidelines can:

- provide recommendations for the treatment and care of people by health professionals
- be used to develop standards to assess the clinical practice of individual health professionals
- be used in the education and training of health professionals
- help patients to make informed decisions
- improve communication between patient and health professional

While guidelines assist the practice of healthcare professionals, they do not replace their knowledge and skills.

We produce our guidelines using the following steps:

- Guideline topic is referred to NICE from the Department of Health
- Stakeholders register an interest in the guideline and are consulted throughout the development process.
- The scope is prepared by the National Clinical Guideline Centre (NCGC)
- The NCGC establishes a guideline development group
- A draft guideline is produced after the group assesses the available evidence and makes recommendations
- There is a consultation on the draft guideline.
- The final guideline is produced.

The NCGC and NICE produce a number of versions of this guideline:

- the full guideline contains all the recommendations, plus details of the methods used and the underpinning evidence
- the NICE guideline lists the recommendations
- the NICE pathway is a practical online resource for healthcare and other professionals that contains all the recommendations from a guideline, as well as any other NICE guidance that is directly relevant to the topic. It also contains links to implementation tools and to related NICE guidance and pathways.
- information for the public (IFP) is written using suitable language for people without specialist medical knowledge.

This version is the full version. The other versions can be downloaded from NICE at www.nice.org.uk

3.2 Remit

NICE received the remit for this guideline from the Department of Health. They commissioned the NCGC to produce the guideline.

The remit for this guideline is:
The Department of Health has asked NICE 'To produce a clinical guideline on the diagnosis and management of suspected idiopathic pulmonary fibrosis.'

3.3 Who developed this guideline?

A multidisciplinary Guideline Development Group (GDG) comprising professional group members and consumer representatives of the main stakeholders developed this guideline (see section on Guideline Development Group Membership and acknowledgements).

The National Institute for Health and Care Excellence funds the National Clinical Guideline Centre (NCGC) and thus supported the development of this guideline. The GDG was convened by the NCGC and chaired by Nik Hirani in accordance with guidance from the National Institute for Health and Care Excellence (NICE).

The group met every 4-5 weeks during the development of the guideline. At the start of the guideline development process all GDG members declared interests including consultancies, fee-paid work, share-holdings, fellowships and support from the healthcare industry. At all subsequent GDG meetings, members declared arising conflicts of interest, which were also recorded (Appendix B).

Members were either required to withdraw completely or for part of the discussion if their declared interest made it appropriate. The details of declared interests and the actions taken are shown in Appendix B.

Staff from the NCGC provided methodological support and guidance for the development process. The team working on the guideline included a project manager, systematic reviewers, health economists and information scientists. They undertook systematic searches of the literature, appraised the evidence, conducted meta-analysis and cost effectiveness analysis where appropriate and drafted the guideline in collaboration with the GDG.

3.4 What this guideline covers

This guideline covers the following populations:

Adults (18 and older) with suspected or diagnosed idiopathic pulmonary fibrosis in all settings in which NHS healthcare is provided.

The following clinical issues are covered:

- Diagnosis:
    - high resolution computed tomography (HRCT) scanning
    - biopsy (bronchoalveolar lavage and surgical lung biopsy)
    - multidisciplinary teams to achieve a consensus diagnosis
    - pulmonary function tests.
- Prognosis:
    - pulmonary function tests (resting spirometric and gas transfer measurement)
    - sub-maximal exercise testing
    - echocardiography.
- Treatment of the disease with the following drugs:
    - prednisolone
    - mycophenolate mofetil
    - warfarin
    - azathioprine

    - N-acetyl cysteine
    - proton-pump inhibitors
    - co-trimoxazole
    - ambrisentan
    - bosentan
    - sildenafil
- Symptom relief:
    - lung transplantation timing and referral
    - best supportive care (benzodiazepines, oxygen therapy and palliative care)
    - non-invasive and invasive ventilation
    - pulmonary rehabilitation (breathlessness management).
- Patient review and follow-up.

For further details please refer to the scope in Appendix A [and review questions in section 3.1].

3.5 What this guideline does not cover

This guideline does not cover:

- Children and young people (younger than 18).
- People with a diagnosis of pulmonary fibrosis as a complication of:
    - connective tissue disorders (for example, systemic lupus erythematosus, rheumatoid arthritis, scleroderma, polymyositis and dermatomyositis)
    - a known exogenous agent (for example, drug-induced disease or asbestosis).
- Therapies for pulmonary hypertension as a complication of idiopathic pulmonary fibrosis.
- Treatment of lung cancer as a complication of idiopathic pulmonary fibrosis.
- Lung transplantation, other than timing and referral.

3.6 Relationships between the guideline and other NICE guidance

Related NICE Clinical Guidelines:

Opioids in palliative care. NICE clinical guideline 140 (2012).

Patient experience in adult NHS service. NICE clinical guideline 138 (2012).

Lung cancer. NICE clinical guideline 121 (2011).

Tuberculosis. NICE clinical guideline 117 (2011).

Chronic obstructive pulmonary disease. NICE clinical guideline 101 (2010).

Dyspepsia. NICE clinical guideline 17 (2004).

Related NICE Public Health Guidance:

Smoking cessation services. NICE public health guidance 10 (2008).

Brief interventions and referral for smoking cessation. NICE public health guidance 1 (2006).

Related NICE Technology Appraisal:

Smoking cessation – varenicline. NICE technology appraisal 123 (2007).

Pirfenidone for the treatment of idiopathic pulmonary fibrosis. NICE technology appraisal 282. (2013).

NICE Related Guidance currently in development:

Dyspepsia/GORD. NICE clinical guideline. Publication date to be confirmed.

Tuberculosis (update). NICE clinical guideline. Publication date to be confirmed.

4 Methodology

This guidance was developed in accordance with the methods outlined in the NICE Guidelines Manual 2012[84].

4.1 Developing the review questions and outcomes

Review questions were developed in a PICO framework (patient, intervention, comparison and outcome) for intervention reviews, using population, presence or absence of factors under investigation (for example prognostic factors) and outcomes for prognostic reviews. This use of a framework guided the literature searching process, critical appraisal and synthesis of evidence, and facilitated the development of recommendations by the Guideline Development Group (GDG). The review questions were drafted by the NCGC technical team and refined and validated by the GDG. The GDG chose approximately 7 outcomes identifying which outcomes were critical to their decision making and which were important. This distinction helped the GDG to make judgements about the importance of the different outcomes and their impact on decision making. For example, mortality will usually be considered a critical outcome and would be given greater weight when considering the clinical effectiveness of an intervention than an important outcome with less serious consequences. The GDG decide on the relative importance in the review protocol before seeing the review.

For questions on prognostic factors, protocols stated the risk factor that would be searched for instead of the intervention and comparison.

The GDG agreed that for the clinical questions identified, the following critical outcomes were considered the most important for decision making when concluding on the efficacy of clinical interventions:

- All cause and IPF related mortality
- Survival
- Change in percentage predicted forced vital capacity (FVC)

Whilst significant change in all-cause mortality should be the 'gold standard' clinically meaningful end-point in phase 3 trials, it may be impractical given the large number of patients who would need to be enrolled and length of time required for follow-up. For this reason, serial trend in FVC is considered by many as an acceptable and practical marker of improvement or decline. Such trends may also be the most effective way to confirm disease stability or measure small, incremental improvements both of which should be considered beneficial effects in a condition with such a high mortality.

Table 1: Review questions

Chapter	Review questions	Outcomes
Diagnosis	In suspected IPF what is the value of adding biopsy to clinical evaluation, PFTs, CT +/- bronchoalveolar lavage for confirming the diagnosis of IPF?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Sensitivity Specificity Other outcomes Adverse events

Chapter	Review questions	Outcomes
		Improvement in health-related quality of life
	In suspected IPF what is the value of adding multidisciplinary team (MDT) consensus to clinical assessment, PFTs and CT in the diagnosis of IPF?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Sensitivity Specificity Inter-observer agreement Improvement in health-related quality of life
	How and by whom a MDT diagnostic consensus is best achieved (i.e. constituency of the MDT, specialist clinics, and networks)?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Sensitivity Specificity Inter-observer agreement Improvement in health-related quality of life
Prognosis	Do serial pulmonary function tests (resting spirometric, gas transfer measurement and oxygen saturation) predict prognosis of IPF?	Critical outcomes Mortality or survival (time to event) Other outcomes Progression free survival Acute exacerbation (time to event) Respiratory hospitalisations (surrogate outcome for acute exacerbation) Eligibility for lung transplantation
	Does baseline sub-maximal exercise testing predict prognosis of IPF?	Critical outcomes Mortality or survival (time to event) Other outcomes Progression free survival Acute exacerbation (time to event) Respiratory hospitalisations (surrogate outcome for acute exacerbation) Eligibility for lung transplant
	Does baseline echocardiography predict prognosis of IPF?	Critical outcomes Mortality or survival (time to event) Other outcomes Progression free survival Acute exacerbation (time to event) Respiratory hospitalisations (surrogate outcome for acute exacerbation) Eligibility for lung transplant
	Do baseline CT scores predict prognosis of IPF?	Critical outcomes Mortality or survival (time to event) Other outcomes Progression free survival Acute exacerbation (time to event) Respiratory hospitalisations (surrogate outcome for acute exacerbation)

Chapter	Review questions	Outcomes
		Eligibility for lung transplant
Best supportive care	What is the clinical and cost effectiveness of best supportive care (palliation of cough, breathlessness and fatigue, and oxygen management) in the symptomatic relief of people with IPF?	Critical outcomes Improvement in health-related quality of life Other outcomes All cause and IPF related mortality Hospitalisations due to IPF complications (including IPF exacerbations) Improvement in cough and breathlessness Improvement in psychosocial health (including depression) Performance on sub-maximal walk test (distance walked and lowest oxygen saturation (SaO_2)) Symptom relief
Psychosocial support	What is the specific type of psychosocial support and information that should be provided for patients diagnosed with IPF?	Critical outcomes Improvement in health-related quality of life Other outcomes Dyspnoea Improvement in psychosocial health (including depression)
Pulmonary rehabilitation	What are the benefits of pulmonary rehabilitation programmes for patients with confirmed IPF?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Dyspnoea Hospitalisations due to IPF complications (including IPF exacerbations) Improvement in cough and breathlessness Improvement in health-related quality of life Performance on sub-maximal walk test (distance walked and lowest SaO_2) Improvement in psychosocial health (including depression)
	What is the optimal course content, setting and duration for patients referred for pulmonary rehab programmes?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Dyspnoea Hospitalisations due to IPF complications (including IPF exacerbations) Improvement in cough and breathlessness Improvement in health-related quality of life Performance on sub-maximal walk test (distance walked and lowest SaO_2) Improvement in psychosocial health (including depression)
Pharmacological interventions	Which drug should be initiated first, for how long, and what combination in the treatment of IPF?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes

Chapter	Review questions	Outcomes
	What is the clinical and cost effectiveness of pharmacological interventions to manage patients with suspected or confirmed IPF?	Adverse events (please see adverse events table listed in Appendix N) Dyspnoea Change in percentage predicted carbon monoxide diffusing capacity (DLCO) Hospitalisations due to IPF complications, including IPF exacerbations Improvement in health-related quality of life Change in percentage predicted FVC Performance on sub-maximal walk test (distance walked and lowest SaO_2)
	Which measures can be taken to minimize the occurrence/severity of adverse events when undergoing pharmacological treatment for IPF?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Adverse events (please see adverse events table listed in Appendix N) Dyspnoea Hospitalisations due to IPF complications, including IPF exacerbations Improvement in health-related quality of life Performance on sub-maximal walk test (distance walked and lowest SaO_2)
Lung transplantation	What is the optimal timing to consider a patient with IPF for lung transplantation referral?	Critical outcomes All cause and IPF related mortality 1 and 3 year survival rates Other outcomes Cross-over time Hospitalisations due to IPF complications (including IPF exacerbations) Improvement of health-related quality of life Occurrence lung transplantation
Ventilation	In acute or acute-on chronic respiratory failure in people with IPF, what is the value of non-invasive and invasive ventilation?	Critical outcomes Mortality (in hospital and post discharge) Other outcomes Improvement of health-related quality of life Hospital length of stay
Patient review and follow up	How often should a patient with confirmed diagnosis of IPF be reviewed?	Critical outcomes Change in percent predicted DLCO Change in percent predicted FVC Other outcomes Oxygen saturation at rest Oxygen saturation on exertion Distance walked on 6 min walk or incremental shuttle walk test Eligibility for lung transplant
	In which healthcare setting and by whom should a review appointment for patients with	Critical outcomes Change in percent predicted DLCO Change in percent predicted FVC

Chapter	Review questions	Outcomes
	confirmed IPF be conducted?	Other outcomes Oxygen saturation at rest Oxygen saturation on exertion Distance walked on 6 min walk or incremental shuttle walk test Eligibility for lung transplant

4.2 Searching for evidence

4.2.1 Clinical literature search

Systematic literature searches were undertaken to identify evidence within published literature in order to answer the review questions as per The Guidelines Manual 2012[83]. Clinical databases were searched using relevant medical subject headings, free-text terms and study type filters where appropriate. Studies published in languages other than English were not reviewed. Where possible, searches were restricted to articles published in English language. All searches were conducted on core databases, MEDLINE, Embase and The Cochrane Library. The additional subject specific databases CINAHL and PsychInfo were used for some questions. All searches were updated on the 1st November 2012. No papers published after this date were considered.

Search strategies were checked by looking at reference lists of relevant key papers, checking search strategies in other systematic reviews and asking the GDG for known studies. The questions, the study types applied, the databases searched and the years covered can be found in Appendix C and D.

During the scoping stage, a search was conducted for guidelines and reports on the websites listed below and on organisations relevant to the topic. Searching for grey literature or unpublished literature was not undertaken. All references sent by stakeholders were considered.

- Guidelines International Network database (www.g-i-n.net)
- National Guideline Clearing House (www.guideline.gov/)
- National Institute for Health and Care Excellence (NICE) (www.nice.org.uk)
- National Institutes of Health Consensus Development Program (consensus.nih.gov/)
- National Library for Health (www.library.nhs.uk/)

4.2.2 Health economic literature search

Systematic literature searches were also undertaken to identify health economic evidence within published literature relevant to the review questions. The evidence was identified by conducting a broad search relating to the guideline population in the NHS economic evaluation database (NHS EED), the Health Economic Evaluations Database (HEED) and health technology assessment (HTA) databases with no date restrictions. Additionally, the search was run on MEDLINE and Embase, with a specific economic filter, from 2010, to ensure recent publications that had not yet been indexed by these databases were identified. Studies published in languages other than English were not reviewed. Where possible, searches were restricted to articles published in English language.

The search strategies for health economics are included in Appendix D. All searches were updated on the 1st November 2012. No papers published after this date were considered.

4.3 Evidence of effectiveness

4.3.1 The literature review

The process for review of evidence of effectiveness is as follows:

The Research Fellows:

- Identified potentially relevant studies for each review question from the relevant search results by reviewing titles and abstracts – full papers were then obtained.
- Reviewed full papers against pre-specified inclusion / exclusion criteria to identify studies that addressed the review question in the appropriate population and reported on outcomes of interest (review protocols are included in Appendix C. To minimise errors and any potential bias in the assessment, two reviewers independently assessed a random selection of studies. Any differences arising from this were then discussed with the GDG.
- Critically appraised relevant studies using the appropriate checklist as specified in The Guidelines Manual[83].
- Extracted key information about the study's methods and results into evidence tables (evidence tables are included in Appendix F).
- Generated summaries of the evidence by outcome (included in the relevant chapter write-ups) and produced evidence statements indicating the number of included studies, sample size (number randomised), direction of effect, uncertainty and GRADE quality rating:
  - randomised studies: meta analysed, where appropriate and reported in GRADE profiles (for clinical studies) – see below for details
  - observational studies: data presented as a range of values in adapted GRADE profiles
  - diagnostic studies: data presented as a range of values in adapted GRADE profiles
  - prognostic studies: data presented as a range of values in adapted GRADE profiles
  - qualitative studies: each study summarised in a table where possible, otherwise presented in a narrative.

4.3.2 Inclusion/exclusion

The inclusion and exclusion criteria were considered according to the PICO used in the protocols, see Appendix C for full details. The GDG were consulted about any uncertainty regarding inclusion/exclusion of selected studies.

A major consideration in determining the inclusion and exclusion criteria in the protocol was the applicability of the evidence to the guideline population. The populations included in the review may differ for each review question, depending on the applicability of the data. See "Indirectness", section 3.3.8. The GDG acknowledged that data from ILD populations would include overlap between non-specific interstitial pneumonia (NSIP) and usual interstitial pneumonia (UIP), but agreed that the differences in clinical features and prognosis would not be a limitation for the evidence for diagnosis, pulmonary rehabilitation, best supportive care, psychosocial support and review and follow-up. However, a confirmed IPF population was specified for prognosis, pharmacological interventions, lung transplantation and ventilation. IPF data from ILD populations was only included in these clinical areas if IPF alone was analysed separately.

Pre-1994 evidence was excluded by limiting searches to post 1994 data for review questions relating to diagnosis and prognosis only, as advances in CT scanning have resulted in more consistent diagnosis of IPF after this time. No date restrictions were applied to any of the other clinical areas covered in this guideline.

Abstracts were included for three clinical areas; best supportive care, pulmonary rehabilitation and pharmacological interventions, on GDG advice due to the lack of evidence. Apart from those clinical areas abstracts were not included as evidence to inform other review questions, as the GDG considered that sufficient published evidence was available to inform decision making.

4.3.3 Methods of combining clinical studies

Data synthesis for intervention reviews

Available case analysis

Estimates of effect from individual studies were based on available case analysis (ACA) where it was possible to extract these data. ACA was defined as analysis using all participants with data available for the outcome being considered. For example, for dichotomous outcomes, the denominator is the number of participants with available data and the numerator is the number who experienced the event. Participants for whom data for that outcome were not available are assumed to be missing at random. Where ACA was not possible data were reported as in the study and this is explained in the introduction of the relevant clinical review.

This method was used rather than intention-to-treat analysis to avoid making assumptions about the participants for whom outcome data were not available, and rather assuming that those who drop out have the same event rate as those who continue. This also avoids incorrectly weighting studies in meta-analysis and overestimating the precision of the effect by using a denominator that does not reflect the true sample size with outcome data available.

ITT analysis is where all participants that were randomised are considered in the final analysis based on the intervention and control groups to which they were originally assigned. It was assumed that participants in the trials lost to follow-up did not experience the outcome of interest (categorical outcomes) and they would not considerably change the average scores of their assigned groups (for continuous outcomes). It is important to note that ITT analyses tend to bias the results towards no difference. ITT analysis is a conservative approach to analyse the data, and therefore the effect may be smaller than in reality.

Meta-analyses

Where possible, meta-analyses were conducted to combine the results of studies for each review question using Cochrane Review Manager (RevMan5) software. Fixed-effects (Mantel-Haenszel) techniques were used to calculate pooled risk ratios (relative risk) for the binary outcomes. The continuous outcomes were analysed using an inverse variance method for pooling weighted mean differences and where the studies had different scales, standardised mean differences were used.

Statistical heterogeneity was assessed by considering the chi-squared test for significance at $p < 0.1$ or an I-squared inconsistency statistic of >50% to indicate significant heterogeneity. Assessments of potential differences in effect between subgroups were based on the chi-squared tests for heterogeneity statistics between subgroups.

The means and standard deviations of continuous outcomes were required for meta-analysis. However, in cases where standard deviations were not reported, the standard error was calculated if the p-values or 95% confidence intervals were reported and meta-analysis was undertaken with the mean and standard error using the generic inverse variance method in Cochrane Review Manager (RevMan5) software. When the only evidence was based on studies which only presented means, this information was summarised in the GRADE tables without calculating the relative and absolute effect.

For binary outcomes, absolute event rates were also calculated using the GRADEpro software using event rate in the control arm of the pooled results.

Data synthesis for prognostic factor reviews

Odds ratio, relative risks or hazard ratios, with their 95% confidence intervals, from multivariate analyses were extracted from the papers, and standard errors were calculated from the 95% confidence intervals. The log of the effect size with its standard error was entered into the generic inverse variance technique in the Cochrane Review Manager (RevMan5) software (http://ims.cochrane.org/revman). Studies were not combined in a meta-analysis for observational studies.

The quality of studies was assessed and presented in an adapted GRADE profile according to criteria stated in the methodology checklist for prognostic studies in the guidelines manual. Results were reported as ranges.

Data synthesis for diagnostic test accuracy review

Evidence for diagnostic data was evaluated by study, using version two of the Quality Assessment of Diagnostic Accuracy Studies checklists (QUADAS-2) (http://www.bris.ac.uk/quadas/quadas-2). For diagnostic test accuracy studies, the following outcomes were reported: sensitivity, specificity, positive predictive value and negative predictive value. In cases where the outcomes were not reported, 2 by 2 tables were constructed from raw data to allow calculation of these accuracy measures. Summary receiver operative characteristic (ROC) curves, would have been generated if appropriate, however there were no data in the diagnostic reviews included in this guideline that could be combined to produce an ROC curve or diagnostic meta-analysis.

Data synthesis for qualitative review

Themes were identified from these studies by two reviewers independently, and then verified jointly. These themes were supplemented with data from surveys where available. Common themes relevant to the question are reported in a narrative in the guideline text.

4.3.4 Appraising the quality of evidence by outcomes

The evidence for outcomes from the included RCT and observational studies were evaluated and presented using the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group (http://www.gradeworkinggroup.org/). The software (GRADEpro) developed by the GRADE working group was used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results. The summary of findings was presented as two separate tables in this guideline. The 'Clinical/Economic Study Characteristics' table includes details of the quality assessment while the 'Clinical /Economic Summary of Findings' table includes pooled outcome data, where appropriate, an absolute measure of intervention effect and the summary of quality of evidence for that outcome. In this table, the columns for intervention and control indicate the sum of the sample size for continuous outcomes. For binary outcomes such as number of people with an adverse event, the event rates (n/N: number of people with events divided by sum of number of people) are shown with percentages. Reporting or publication bias was only taken into consideration in the quality assessment and included in the Clinical Study Characteristics table if it was apparent. Each outcome was examined separately for the quality elements listed and defined in Table 2 and each graded using the quality levels listed in Table 3. The main criteria considered in the rating of these elements are discussed below (see section 2.8.4 Grading of Evidence). Footnotes were used to

describe reasons for grading a quality element as having serious or very serious problems. The ratings for each component were summed to obtain an overall assessment for each outcome.

Table 2: Description of quality elements in GRADE for intervention studies

Quality element	Description
Limitations	Limitations in the study design and implementation may bias the estimates of the treatment effect. Major limitations in studies decrease the confidence in the estimate of the effect.
Inconsistency	Inconsistency refers to an unexplained heterogeneity of results.
Indirectness	Indirectness refers to differences in study population, intervention, comparator and outcomes between the available evidence and the review question, or recommendation made.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect relative to the clinically important threshold.
Publication bias	Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies.

Table 3: Levels of quality elements in GRADE

Level	Description
None	There are no serious issues with the evidence
Serious	The issues are serious enough to downgrade the outcome evidence by one level
Very serious	The issues are serious enough to downgrade the outcome evidence by two levels

4.3.5 Grading the quality of clinical evidence

After results were pooled, the overall quality of evidence for each outcome was considered. The following procedure was adopted when using GRADE:

1. A quality rating was assigned, based on the study design. RCTs start HIGH and observational studies as LOW, uncontrolled case series as LOW or VERY LOW.
2. The rating was then downgraded for the specified criteria: Study limitations, inconsistency, indirectness, imprecision and reporting bias. These criteria are detailed below. Observational studies were upgraded if there was: a large magnitude of effect, dose-response gradient, and if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect. Each quality element considered to have 'serious' or 'very serious' risks of bias were rated down -1 or -2 points respectively.
3. The downgraded/upgraded marks were then totalled and the overall quality rating was revised. For example, all RCTs started as HIGH and the overall quality became MODERATE, LOW or VERY LOW if 1, 2 or 3 points were deducted respectively.
4. The reasons or criteria used for downgrading were specified in the footnotes.

The details of criteria used for each of the main quality element are discussed further in the following sections 1.3.6 to 1.3.9.

Table 4: Overall quality of outcome evidence in GRADE

Level	Description
High	We are very confident that the true effect lies close to that of the estimate of the effect.

Level	Description
Moderate	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low	We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

4.3.6 Study limitations

The main limitations for randomised controlled trials are listed in Table 5.

Table 5: Study limitations of randomised controlled trials

Limitation	Explanation
Allocation concealment	Those enrolling patients are aware of the group to which the next enrolled patient will be allocated (major problem in "pseudo" or "quasi" randomised trials with allocation by day of week, birth date, chart number, etc.)
Lack of blinding	Patient, caregivers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated.
Incomplete accounting of patients and outcome events	Loss to follow-up not accounted
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results.
Other limitations	For example: Use of invalidated patient-reported outcomes

4.3.7 Inconsistency

Inconsistency refers to an unexplained heterogeneity of results. When estimates of the treatment effect across studies differ widely (i.e. heterogeneity or variability in results), this suggests true differences in underlying treatment effect. When heterogeneity exists (Chi square $p<0.1$ or I- squared inconsistency statistic of >50%), but no plausible explanation can be found, the quality of evidence was downgraded by one or two levels, depending on the extent of uncertainty to the results contributed by the inconsistency in the results. In addition to the I- square and Chi square values, the decision for downgrading was also dependent on factors such as whether the intervention is associated with benefit in all other outcomes or whether the uncertainty about the magnitude of benefit (or harm) of the outcome showing heterogeneity would influence the overall judgment about net benefit or harm (across all outcomes).

4.3.8 Indirectness

Directness refers to the extent to which the populations, intervention, comparisons and outcome measures are similar to those defined in the inclusion criteria for the reviews. Indirectness is important when these differences are expected to contribute to a difference in effect size, or may affect the balance of harms and benefits considered for an intervention.

4.3.9 Imprecision

Imprecision refers to the certainty in the effect for the outcome. When results are imprecise or very imprecise we are uncertain if there is an important difference between interventions or not.

Minimally importance difference (MID)

The thresholds of important benefits or harms, or the MID (minimally important difference) for an outcome are important considerations for determining whether there is a "clinically important" difference between interventions, and in assessing imprecision. For continuous outcomes, the MID is defined as "the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in the management" [37,53,113,114]. An effect estimate larger than the MID is considered to be "clinically important". For dichotomous outcomes, the MID is considered in terms of changes in both relative and absolute risk.

A literature search was conducted to pick up any relevant studies on MIDs in IPF as established MIDs are likely to be published and have probably been around long enough to be seen and accepted by clinical community. Given the poor-indexing in this field the GDG were also asked if they were aware of any published values for MIDs for the guideline outcomes. The following thresholds were identified and agreed with the GDG as the MIDs for the outcomes in this guideline:

- Six minute walk distance – 24-45m[28,43,123]
- Lung capacity (VC/FVC) – 2-6%[28]
- Transfer factor of the lung for carbon monoxide (TLCO) or DLCO – approximately 15%[17,31,56,64,68]
- SF-36 –2-4 points[123]
- St Georges respiratory questionnaire (SGRQ) – 5-8 points[123]
- EuroQol group 5-dimension self-reported questionnaire - approximately 0.08 for the self-report questionnaire and 7 points for the visual-analogue scale [100]

For several of the outcomes, there were no published MIDs. The GDG agreed that the default values stated in the GRADEpro were appropriate for these outcomes (see below). The default thresholds suggested by GRADE are a relative risk reduction of 25% (relative risk of 0.75 for negative outcomes) or a relative risk increase of 25% (risk ratio 1.25 for positive outcomes) for dichotomous outcomes.

For continuous outcome variables the MID is taken as half the median baseline standard deviation of that variable, across all studies in the meta-analysis. For example, if the median value of baseline standard deviations across all the meta-analysis studies is 10, then the MID will be +5. In such a case, the MID denoting the minimum clinically significant benefit will be +5 for a positive" outcome (for example, a quality of life measure where a higher score denotes better health), or -5 for a "negative" outcome (for example, a VAS pain score). Clinically significant harms will be the converse of these. If baseline values are unavailable, then half the median comparator group standard deviation of that variable will be taken as the MID.

If standardised mean differences have been used, then the MID will be set at the absolute value of + 0.5. This follows because standardised mean differences are mean differences normalised to the pooled standard deviation of the two groups, and are thus effectively expressed in units of "number of standard deviations". The 0.5 value in this context therefore indicates half a standard deviation, the same definition of MID as used for non-standardised mean differences.

For mortality, the GDG agreed to consider any reduction in mortality as a clinically important difference for patients with IPF.

The following are outcomes agreed by the GDG where the default MID would be applicable to assess imprecision and inform discussions on clinical importance:

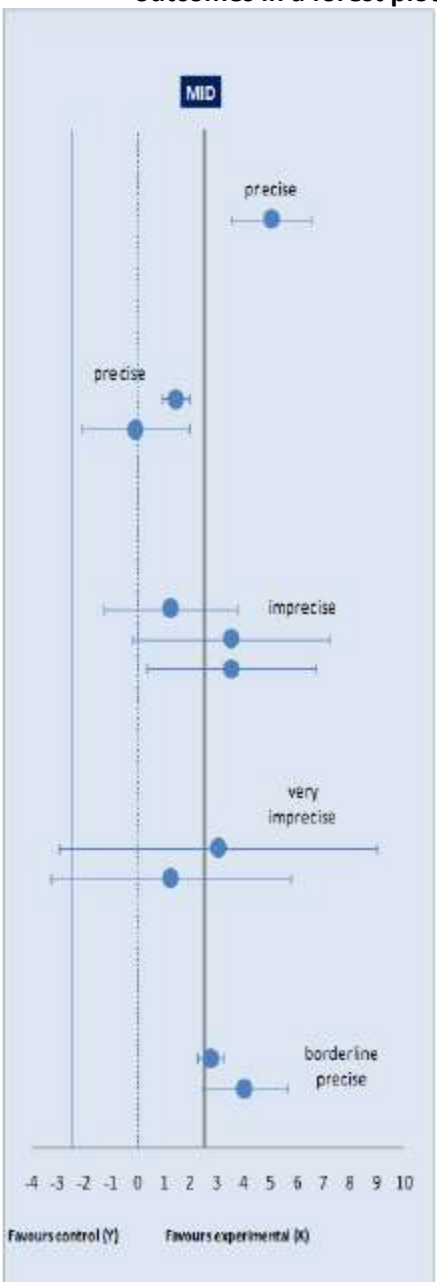
- Survival
- Hospitalisations due to IPF complications (including IPF exacerbations)
- Dyspnoea
- Time to disease progression
- Progression free survival

Assessing clinical importance and imprecision

The confidence interval for the pooled or best estimate of effect was considered in relation to the MIDs to assess imprecision. If the confidence interval crossed the MID threshold, there was uncertainty in the effect estimate supporting our recommendation (because the CI was consistent with two decisions) and the effect estimate was rated as having serious imprecision. If both MIDs were crossed, the effect estimate was rated as having very serious imprecision. In cases, where it was not possible to calculate imprecision (i.e. no baseline data was provided for a randomised controlled trial), then the effect estimate was rated as having very serious imprecision.

For the purposes of this guideline, clinical importance was assessed by comparing the relative effect estimate against the MID and reviewing the absolute effect reported in the GRADE summary table. For example, if the effect size was small (less than the MID), this finding suggests that there may not be enough difference to recommend one intervention over the other based on that outcome, unless in exceptional circumstances, the GDG agreed that the absolute effect was great enough to reach clinical importance. An effect estimate larger than the MID is considered to be clinically important. Figure 1 illustrates how the clinical importance of effect estimates was considered along with imprecision. This is documented in the evidence statements throughout this guideline.

Figure 1: Illustration of precise and imprecise outcomes based on the confidence interval of outcomes in a forest plot in relation to the MID



Evidence statements

Evidence statements were formed for each outcome indicating the quantity and quality of evidence available, and the outcome and population to which they relate.

4.4 Evidence of cost-effectiveness

Evidence on cost-effectiveness related to the key clinical issues being addressed in the guideline was sought. The health economist:

- Undertook a systematic review of the economic literature
- Undertook new cost-effectiveness analyses in priority areas

4.4.1 Literature review

The Health Economist:

- Identified potentially relevant studies for each review question from the economic search results by reviewing titles and abstracts – full papers were then obtained.
- Reviewed full papers against pre-specified inclusion / exclusion criteria to identify relevant studies (see below for details).
- Critically appraised relevant studies using the economic evaluations checklist as specified in The Guidelines Manual [83].
- Extracted key information about the study's methods and results into evidence tables (evidence tables are included in Appendix F.
- Generated summaries of the evidence in NICE economic evidence profiles (included in the relevant chapter write-ups) – see below for details.

4.4.1.1 Inclusion/exclusion

Full economic evaluations (studies comparing costs and health consequences of alternative courses of action: cost–utility, cost-effectiveness, cost-benefit and cost-consequence analyses) and comparative costing studies that addressed the review question in the relevant population were considered potentially applicable as economic evidence.

Abstracts were assessed for applicability and included in the clinical review and for economic evidence for three clinical areas (best supportive care, pulmonary rehabilitation and pharmacological interventions). If assessed as potentially applicable, the authors were contacted for further information.

Studies were excluded which only reported cost per hospital (not per patient), or only reported the average cost effectiveness without disaggregated costs and effects. Posters, reviews, letters/editorials, foreign language publications and unpublished studies were also excluded. Studies judged to have an applicability rating of 'not applicable' were excluded (this included studies that took the perspective of non-organisation for economic co-operation and development countries).

Remaining studies were prioritised for inclusion based on their relative applicability to the development of this guideline and the study limitations. For example, if a high quality, directly applicable UK analysis was available other less relevant studies may not have been included. Where exclusions occurred on this basis, this is noted in the relevant section.

For more details about the assessment of applicability and methodological quality see the economic evaluation checklist (The Guidelines Manual [83]) and the health economics research protocol in Appendix R.

When no relevant economic analysis was found from the economic literature review, relevant UK NHS unit costs related to the compared interventions were presented to the GDG to inform the possible economic implication of the recommendation to make.

4.4.1.2 NICE economic evidence profiles

The NICE economic evidence profile has been used to summarise cost and cost-effectiveness estimates. The economic evidence profile shows, for each economic study, an assessment of applicability and methodological quality, with footnotes indicating the reasons for the assessment. These assessments were made by the health economist using the economic evaluation checklist from The Guidelines Manual [83]. It also shows incremental costs, incremental outcomes (for example, QALYs) and the incremental cost-effectiveness ratio from the primary analysis, as well as information about the assessment of uncertainty in the analysis. See Table 6 for more details.

If a non-UK study was included in the profile, the results were converted into pounds sterling using the appropriate purchasing power parity [95].

Table 6: Content of NICE economic profile

Item	Description
Study	First author name, reference, date of study publication and country perspective.
Limitations	An assessment of methodological quality of the study*: Minor limitations – the study meets all quality criteria, or the study fails to meet one or more quality criteria, but this is unlikely to change the conclusions about cost effectiveness. Potentially serious limitations – the study fails to meet one or more quality criteria, and this could change the conclusion about cost effectiveness. Very serious limitations – the study fails to meet one or more quality criteria and this is very likely to change the conclusions about cost effectiveness. Studies with very serious limitations would usually be excluded from the economic profile table.
Applicability	An assessment of applicability of the study to the clinical guideline, the current NHS situation and NICE decision-making*: Directly applicable – the applicability criteria are met, or one or more criteria are not met but this is not likely to change the conclusions about cost effectiveness. Partially applicable – one or more of the applicability criteria are not met, and this might possibly change the conclusions about cost effectiveness. Not applicable – one or more of the applicability criteria are not met, and this is likely to change the conclusions about cost effectiveness.
Other comments	Particular issues that should be considered when interpreting the study.
Incremental cost	The mean cost associated with one strategy minus the mean cost of a comparator strategy.
Incremental effects	The mean QALYs (or other selected measure of health outcome) associated with one strategy minus the mean QALYs of a comparator strategy.
ICER	Incremental cost-effectiveness ratio: the incremental cost divided by the respective QALYs gained.
Uncertainty	A summary of the extent of uncertainty about the ICER reflecting the results of deterministic or probabilistic sensitivity analyses, or stochastic analyses of trial data, as appropriate.

*Limitations and applicability were assessed using the economic evaluation checklist from The Guidelines Manual [83]

Where economic studies compare multiple strategies, results are not reported in the standard economic profile but are instead presented at the end of the relevant chapter in an alternative table. The study is summarised as a whole in a descriptive manner.

4.4.2 Undertaking new health economic analysis

As well as reviewing the published economic literature for each review question, as described above, new economic analyses were undertaken by the Health Economist in priority areas. Priority areas for new health economic analysis were agreed by the GDG after formation of the review questions and consideration of the available health economic evidence.

Additional data for the analyses were identified as required through additional literature searches undertaken by the Health Economist, and discussion with the GDG. Model structure, inputs and assumptions were explained to and agreed by the GDG members during meetings, and they commented on subsequent revisions.

See Appendices I, K, J, L, M, O, for details of the health economic analyses undertaken for the guideline.

4.4.3 Cost-effectiveness criteria

NICE's report 'Social value judgements: principles for the development of NICE guidance' sets out the principles that GDGs should consider when judging whether an intervention offers good value for money[82].

In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

a. The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
b. The intervention cost less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

If the GDG recommended an intervention that was estimated to cost more than £20,000 per QALY gained, or did not recommend one that was estimated to cost less than £20,000 per QALY gained, the reasons for this decision are discussed explicitly in the 'from evidence to recommendations' section of the relevant chapter with reference to issues regarding the plausibility of the estimate or to the factors set out in the 'Social value judgements: principles for the development of NICE guidance'[81].

If a study reported the cost per life year gained but not QALYs, the cost per QALY gained was estimated by multiplying by an appropriate utility estimate to aid interpretation. The estimated cost per QALY gained is reported in the economic evidence profile with a footnote detailing the life-years gained and the utility value used. When QALYs or life years gained are not used in the analysis, results are difficult to interpret unless one strategy dominates the others with respect to every relevant health outcome and cost.

4.5 Developing recommendations

Over the course of the guideline development process, the GDG were presented with:

- Evidence tables of the clinical and economic evidence reviewed from the literature. All evidence tables are in Appendix F and G.
- Summary of clinical and economic evidence and quality (as presented in chapters 5 to 13).
- Forest plots (Appendix E).
- A description of the methods and results of the cost-effectiveness analysis undertaken for the guideline (Appendix L).

Recommendations were drafted on the basis of the GDG interpretation of the available evidence, taking into account the balance of benefits and harms, quality of evidence, and costs. When clinical and economic evidence was of poor quality, conflicting or absent, the GDG drafted recommendations based on consensus. Expert advisors were invited to provide advice on how to interpret the identified evidence. The considerations for making consensus based recommendations included the balance between potential harms and benefits, economic implications compared to the benefits, current practices, recommendations made in other relevant guidelines, patient preferences and equality issues. The consensus recommendations were done through discussions in the GDG, or methods of formal consensus were applied. The GDG considered whether the uncertainty was

sufficient to justify delaying making a recommendation to await further research, taking into account the potential harm of failing to make a clear recommendation.

The main considerations specific to each recommendation are outlined in the Evidence to Recommendation Sections preceding the recommendation section in each chapter.

This guideline recommends some drugs for indications for which they do not have a UK marketing authorisation at the date of publication, if there is evidence to support that use [7].

4.5.1 Research recommendations

When areas were identified for which good evidence was lacking, the guideline development group considered making recommendations for future research. Decisions about inclusion were based on factors such as:

- the importance to patients or the population
- national priorities
- potential impact on the NHS and future NICE guidance
- ethical and technical feasibility.

4.5.2 Validation process

The guidance is subject to a six week public consultation and feedback as part of the quality assurance and peer review the document. All comments received from registered stakeholders were responded to in turn and posted on the NICE website.

4.5.3 Updating the guideline

Following publication, the guideline will be reviewed and updated in line with the arrangements described in the Guidelines Manual.

4.5.4 Disclaimer

Healthcare providers need to use clinical judgement, knowledge and expertise when deciding whether it is appropriate to apply guidelines. The recommendations cited here are a guide and may not be appropriate for use in all situations. The decision to adopt any of the recommendations cited here must be made by the practitioners in light of individual patient circumstances, the wishes of the patient, clinical expertise and resources.

The National Clinical Guideline Centre disclaims any responsibility for damages arising out of the use or non-use of these guidelines and the literature used in support of these guidelines.

4.5.5 Funding

The National Clinical Guideline Centre was commissioned by the National Institute for Health and Care Excellence to undertake the work on this guideline.

5 Guideline summary

5.1 Algorithms

Algorithm to be developed as part of NICE pathways.

5.2 Key priorities for implementation

From the full set of recommendations, the GDG selected eleven key priorities for implementation. The criteria used for selecting these recommendations are listed in detail in The Guidelines Manual[83]. The reasons that each of these recommendations was chosen are shown in the table linking the evidence to the recommendation in the relevant chapter. The recommendations are listed in the order they appear in the NICE guideline.

Awareness of clinical features of idiopathic pulmonary fibrosis

- Be aware of idiopathic pulmonary fibrosis when assessing a patient with the clinical features listed below and when considering requesting a chest X-ray or referring to a specialist:
  - age over 45 years
  - persistent breathlessness on exertion
  - persistent cough
  - bilateral inspiratory crackles when listening to the chest
  - clubbing of the fingers
  - normal spirometry or impaired spirometry usually with a restrictive pattern but sometimes with an obstructive pattern.

Diagnosis

- Diagnose idiopathic pulmonary fibrosis only with the consensus of the multidisciplinary team (listed in table 1), based on:
  - the clinical features, lung function and radiological findings (see recommendation 3)
  - pathology when indicated (see recommendation 6).

Table 1 Minimum composition of multidisciplinary team involved in diagnosing idiopathic pulmonary fibrosis

Stage of diagnostic care pathway	Multidisciplinary team composition (all healthcare professionals should have expertise in interstitial lung disease)[a]
After clinical evaluation, baseline lung function and CT	Consultant respiratory physician Consultant radiologist Interstitial lung disease specialist nurse Multidisciplinary team coordinator
When considering performing bronchoalveolar lavage, and/or transbronchial biopsy or surgical lung biopsy Only some patients will have bronchoalveolar lavage or transbronchial biopsy but they may be being considered for surgical lung biopsy	Consultant respiratory physician Consultant radiologist Consultant histopathologist Thoracic surgeon as appropriate Interstitial lung disease specialist nurse Multidisciplinary team coordinator
When considering results of bronchoalveolar lavage, transbronchial biopsy or surgical lung biopsy	Consultant respiratory physician Consultant radiologist Consultant histopathologist Interstitial lung disease specialist nurse Multidisciplinary team coordinator

Information and support

- The consultant respiratory physician or interstitial lung disease specialist nurse should provide accurate and clear information (verbal and written) to people with idiopathic pulmonary fibrosis, and their families and carers with the person's consent. This should include information about investigations, diagnosis and management.

- An interstitial lung disease specialist nurse should be available at all stages of the care pathway to provide information and support to people with idiopathic pulmonary fibrosis and their families and carers with the person's consent.

[a] See chapter 6.5 (Multidisciplinary Team) for more information on the expertise of the multidisciplinary team.

Pulmonary rehabilitation

- Assess people with idiopathic pulmonary fibrosis for pulmonary rehabilitation at the time of diagnosis. Assessment may include a 6-minute walk test (distance walked and oxygen saturation measured by pulse oximetry) and a quality-of-life assessment.

Best supportive care

- Offer best supportive care to people with idiopathic pulmonary fibrosis from the point of diagnosis. Best supportive care should be tailored to disease severity, rate of progression, and the person's preference, and should include if appropriate:
  - information and support (see recommendation 2)
  - symptom relief
  - management of comorbidities
  - withdrawal of therapies suspected to be ineffective or causing harm
  - end of life care.

- If the person is breathless on exertion consider assessment for:
  - the causes of breathlessness and degree of hypoxia and
  - ambulatory oxygen therapy and long-term oxygen therapy and/or
  - pulmonary rehabilitation.

Disease-modifying pharmacological interventions

- For guidance on pirfenidone, see the NICE technology appraisal on pirfenidone for the treatment of idiopathic pulmonary fibrosis. For guidance on nintedanib, see the NICE technology appraisal on nintedanib for the treatment of idiopathic pulmonary fibrosis.

- Do not use any of the drugs below, either alone or in combination, to modify disease progression in idiopathic pulmonary fibrosis:
  - ambrisentan
  - azathioprine
  - bosentan
  - co-trimoxazole
  - mycophenolate mofetil
  - prednisolone

- sildenafil
- warfarin.

Lung transplantation

- Refer people with idiopathic pulmonary fibrosis for lung transplantation assessment if they wish to explore lung transplantation and if there are no absolute contraindications. Ask the transplant centre for an initial response within 4 weeks.

Review and follow-up

- In follow-up appointments for people with idiopathic pulmonary fibrosis:
  - assess lung function
  - assess for oxygen therapy
  - assess for pulmonary rehabilitation
  - offer smoking cessation advice, in line with Smoking cessation services (NICE public health guidance 10)
  - identify exacerbations and previous respiratory hospital admissions
  - consider referral for assessment for lung transplantation in people who do not have absolute contraindications (see recommendations 32 and 33)
  - consider psychosocial needs and referral to relevant services as appropriate
  - consider referral to palliative care services
  - assess for comorbidities (which may include anxiety, bronchiectasis, depression, diabetes, dyspepsia, ischaemic heart disease, lung cancer and pulmonary hypertension).

5.3 Full list of recommendations

Awareness of clinical features of idiopathic pulmonary fibrosis

1. Be aware of idiopathic pulmonary fibrosis when assessing a patient with the clinical features listed below and when considering requesting a chest X-ray or referring to a specialist:
   - age over 45 years
   - persistent breathlessness on exertion
   - persistent cough
   - bilateral inspiratory crackles when listening to the chest
   - clubbing of the fingers

- normal spirometry or impaired spirometry usually with a restrictive pattern but sometimes with an obstructive pattern.

Diagnosis

2. The consultant respiratory physician or interstitial lung disease specialist nurse should provide accurate and clear information (verbal and written) to people with idiopathic pulmonary fibrosis, and their families and carers with the person's consent. This should include information about investigations, diagnosis and management.
3. Assess everyone with suspected idiopathic pulmonary fibrosis by:
   - taking a detailed history, carrying out a clinical examination (see recommendation 1 for clinical features) and performing blood tests to help exclude alternative diagnoses, including lung diseases associated with environmental and occupational exposure, with connective tissue diseases and with drugs, and
   - performing lung function testing (spirometry and gas transfer) and
   - reviewing results of chest X-ray and
   - performing CT of the thorax (including high-resolution images).
4. Diagnose idiopathic pulmonary fibrosis only with the consensus of the multidisciplinary team (listed in table 19), based on:
   - the clinical features, lung function and radiological findings (see recommendation 3)
   - pathology when indicated (see recommendation 6).
5. At each stage of the diagnostic care pathway the multidisciplinary team should consist of a minimum of the healthcare professionals listed in table 19, all of whom should have expertise in interstitial lung disease.

Table 19: Minimum composition of multidisciplinary team involved in diagnosing idiopathic pulmonary fibrosis

Stage of diagnostic care pathway	**Multidisciplinary team composition (all healthcare professionals should have expertise in interstitial lung disease)**[b]
After clinical evaluation, baseline lung function and CT	• **Consultant respiratory physician** • **Consultant radiologist** • **Interstitial lung disease specialist nurse** • **Multidisciplinary team coordinator**

[b] See chapter 6.5 (Multidisciplinary Team) for more information on the expertise of the multidisciplinary team.

When considering performing bronchoalveolar lavage, and/or transbronchial biopsy or surgical lung biopsy **Only some patients will have bronchoalveolar lavage or transbronchial biopsy but they may be being considered for surgical lung biopsy**	• **Consultant respiratory physician** • **Consultant radiologist** • **Consultant histopathologist** • **Thoracic surgeon as appropriate** • **Interstitial lung disease specialist nurse** • **Multidisciplinary team coordinator**
When considering results of bronchoalveolar lavage, transbronchial biopsy or surgical lung biopsy	• **Consultant respiratory physician** • **Consultant radiologist** • **Consultant histopathologist** • **Interstitial lung disease specialist nurse** • **Multidisciplinary team coordinator**

6. If the multidisciplinary team cannot make a confident diagnosis from clinical features, lung function and radiological findings, consider:
   - bronchoalveolar lavage or transbronchial biopsy and/or
   - surgical lung biopsy, with the agreement of the thoracic surgeon.
7. Discuss with the person who may have idiopathic pulmonary fibrosis:
   - the potential benefits of having a confident diagnosis compared with the uncertainty of not having a confident diagnosis and
   - the increased likelihood of obtaining a confident diagnosis with surgical biopsy compared with bronchoalveolar lavage or transbronchial biopsy and
   - the increased risks of surgical biopsy compared with bronchoalveolar lavage or transbronchial biopsy.
8. When considering bronchoalveolar lavage, transbronchial biopsy or surgical lung biopsy take into account:
   - the likely differential diagnoses and
   - the person's clinical condition, including any comorbidities.
9. If a confident diagnosis cannot be made continue to review the person under specialist care.

Prognosis

10. Measure the initial rate of decline in the person's condition, which may predict subsequent prognosis, by using lung function test results (spirometry and gas transfer) at:
    - diagnosis and
    - 6 months and 12 months after diagnosis. Repeat the lung function tests at shorter intervals if there is concern that the person's condition is deteriorating rapidly.

11. Do not use the 6-minute walk distance at diagnosis to estimate prognosis. (The 6-minute walk test may be useful for other purposes, see recommendation 14).
12. The consultant respiratory physician or interstitial lung disease specialist nurse should provide accurate and clear information (verbal and written) to people with idiopathic pulmonary fibrosis, and their families and carers with the person's consent. This should include information about investigations, diagnosis and management.
13. Discuss prognosis with people with idiopathic pulmonary fibrosis in a sensitive manner and include information on:
    - the severity of the person's disease and average life expectancy
    - the varying courses of disease and range of survival
    - management options available.

Pulmonary rehabilitation

14. Assess people with idiopathic pulmonary fibrosis for pulmonary rehabilitation at the time of diagnosis. Assessment may include a 6-minute walk test (distance walked and oxygen saturation measured by pulse oximetry) and a quality-of-life assessment.
15. Repeat the assessment for pulmonary rehabilitation for people with idiopathic pulmonary fibrosis at 6-month or 12-month intervals.
16. If appropriate after each assessment, offer pulmonary rehabilitation including exercise and educational components tailored to the needs of people with idiopathic pulmonary fibrosis in general.
17. Pulmonary rehabilitation should be tailored to the individual needs of each person with idiopathic pulmonary fibrosis. Sessions should be held somewhere that is easy for people with idiopathic pulmonary fibrosis to get to and has good access for people with disabilities.

Best supportive care

18. Offer best supportive care to people with idiopathic pulmonary fibrosis from the point of diagnosis. Best supportive care should be tailored to disease severity, rate of progression, and the person's preference, and should include if appropriate:
    - information and support (see recommendation 2)
    - symptom relief
    - management of comorbidities
    - withdrawal of therapies suspected to be ineffective or causing harm
    - end of life care.
19. If the person is breathless on exertion consider assessment for:
    - the causes of breathlessness and degree of hypoxia and
    - ambulatory oxygen therapy and long-term oxygen therapy and/or
    - pulmonary rehabilitation.
20. If the person is breathless at rest consider:
    - assessment for the causes of breathlessness and degree of hypoxia and

- assessment for additional ambulatory oxygen therapy and long-term oxygen therapy and
- the person's psychosocial needs and offering referral to relevant services such as palliative care services and
- pharmacological symptom relief with benzodiazepines and/or opioids.

21. Assess the oxygen needs of people who have been hospitalised with idiopathic pulmonary fibrosis before they are discharged.

22. If the person has a cough consider:

- treatment for causes other than idiopathic pulmonary fibrosis (such as gastro-oesophageal reflux disease, post-nasal drip)
- treating with opioids if the cough is debilitating
- discussing treatment with thalidomide[c] with a consultant respiratory physician with expertise in interstitial lung disease if the cough is intractable.

23. Ensure people with idiopathic pulmonary fibrosis, and their families and carers, have access to the full range of services offered by palliative care teams. Ensure there is collaboration between the healthcare professionals involved in the person's care, community services and the palliative care team.

24. NICE has produced guidance on the components of good patient experience in adult NHS services. Follow the recommendations in Patient experience in adult NHS services (NICE clinical guideline 138).

25. An interstitial lung disease specialist nurse should be available at all stages of the care pathway to provide information and support to people with idiopathic pulmonary fibrosis and their families and carers with the person's consent.

26. Offer advice, support and treatment to aid smoking cessation to all people with idiopathic pulmonary fibrosis who also smoke, in line with Smoking cessation services (NICE public health guidance 10).

Pharmacological interventions

There is no conclusive evidence to support the use of any drugs to increase the survival of people with idiopathic pulmonary fibrosis.

27. For guidance on pirfenidone, see the NICE technology appraisal on pirfenidone for the treatment of idiopathic pulmonary fibrosis. For guidance on nintedanib, see the NICE technology appraisal on nintedanib for the treatment of idiopathic pulmonary fibrosis.

28. Do not use any of the drugs below, either alone or in combination, to modify disease progression in idiopathic pulmonary fibrosis:

- ambrisentan
- azathioprine
- bosentan

[c] At the time of publication (June 2013), thalidomide did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

- co-trimoxazole
- mycophenolate mofetil
- prednisolone
- sildenafil
- warfarin.

29. Advise the person that oral *N*-acetylcysteine[d] is used for managing idiopathic pulmonary fibrosis, but its benefits are uncertain.

30. If people with idiopathic pulmonary fibrosis are already using prednisolone or azathioprine, discuss the potential risks and benefits of discontinuing, continuing or altering therapy.

31. Manage any comorbidities according to best practice. For gastro-oesophageal reflux disease, see Managing dyspepsia in adults in primary care (NICE clinical guideline 17).

Lung transplantation

32. Discuss lung transplantation as a treatment option for people with idiopathic pulmonary fibrosis who do not have absolute contraindications. Discussions should:

    - take place between 3 and 6 months after diagnosis or sooner if clinically indicated
    - be supported by an interstitial lung disease specialist nurse
    - include the risks and benefits of lung transplantation
    - involve the person's family and carers with the person's consent.

    (See recommendations 18 – 23 about best supportive care.)

33. Refer people with idiopathic pulmonary fibrosis for lung transplantation assessment if they wish to explore lung transplantation and if there are no absolute contraindications. Ask the transplant centre for an initial response within 4 weeks.

Ventilation

34. A respiratory physician or specialist nurse with an interest in interstitial lung disease should discuss the poor outcomes associated with mechanical ventilation (including non-invasive mechanical ventilation) for respiratory failure with people with idiopathic pulmonary fibrosis. These discussions should ideally take place between 3 to 6 months after diagnosis or sooner if clinically indicated. (See recommendations 18 – 23 about best supportive care.)

35. Do not routinely offer mechanical ventilation (including non-invasive mechanical ventilation) to people with idiopathic pulmonary fibrosis who develop life-threatening respiratory failure.

Review and follow-up

36. In follow-up appointments for people with idiopathic pulmonary fibrosis:

[d] At the time of publication (June 2013), *N*-acetylcysteine did not have a UK marketing authorisation. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

- assess lung function
- assess for oxygen therapy
- assess for pulmonary rehabilitation
- offer smoking cessation advice, in line with Smoking cessation services (NICE public health guidance 10)
- identify exacerbations and previous respiratory hospital admissions
- consider referral for assessment for lung transplantation in people who do not have absolute contraindications (see recommendations 32 and 33)
- consider psychosocial needs and referral to relevant services as appropriate
- consider referral to palliative care services
- assess for comorbidities (which may include anxiety, bronchiectasis, depression, diabetes, dyspepsia, ischaemic heart disease, lung cancer and pulmonary hypertension).

37. Consider follow-up of people with idiopathic pulmonary fibrosis:
    - every 3 months or sooner if they are showing rapid disease progression or rapid deterioration of symptoms or
    - every 6 months or sooner if they have steadily progressing disease or
    - initially every 6 months if they have stable disease and then annually if they have stable disease after 1 year.

5.4 Key research recommendations

1. What is the value of bronchoalveolar lavage in people in whom idiopathic pulmonary fibrosis is considered the most likely diagnosis when clinical and CT findings are insufficient to support a confident diagnosis?

2. What is the value of surgical lung biopsy in people in whom idiopathic pulmonary fibrosis is considered the most likely diagnosis when clinical and CT findings are insufficient to support a confident diagnosis?

3. Does pulmonary rehabilitation improve outcomes for people with idiopathic pulmonary fibrosis?

4. Does ambulatory oxygen improve outcomes in idiopathic pulmonary fibrosis?

5. Is anti-reflux therapy an effective treatment for idiopathic pulmonary fibrosis?

6 Diagnosis

6.1 Review introduction

Idiopathic Pulmonary Fibrosis (IPF) is a chronic, progressive fibrotic interstitial lung disease of unknown origin. Each year in the UK, approximately 5,000 new cases are diagnosed and the incidence is rising. People with IPF typically present in their sixties or seventies and it is more common in men than women. The median survival from diagnosis is about three years, a prognosis which is worse than many cancers. The diagnosis of IPF depends on thinking of it as a cause of breathlessness or cough. It may be suspected on the basis of symptoms, signs and abnormalities on a chest radiograph.

Further investigation of IPF requires a more detailed clinical assessment, a CT scan of the thorax and sometimes, surgical lung biopsy. The differential diagnosis includes interstitial lung disease associated with connective tissue disease, occupational and environmental lung disease, drug-induced lung disease, non-specific interstitial pneumonia, and sarcoidosis. These conditions must be excluded in cases of suspected IPF. Bronchoalveolar lavage (BAL) and bronchoscopic/ transbronchial lung biopsy (TBB) can be helpful in this regard. Diagnostic accuracy of IPF increases if a multidisciplinary team (MDT) is involved.

6.1.1 Bronchoalveolar lavage (BAL)

In people with suspected IPF, the differential cell count obtained from BAL may help distinguish IPF from other fibrotic lung diseases. There are no features on the differential cell count that are diagnostic of IPF. In people with suspected IPF where there is diagnostic uncertainty following clinical and CT assessment, a lymphocytosis in BAL fluid may indicate alternative diagnoses such as hypersensitivity pneumonitis or sarcoidosis.

6.1.2 Bronchoscopic biopsy /Transbronchial biopsy (TBB)

The histopathological features of IPF are a usual interstitial pneumonia (UIP) pattern which shows characteristic spatial and temporal heterogeneity within the lung tissue. Bronchoscopic biopsies only sample the large airways and are not useful for diagnosing IPF, but may be useful in supporting an alternative diagnosis such as sarcoidosis. TBBs provide only small samples of lung tissue; abnormalities must be interpreted with caution. Findings on TBB are not useful for diagnosing IPF, but may be helpful in supporting an alternative diagnosis.

6.1.3 Surgical lung biopsy (SLB)

In some people with suspected IPF, a surgical lung biopsy is an appropriate procedure to provide diagnostic information. A histological pattern of UIP is required to support the diagnosis of IPF in this context.

6.1.4 Multidisciplinary team (MDT)

The diagnosis of IPF requires careful integration of clinical, radiological and sometimes histological findings. In this regard, the process is not dissimilar to the diagnostic pathways used for example in lung cancer in which engaging a multidisciplinary team is usual practice. An MDT for IPF might include a chest physician, a radiologist and a histopathologist with expertise in ILD, a thoracic surgeon with whom surgical biopsy can be discussed and specialist ILD nurse.

6.2 Clinical questions and review methodology

The following clinical questions were included in this chapter.

For full details see review protocols in Appendix C.

6.2.1 Biopsy/BAL

In suspected IPF what is the value of **adding biopsy to clinical evaluation, PFTs, CT +/- bronchoalveolar lavage** for confirming the diagnosis of IPF?

Table 7: PICO characteristics of biopsy review question

Population:	Adults with suspected ILD
Intervention:	• Baseline clinical assessment (history, PFTs, CT) and: o +/-Bronchoalveolar lavage • Bronchoscopic biopsy/transbronchial biopsy • Surgical biopsy (open lung biopsy (OLB) or video assisted thoracic surgery (VATS))
Comparison:	• Baseline clinical assessment (history, PFTs, CT, +/- BAL)
Outcomes:	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates • Sensitivity • Specificity Other outcomes • Adverse events • Improvement in health-related quality of life
Study design:	Cohort studies

6.2.2 MDT

In suspected IPF what is the value of **adding multidisciplinary team (MDT) consensus** to clinical assessment, PFTs and CT in the diagnosis of IPF?

Table 8: PICO characteristics of MDT review question

Population:	Adults with suspected ILD
Intervention:	• MDT 1: Clinical assessment + radiological assessment + MDT consensus • MDT 2: Clinical assessment + radiological assessment +/- bronchoalveolar lavage + MDT consensus • MDT 3: Clinical assessment + radiological assessment +/- bronchoalveolar lavage + bronchoscopic/ transbronchial biopsy +/- surgical biopsy (open-lung or video assisted biopsy) + MDT
Comparison:	The following procedures alone or in combination: • Clinical assessment • Radiological assessment • Bronchoalveolar lavage, bronchoscopic/ transbronchial biopsy • Surgical lung biopsy (open lung and video assisted biopsy)

Outcomes:	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates • Sensitivity • Specificity Other outcomes • Adverse events • Improvement in health-related quality of life
Study design:	Cohort studies

How and by whom is a MDT diagnostic consensus best achieved (i.e. constituency of the MDT, specialist clinics, networks)?

Table 9: PICO characteristics of review question

Population:	Adults with suspected ILD
Intervention:	MDT consisting of Respiratory Physician (RP) + Radiologist (R) + Histopathologist (HP) in tertiary referral hub as part of wider network

Comparison:	• Health professionals (RP or R or HP) in isolation
	• Health professionals (+/- RP +/-R +/- HP) in MDT
	  ○ secondary care
	  ○ tertiary care
	  ○ network of referral between secondary hospitals
	  ○ network of referral between secondary and tertiary hospitals

Outcomes:	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates • Sensitivity • Specificity Other outcomes • Adverse events • Improvement in health-related quality of life
Study design:	Cohort studies

The objectives of the clinical questions were to determine:

- the added benefit of a biopsy (bronchoalveolar lavage +/- bronchoscopic biopsy/ transbronchial biopsy or surgical lung biopsy) in the diagnosis of a patient with suspected IPF, when clinical history, pulmonary function tests (PFTs), and CT +/- bronchoalveolar lavage have all been conducted.
- whether MDT consensus provides an additional benefit to diagnosis of people with IPF.
- what requirements an MDT should fulfil in order to provide optimal clinical care to people with IPF.

The literature was searched for all years for studies assessing the additional value of adding biopsy and MDT consensus to standard clinical assessment in the diagnosis of IPF, as well as the MDT constituency.

Inclusion criteria were as follows:

- Any duration of follow-up
- Any sample size
- Population ≥18 years, with suspected interstitial lung disease (ILD)
- Study design: diagnostic cohorts, (prospective and retrospective)
- Studies published post 1994 (studies that span inclusion of subjects pre 1994 are also included).

Note: A modified version of GRADE profile and an additional narrative summary has been used in this evidence review to analyse and present the evidence. The statistics used for this diagnostic review differ from those used in intervention reviews, and a definition for each of them is provided below (Table 11).

Table 10: Definitions of diagnostic terms

Term	Definition
Index test	Test of interest
Reference standard	Best available method of determining disease status

Table 11: Definitions of summary statistics for diagnostic accuracy studies

Measure	Definition
True positives (TP)	Correct positive test result - number of people diagnosed with IPF with a positive index test result
True negatives (TN)	Correct negative test results - number of people diagnosed as not having IPF with a negative index test result
False positives (FP)	Incorrect positive test result - number of people diagnosed as not having IPF with a positive index test result
False negatives (FN)	Incorrect negative test result - number of people with IPF with a negative index test result
Sensitivity	Proportion of those with the disease (based on reference standard) who are *positive* on the index test
Specificity	Proportion of those without the disease (based on reference standard) who are *negative* on the index test
Positive predictive value (PPV)	Probability of having the disease in a patient with a *positive* index test result
Negative predictive value (NPV)	Probability of not having the disease in a patient with a *negative* index test result

Note: Positive and negative predictive values are dependent on disease prevalence (pre-test probability) and so need to be interpreted together with prevalence, in the context of how test results modify the probability of disease (post-test probabilities). Consider that the lower the prevalence of disease the more certain we can be that a negative test indicates no disease, and the less certain that a positive result truly indicates the presence of disease. A note on how to interpret post-test probabilities/predictive values in the light of the disease prevalence is provided in Appendix H.

A summary of the included index tests is provided in Table 13.

Table 12: Description of index tests being assessed for diagnostic accuracy

Test	Description
Clinical evaluation	Basic clinical examination
Pulmonary Function Tests (PFTs)	Forced vital capacity (FVC), gas transfer of carbon monoxide (DLCO)
High resolution CT	May also be referred to as CT

6.3 Clinical evidence

6.3.1 Bronchoalveolar Lavage

6.3.1.1 Overview

One study was found: Ohshimo 2009[91]. This investigated the use of BAL in diagnosing people with IPF.

6.3.1.2 Quality (QUADAS II)

The study was retrospective; included patients had IPF diagnosed by the American Thoracic Society (ATS)/ European Respiratory Society (ERS) criteria.

The index test was CT findings blinded to BAL and clinical results.

The reference standard used in the study was BAL. It was unclear whether these results were interpreted blinded to the results of the index test.

The study largely avoided verification bias (i.e. all patients in the study received BAL, regardless of initial results, and were included in the analysis).

There was an unclear period of time between the index test and reference standard.

6.3.1.3 Results

The final diagnosis was IPF in 68 patients, non-specific interstitial pneumonia (NSIP) in 3 patients and extrinsic allergic alveolitis (EAA) in also 3 patients. Six patients had a change in diagnosis following BAL.

The study did not provide enough detail to calculate sensitivity, specificity, NPV and PPV.

6.3.2 Bronchoscopic/ transbronchial biopsy

6.3.2.1 Overview

One study was identified: Oliveira 2011[92]

6.3.2.2 Quality (QUADAS II)

The study was a retrospective cohort. Some patients had already undergone biopsy prior to entering the study. All patients received a transbronchial biopsy as the reference standard. It was unclear whether the results of this were interpreted blinded to the results of the index test.

The study avoided verification bias (i.e. all patients in the studies received a biopsy, regardless of initial results, and were included in the analysis).

6.3.2.3 Results

The study did not provide enough detail to calculate sensitivity, specificity, positive predictive values (PPVs) and negative predictive values (NPV).

6.3.3 Surgical lung biopsy (video-assisted and open lung biopsy)

6.3.3.1 Overview

Sixteen studies were identified: Aalokken 2012[4], Coutinho 2008[18], Ishie 2009[51], Flaherty 2002[33], Jamaati 2006 [54], Lettieri 2005A[69], Lettieri 2005[70], Oliveira 2011[92], Ooi 2005[93], Peckham 2004[98], Rena 1999[106], Sigurdsson 2009[116], Slodkowska 2000[117], Trahan 2008A[128], Vansteenkiste 1999[129] and Yamaguchi 2004[131]. Three of the studies were pre-2002 and therefore did not use ATS/ERS diagnostic criteria: Rena 1999[106], Vansteenkiste 1999[129] and Slodkowska 2000[117].

6.3.3.2 Quality (QUADAS II)

Most of the studies were retrospective cohorts, except for Rena 1999[106] which was a prospective cohort. The majority of studies included people with suspected ILD. Jamaati 2006[54] included people with suspected IPF. Slodkowska 2000[117] included patients already diagnosed with IPF/UIP.

The index test was clinical findings, PFTs and CT findings in most papers. Coutinho 2008[18], Lettieri 2005A[69], Lettieri 2005[70] and Slodkowska 2000[117] did not specify whether all patients received the index tests. In Oliveira 2011[92] and Rena 1999[106], patients had already undergone biopsy before entering the study.

Aalokken 2012[4] was a comparison of SLB against the reference standard of an MDT.

Studies which did not report American Thoracic Society (ATS) diagnostic criteria and studies which were conducted pre-ATS (pre-2002) were downgraded, but included in this review.

6.3.3.3 Results

Most studies did not provide enough detail to calculate sensitivity, specificity, NPV and PPV. However, for a clinical diagnosis, Coutinho 2008[18] reported a sensitivity of 67% (57-75), a specificity of 90% (85-93), a PPV of76% (67-84) and an NPV of 85% (80-89).

Peckham 2004[98] reported (for CT), a sensitivity of 71% (51-92), a specificity of 67% (39-86%), a PPV of 71% (51-92%) and an NPV of 76% (39-86%). For ATS clinical criteria a sensitivity of 71% (51-92), a specificity of 75% (47-92%), a PPV of 77% (50-92%) and an NPV of73% (54-86%) were reported.

Coutinho 2008[18] reported a "correct diagnosis" in 76% (n=80), a "new diagnosis" in 21% (n=22) and the biopsy was "inconclusive" in 3% (n=3).

Aalokken 2012[4] reported the sensitivity of a histological diagnosis of UIP to be 73%, the specificity 74%, the PPV 83% and the NPV 61%. This was against the reference standard of MDT consensus consisting of a radiologist and histopathologist who were blinded to results of the initial diagnosis.

The final diagnosis varied between IPF, interstitial fibrosis, UIP and idiopathic interstitial pneumonia (IIP) in the papers. A diagnosis of another ILD was provided by some papers.

6.3.4 Multidisciplinary Team (MDT)

6.3.4.1 Overview

From the initial search 13 papers were identified as MDT related, and of these 4 papers were excluded (see Appendix R).

The 9 papers which were included in the review were; Hunninghake 2001[49], Flaherty 2003A[32], Flaherty 2007[34], Lynch 2005[72], Sumikawa 2008[121], Sverzellati 2010[122], Thomeer 2008 [126], Spencer 2011[119] and Raghu 1999[103].

6.3.4.2 Quality

The studies were largely retrospective cohorts, with the exception of Raghu 1999[103] and Hunninghake 2001[49] which were prospective cohorts.

There were variable patient selection criteria; only 3 papers included people with suspected IPF, IIP or ILD [49,34,103]. The other 6 papers included patients who had a confirmed histological diagnosis of IPF, in 2 of these papers patients had been enrolled onto clinical trials [126, 72].

The index test was radiological diagnosis of IPF with or without clinical information, in the majority of papers [49,32,121,122,126,103]. Other papers used the level of experience and expertise of the assessor, the index test being the diagnosis of assessor with the least experience, which was either shown through setting (community vs. academic) or number of years of experience[34, 72]. Spencer 2011[119] had the referral centre as the index test.

The reference standard was pathological/histological diagnosis in most papers [49,32,121,122,126,103]. Spencer 2011[119] had an MDT in a tertiary centre as the reference standard. Flaherty 2007 [34] and Lynch 2005[72] used the level of experience and expertise of the assessor, the reference standard being the diagnosis of the assessor with the most experience, which was either shown through setting (community vs. academic) or number of years of experience.

The studies avoided verification bias (i.e. all patients in the studies received the same comparison tests, regardless of initial results, and were included in the analysis).

There was an unclear period of time between the index test and reference standard in the majority of studies. However most studies were retrospective reviews of patient data therefore the flow and timing weren't an important consideration for quality.

6.3.4.3 Results

Two studies were identified[49,126] that explored the level of accuracy displayed by agreement between clinicians. Hunninghake 2001[49] suggested that a single clinician in a referral centre was more likely to diagnose an ILD patient with IPF than a group of clinicians of the same speciality (clinical cores) in liaison with each other. This meant that although the positive predictive power of the referring clinician was lower than the clinical cores, the negative predictive power of the referring clinician was comparable or higher than the cores if the same starting prevalence of IPF was assumed. By placing the referring clinician at the first stage in a diagnostic pathway (where those identified without IPF left the diagnostic pathway but those identified with IPF remained), this would have the consequence of screening people who do not have IPF and improve the positive predictive power of subsequent diagnostic interventions.

Thomeer and colleagues [126] showed that when findings of a radiologist and a histopathologist were consulted, it was more likely that the radiologist would give a diagnosis of IPF, whereas the histopathologist was more likely to state that IPF was absent. According to this evidence, if biopsy was placed at the second stage of the diagnostic pathway, we would expect the number of IPF

diagnoses to decrease – potentially resulting in fewer false positives (if biopsy is believed to be the gold standard) or more false negatives (if the CT findings are to be believed).

Sumikawa[121] suggested that in those with IPF, the CT findings would not concur with the findings of the biopsy in 29% of cases. Sverzellati [122] suggested that where the probability of IPF was 45%, CT and biopsy findings would strongly conflict in at least 64% of IPF cases, and moderately conflict in a further 10% of IPF cases. At this stage of the diagnostic pathway we would expect the level of agreement between the specialities to be moderate to low (i.e. with a kappa statistic of approximately 0.4) (Flaherty 2007[34]).

However, inter-observer agreement between specialities can increase if specialist cores (i.e. more than one clinician of the same speciality) are involved in the multidisciplinary agreement exercise (similar to a second opinion rather than consensus discussion). For example, Lynch (2005)[72] showed that if CT scans were interpreted by two radiologists, CT findings and biopsy findings agreed 88.3% of the time, with 11.7% of findings conflicting. These authors suggest that interpretation of CT by a clinical core is more likely to agree with biopsy results than a sole clinician operating at a study site.

There is conflicting evidence to suggest whether levels of expertise within the clinicians influence the degree of agreement within and between the specialities. Lynch (2005)[72] suggests that the level of agreement between members of a core is not likely to be influenced by whether the clinician is in an academic or in a local community setting; however, Flaherty and colleagues (2007)[34] suggest otherwise. These authors demonstrate substantially higher agreement within academic specialists' cores at each stage of a diagnostic pathway than community specialists (i.e. a kappa statistic of 0.71 for an academic specialist core post MDT consensus versus a kappa statistic of 0.44 for a community specialist core post MDT consensus).

If we assume that a higher level of agreement within speciality cores infers a higher level of precision when interpreting results (therefore affecting the accuracy of the final diagnosis), then referring people with discordant findings between CT and biopsy to another clinician of the same speciality in an academic centre is likely to reduce the number of inaccurate diagnoses of IPF.

Using data reported by Flaherty et al (2007)[34]ranges for sensitivity and specificity can be calculated. The data shows that sensitivity and specificity in academic and community settings, increases with MDT consensus for the diagnosis of IPF (see extraction table in Appendix F)

Flaherty 2003[32] reported the course of disease progression – which may be used as a reference standard to establish whether a correct diagnosis may have been made. In this paper, we can see that for those who were diagnosed with IPF by a UIP on biopsy, 37% of CT findings would have strongly disagreed and a further 27% of CT findings would have moderate disagreement. Where biopsy failed to identify IPF, CT findings would moderately disagree with 22% of cases. However, in cases where the specialities did not agree, a different trajectory of survival could be observed. It could be argued that multidisciplinary discussion may not only improve the certainty of a diagnosis through agreement, but in cases where no consensus can be reached it can inform on prognosis.

Spencer 2011[119] reported the number of cases diagnosed as having either 'probable' or 'definite' IPF by the referring centre, being confirmed or having a change in diagnosis by the MDT in the tertiary centre.

Evidence is summarised in the modified GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

6.3.5 Summary of diagnostic accuracy and quality of studies for BAL, biopsy and MDT

Studies which only gave diagnostic yield were considered very low quality and were not graded and have only been included for information (see table 15).

No studies reported diagnostic accuracy, only diagnostic yield for BAL or TBB (see table 15)

6.3.5.1 Evidence profile

Table 13: Modified GRADE table for the diagnostic accuracy of BAL, Biopsy and MDT

Study ID	Design	No. of patients	Limitation	Inconsistency	Indirectness	Imprecision	Pre-test probability	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Quality
Surgical lung biopsy												
Aalokken 2012[4]	Retrospective (histological diagnosis)	64	Serious	N/A	N/A	N/A	NR	73%	74%	83%	61%	Moderate
Coutinho 2008[18]	Retrospective	120	Serious	N/A	Serious	N/A	NR	Clinical diagnosis: 67 (57-75)	Clinical diagnosis: 90 (85-93)	Clinical diagnosis: 76 (67-84)	Clinical diagnosis: 85 (80-89)	Low
Peckham 2004[98]	Retrospective	26	Serious	N/A	Serious	N/A	NR	CT:71% (51-92) ATS clinical criteria: 71 (51-92)	CT: 67% (39-86%) ATS clinical criteria: 75% (47-92%)	CT: 71% (51-92%) ATS clinical criteria: 77% (50-92%)	CT: 76% (39-86%) ATS clinical criteria: 73% (54-86%)	Low
MDT												

Study ID	Design	No. of patients	Limitation	Inconsistency	Indirectness	Imprecision	Pre-test probability	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Quality
Hunninghake 2001[49]	Prospective Results from referring centre of overall IPF diagnosis	91	Serious	N/A	N/A	N/A	NR	46/54 (85%)	16/37 (43%)	46/67 (69%)	NR	Moderate
Hunninghake 2001[49]	Prospective Results from clinical core of overall IPF diagnosis	91	Serious	N/A	N/A	N/A	NR	39/54 (72%)	31/37 (84%)	39/45 (87%)	NR	Moderate
Hunninghake 2001[49]	Prospective Results from radiological core of overall IPF diagnosis	91	Serious	N/A	N/A	N/A	NR	41/53 (77%)	26/36 (72%)	67/89 (75%)	NR	Moderate
Flaherty 2003A[32]	Retrospective (radiologists)	73 (UIP) and 23 (NSIP)	Serious	N/A	serious	N/A	NR	37%	100%	NR	NR	Low
Flaherty 2007[34]	Community: Clinicians, radiologists and histopathologist: without MDT	39	Serious	N/A	N/A	N/A	NR	82%- 87%	:64%- 72%	N/A	N/A	Low
Flaherty 2007[34]	Academic: Clinicians, radiologists and histopathologist: without MDT	39	Serious	N/A	N/A	N/A	NR	72%-78%	83%-90%	N/A	N/A	Low

Study ID	Design	No. of patients	Limitation	Inconsistency	Indirectness	Imprecision	Pre-test probability	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Quality
Flaherty 2007[34]	Overall: Clinicians, radiologists and histopathologist: without MDT	39	Serious	N/A	N/A	N/A	NR	76%-81%	78%-84%	NR	NR	Low
Flaherty 2007[34]	Community: Clinicians, radiologists and histopathologist: with MDT	39	Serious	N/A	N/A	N/A	NR	89%- 92%	64%- 81%	NR	NR	Low
Flaherty 2007[34]	Academic: Clinicians, radiologists and histopathologist: with MDT	39	Serious	N/A	N/A	N/A	NR	73%-96%	87%-95%	NR	NR	Low
Flaherty 2007[34]	Overall: Clinicians, radiologists and histopathologist: with MDT	39	Serious	N/A	N/A	N/A	NR	79%-94%	78%-90%	NR	NR	Low
Raghu 1999[103]	Prospective (clinical diagnosis)	59	Serious	N/A	N/A	N/A	NR	62%	97%	95%	73%	Moderate
Raghu 1999[103]	Prospective (radiological diagnosis)	59	Serious	N/A	N/A	N/A	NR	78.5%	90%	88%	82%	Moderate

6.3.5.2 Evidence summary of diagnostic yield studies

Table 14: Summary results of diagnostic yield studies

Study	Number of patients diagnosed with IPF	Number of people diagnosed with 'NOT' IPF	Adverse Events	Mortality
BAL				
Ohshimo 2009[91]	68/74	6/74	NR	NR
Transbronchial biopsy				
Oliveira 2011[92]	11/56	45/56	NR	NR
Surgical lung biopsy				
Coutinho 2008[18]	42/120 IIP	78/120	NR	None
Flaherty 2002[33]	106/168 UIP	61/168	NR	NR
Ishie 2009[51]	14/48	33/48	1/48 (residual pneumothorax after chest drain removal)	NR
Jamaati 2006[54]	50/50 UIP	0/50	NR	NR
Lettieri 2005A[69]	42/83	41/83	7/83 (8.4%) (2 acute MI, 2 nosocomial pneumonia, 1 stroke, 1 pancreatitis, 1 prolonged mechanical ventilation)	4/83 at 30 days 5/83 at 90 days
Lettieri 2005[70]	17/44 UIP (specialists), 22/44 UIP (generalists)	28/44 (specialists), 22/44 (generalists)	NR	NR
Ooi 2005[93]	26/70 UIP	44/70	4/70 (OLB: 0 events, VATS: 1 pneumothorax, 1 haemothorax, 2 urinary retention)	1 (VATS)
Peckham 2004[98]	14/26 UIP	11/26	NR	NR
Rena 1999[106]	14/58	44/58	2/58 (prolonged air leak >5 days)	None

Study	Number of patients diagnosed with IPF	Number of people diagnosed with 'NOT' IPF	Adverse Events	Mortality
Sigurdsson 2009A[116]	23/72 UIP	8/73	12/73 (prolonged air leakage: 9, need for mechanical ventilation: 3, pneumonia: 3, acute exacerbation of respiratory failure: 2, other: 1)	NR
Slodkowska 2000[117]	7/14 UIP	NR	NR	NR
Trahan 2008A[128]	5/15 UIP	10/15	NR	NR
Vansteenkiste 1999[129]	4/24	20/24	11/24 (air leak: 7,bleeding: 1, fever: 3)	3/24
Yamaguchi 2004[131]	12/30 IPF	18/30	3/30 (2 acute respiratory failure, 1 prolonged air leak)	None
MDT				
Thomeer 2008[126]	156/179	23/179	NR	NR
Lynch 2005[72]	181/205	24/205	NR	NR
Sverzellati 2010[122] (probability of IPF diagnosis high /intermediate and low)	15/55	40/55	NR	NR
Flaherty 2003A[32]	27/73 (UIP)	26/73 (NSIP)	NR	NR
Flaherty 2007[34]	13/39	23/39	NR	NR
Spencer 2011[119]	40/67	27/67	NR	NR
Sumikawa 2008[121] (radiologist classification: definite + consistent/suggestive of alternative diagnosis+ unclassified)	69/112 (UIP)	29/112(UIP)	NR	NR
Raghu 1999[103]	29/59	30/59	NR	NR

6.3.5.3 Study quality for BAL, Biopsy and MDT

Table 15: Study quality for studies using QUADAS II

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
BAL							
Ohshimo 2009[91]	Retrospective	CT findings assessed-blinded to BAL and clinical results	BAL-unclear if results interpreted blinded	All patients received BAL and were included in analysis	IPF diagnosed by ATS/ERS criteria	Index test was CT without clinical information	Reference standard used in study matched protocol
Transbronchial Biopsy							
Oliveira 2011[92]	Retrospective	Some patients had already undergone diagnostic biopsy	TBB	All patients had TBB	Suspected ILD	Patients had already undergone biopsy	Reference standard used in study matched protocol
Surgical lung biopsy							
Aalokken 2012[4]	Retrospective	Histological diagnosis, without knowledge of the final diagnosis	MDT consensus, blinded to results of index test	All patients were discussed at MDT	Suspected ILD	Index test in study matched protocol	Reference standard used in study matched protocol
Coutinho 2008[18]	Retrospective	Any of: Clinical, History, PFTs, CXR, CT, BAL, TBB, culture for microbiology	SLB; unclear if results of index test were known	Unclear	Unclear	Clinical assessment, X ray, CT (not clear if everyone had CT)	Reference standard used in study matched protocol
Flaherty 2002[33]	Retrospective	Clinical, PFTs, CT	SLB; unclear if results of index test were known	All patients had SLB	Suspected IIP	Index test in study matched protocol	Reference standard used in study matched protocol
Ishie 2009[51]	Retrospective	VATS	Unclear	All patients had VATS	Suspected DPLD	Index test in study matched protocol	Unclear

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
Jamaati 2006[54]	Retrospective	Clinical, CT	OLB, TBB, VATS	TBB, OLB and VATS used	Suspected IPF	Index test in study matched protocol	Reference standard used in study matched protocol
Lettieri 2005A[69]	Retrospective	Unclear	SLB; unclear if results of index test were known	SLB- OLB and VATS	Suspected ILD	Unclear	Reference standard used in study matched protocol
Lettieri 2005[70]	Retrospective	PFTs, interpreted without results of reference standard	SLB	All patients had SLB	Suspected ILD	Unclear	Reference standard used in study matched protocol
Oliveira 2011[92]	Retrospective	Some patients had already undergone diagnostic biopsy	TBB	All patients had TBB	Suspected ILD	Patients had already undergone biopsy	Reference standard used in study matched protocol
Ooi 2005[93]	Retrospective	Clinical, CT	VATS	All patients had VATS	Suspected ILD	Index test in study matched protocol	Reference standard used in study matched protocol
Peckham 2004[98]	Retrospective	CT, ATS criteria	SLB	15 cases were excluded due to incomplete data	Suspected ILD	Unclear	Reference standard used in study matched protocol
Rena 1999[106]	Prospective	Clinical, PFTS, CT, blood tests, bronchoscopy, BAL	VLTB	All patients had VLTB	ILD of unknown aetiology	Patients also had bronchoscopy and BAL	Reference standard used in study matched protocol
Sigurdsson 2009A[116]	Retrospective	Clinical, PFTs, CT,	VATS/OLB	VATS/OLB	Suspected ILD	Not all patients	Reference

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
		bronchoscopy				had a CT scan prior to biopsy	standard used in study matched protocol
Slodkowska 2000[117]	Retrospective	Clinical symptoms, chest radiographs, CT and lung function tests	Open lung biopsy and CT separately	All patients received a separate histological and CT re-examination. Follow up ranged from 1-4 years	Diagnosed IPF/UIP	Not clear if patients had same baseline tests	Reference standard used in study matched protocol
Trahan 2008A[128]	Retrospective	Clinical , PFTs, CT	SLB, results of index test were not known	SLB	Clinical diagnosis of chronic hypersensitivity disorder	Index test matched clinical question in protocol	Reference standard used in study matched protocol
Vansteenkiste 1999[129]	NR, consecutive patients	Clinical, CT, BAL, TBB	OLB/ VATS-histopathologists blinded to clinical info	OLB/ VATS	ILD, not specified after clinical assessment	Patients also had BAL/ TBB prior to biopsy	Reference standard used in study matched protocol
Yamaguchi 2004[131]	Retrospective, consecutive patients	Clinical, PFTs, CXR, CT	VATLB	VATLB	ILD diagnosed by CXR/CT	Index test matched clinical question in protocol	Reference standard used in study matched protocol
MDT							
Hunninghake 2001[49]	Prospective	Clinical diagnosis, radiological diagnosis,	Pathology diagnosis of IPF	All patients underwent a TBB/SLB and CT	All patients suspected of having IPF	MDT composition is not as described in the protocol, blinding not reported to	Cannot be certain that the reference standard is 100% accurate in

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
						knowledge of reference standard, clinicians did not examine the patients themselves only reviewed available data including CT, no clinical information was provided to the radiologists	diagnosing IPF
Flaherty 2003A[32]	Retrospective	CT diagnosis of IPF	Pathological diagnosis of IPF	Non applicable	People with a histological diagnosis of UIP	No concerns	Cannot be certain that the reference standard is 100% accurate in diagnosing IPF
Flaherty 2007[34]	Retrospective	MDT setting Academic/ community	Level of agreement	Non applicable	Suspected IIP	No concerns	Does a higher level of inter observer agreement mean an accurate diagnosis of IPF
Lynch 2005[72]	Retrospective	Study site diagnosis (less experienced radiologist)	Core radiologist diagnosis (higher level of experience)	Non applicable	Patients already diagnosed with IPF and enrolled into a phase 3	No details provided on the level of experience of the	No details provided on the level of experience of the

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
					pharma trial	study site radiologists	core radiologists No blinding of index test results – aware all patients had been diagnosed with IPF
Spencer 2011[119]	Prospective	Diagnostic accuracy of referring centre in diagnosing 'definite' IPF	MDT consensus in a tertiary centre	Non applicable	'definite' IPF	MDT was aware of results of index test	Biopsy was not available in the majority of cases
Sumikawa 2008[121]	Retrospective	Radiological diagnosis of UIP	Pathological diagnosis of UIP	Non applicable	Confident diagnosis of UIP (by second opinion)	No blinding – radiologist were informed about the pathological and clinical UIP diagnosis	No concerns
Sverzellati 2010[122]	Retrospective	CT diagnosis of IPF	SLB diagnosis of IPF	Non applicable	Patients diagnosed with IPF (typical and atypical) mixed with people who do not have IPF	Based on radiologists experience	Based on set criteria
Thomeer 2008[126]	Retrospective	Diagnostic accuracy of respiratory physicians in diagnosing IPF	CT diagnosis and pathological diagnosis	Non applicable	Patients already diagnosed with IPF by a specialist respiratory physician. Patients had already taken part in a pharma	Not reported	Cannot be certain that the reference standard is 100% accurate in diagnosing IPF

Study	Risk of bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Patient Selection	Index Test	Reference Standard
					trial		
Raghu 1999[103]	Prospective	Clinical diagnosis based on thorough assessment including CT and TBB	SLB diagnosis	Appropriately spaced	Untreated symptomatic patients suspect of ILD	No concerns	No concerns

Abbreviations: ATS = American thoracic society, CT = computed tomography, CXR = chest X ray, DPLD = diffuse parenchymal lung disease, ILD = interstitial lung disease, IPF = idiopathic pulmonary fibrosis, MDT = multidisciplinary team, OLB = open lung biopsy, PFTs = pulmonary function tests, SLB = surgical lung biopsy, TBB = transbronchial biopsy, VATLB = video assisted thoracic lung biopsy, UIP = usual interstitial pneumonia, VATS = video assisted thoracic surgery.

6.4 Economic evidence summary

6.4.1 Literature review

One study was identified that included a relevant comparison of video-assisted thoracic surgery (VATS) to limited thoracotomy as a means of obtaining a biopsy sample for the diagnosis of an ILD (interstitial lung disease) patient [76]. This was selectively excluded on the account of having very serious limitations. It is summarised in Appendix R.

No relevant economic evaluations were identified that assessed the value of a multidisciplinary team consensus in the diagnosis of IPF or how this should best be achieved.

6.4.2 Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided below to aid consideration of cost effectiveness. Further details of the unit costs for the cadre of staff that may be involved in an MDT are presented in Appendix J.

Table 16: Unit cost of interventions in the IPF diagnostic pathway

Item	Cost (Inter quartile range)	Notes
Local chest clinic i.e. secondary care diagnostic work up		
X ray	£29 (£23 to £33)	Direct access plain film (First Consultant Led Outpatient attendance, procedure code: DAPF)
Outpatient appointment	£162 (£136 to £231)	Consultant led, face to face, Outpatient procedure code: 340
Computerised Tomography Scan, one area, no contrast	+£95 (£73 to £106)	Outpatient procedure; HRG code RA08Z. Note this is an unbundled cost so only represents the additional cost of the scan and not associated cost of the consultation etc.
Lung Volume Studies	£187 (£122 to £298)	Outpatient procedure; HRG code DZ45Z
Simple airflow study	£168 (£135 to £195)	Outpatient procedure; HRG code DZ44Z. Note that this procedure is likely to be within the same episode as the lung function study, and would be coded and included under the cost of the lung volume study
Simple Gas Exchange Studies	£146 (£124 to £183)	Outpatient procedure; HRG code DZ40Z Note that this procedure is likely to be within the same episode as the lung function study, and would be coded and included under the cost of the lung volume study
Simple Lung Function Exercise Testing e.g. six minute walk, shuttle walk	£269 (£188 to £263)	Outpatient procedure; HRG code DZ32Z. This intervention may or may not be included in the diagnostic work up and is likely to occur in a separate episode to that where lung volume, airflow and gas exchange is studied
Biopsy		
Bronchoalveolar lavage outpatient	£249 (£118 to £305)	Outpatient procedure; HRG code DZ07A - E49.2
Biopsy using Video-assisted thoracic surgery	£2262 (£368 to £3006)	Inpatient procedure, excess bed days not included; HRG code DZ06Z - E59.3+Y744
Multidisciplinary team involvement throughout the diagnostic pathway		

Item	Cost (Inter quartile range)	Notes
Involvement of 2 local centre consultants and three specialist level consultants, MDT coordinator and specialist ILD nurse.	£227 per patient	As costed using personal social services research unit (PSSRU) staff unit costs, including qualifications and strip-end for audio visual equipment that assumes network can "piggy back" on arrangements already in place for other clinical networks. MDTs assumed to operate within a specialist ILD network, with 6 local level MDTs feeding into a central specialist referral hub, covering a population of 1.5 million. Patients may be reviewed up to 3 times in the specialist hub. Please refer to Appendix J.

Abbreviations: HRG = Health Resource Group
Source: NHS Reference costs 2010-2011[25] PSSRU 2010 [99]

6.4.3 New cost-effectiveness analysis

New analysis was not prioritised for this question as the health benefit and cost associated with the outcome of the diagnostic intervention remains unclear. However, in order to aid consideration of cost effectiveness a detailed costing which estimates the incremental cost of adding MDTs to the diagnostic pathway is presented in Appendix J.

In addition, a simple decision analysis that placed the costing and evidence from the clinical review in an economic framework for decision making is presented in Appendix K. The analysis explores eight different diagnostic strategies, half of which have MDT involvement. The impact of different QALY weights being associated with each diagnostic outcome in relation to the cost of each diagnostic strategy is explored. The results give strength to the argument that biopsy should only be offered to patients that have an unconfident diagnosis based on CT findings alone. The results also suggest that MDT involvement gives value by improving precision in interpretation of diagnostic findings, which in turn improves diagnostic yield by reducing the number of cases where clinicians cannot agree on the diagnosis. Please refer to the table below for a summary of findings and to Appendix K for the full report.

Table 17: Economic evidence profile: Diagnostic decision tree comparing 4 potential diagnostic pathway scenarios.

Study	Applicability	Limitations	Other comments	Total cost per patient	Total effects (with QALY weight[c])	Cost effectiveness	Uncertainty
NCGC economic costing	Directly applicable (a)	Potentially serious limitations (b)	Diagnostic decision tree comparing 4 potential diagnostic pathway scenarios [d], with and without MDT involvement	1: £480 1+MDT: £518 2: £605 2+MDT £1,006 3: £1,293 3+MDT: £1,844 4: £2,118 4+MDT: £3,106	1: 0.0448 1+MDT: 0.0495 2: 0.0464 2+MDT: 0.0521 3: 0.0421 3+MDT: 0.0557 4: 0.0444 4+MDT: 0.0560	Non-dominated strategies were scenarios with MDT. The base case analysis suggests the most likely cost effective option is to have a clinical exam, PFTs, and HRCT with a multidisciplinary discussion at local level (scenario 1 with MDT)	In deterministic analysis, the analysis explored different estimates of diagnostic accuracy as derived from literature of the clinical review. Scenario 3 without MDT and scenario 4 without MDT remained dominated options in all of these sensitivity analyses. It is therefore unlikely these strategies are cost effective. Additionally the analysis shows that staff working in academic settings achieve greater diagnostic success than those working in the community both with and without an MDT. Scenarios with an academic MDT in this analysis dominate scenarios without MDT or community MDT. No analysis suggested that a strategy with biopsy was optimal. If a greater QALY gain could be associated with a correct diagnostic outcome; or alternatively a greater QALY or monetary loss could be associated with an incorrect diagnostic outcome, strategies involving biopsy would become more cost effective. Scenario 4 ranked less optimal than scenario 3 in all analyses when using a threshold of £20,000. Varying the time required to review a patient in a local MDT and specialist MDT to 15 minutes respectively did not change the

Study	Applicability	Limitations	Other comments	Total cost per patient	Total effects (with QALY weight[c])	Cost effectiveness	Uncertainty
							conclusions of the results.

(a) From UK perspective with use of NHS published costs.

(b) Treatment effect from clinical evidence identified by systematic review. Downstream benefit was considered in simplistic fashion by applying a QALY weight to each potential outcome, downstream costs were not considered. A QALY gain or loss was not associated with indeterminate cases. Findings of this analysis may not be reflective of a scenario where there is substantial cost associated with effective treatment of IPF patients. No consideration of cost difference between academic and community settings.

(c) In the base case a QALY weight of 0.08 was given to every correct diagnosis, and a QALY weight of 0.08 to every incorrect diagnosis.

(d) Scenario 1: Clinical examination (including PFTs) and HRCT only; Scenario 2: Clinical examination (including PFTs) and HRCT, with the addition of BAL for any patients with unconfident diagnosis using HRCT; Scenario 3: Clinical examination (including PFTs) and HRCT, with the addition of BAL for any patients with unconfident diagnosis using HRCT findings. Where BAL could not exclude IPF with certainty, these patients would have a biopsy. Only those with unconfident diagnosis referred for biopsy; Scenario 4: Clinical examination (including PFTs) and HRCT, with the addition of BAL for any patients which could not have a confident diagnosis using HRCT findings. With the exception of patients that were diagnosed with an alternative ILD at BAL, all patients have a biopsy to confirm diagnosis of HRCT.

6.4.4 Clinical evidence statements

The following statements are organised by outcome and ordered to list the tests in order from the best to the worst diagnostic accuracy according to that measure.

Sensitivity was highest for histological diagnosis, as reported by Aalokken 2012[4].

Moderate quality evidence showed the sensitivity of histological diagnosis of UIP to be 73% in people with idiopathic interstitial pneumonia. (One study, N=64) Aalokken 2012[4].

Low quality evidence showed the sensitivity of CT to be 71% (51-92%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

Low quality evidence showed the sensitivity of ATS clinical criteria to be 71% (51-92%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

Low quality evidence showed the sensitivity of a clinical diagnosis to be 67% (57-75%) in people with diffuse parenchymal lung disease (one study, N=120) Coutinho 2008[18]

Specificity was highest for a clinical diagnosis, as reported by Coutinho 2008[18].

Low quality evidence showed the sensitivity of a clinical diagnosis to be 90% (85-93%) in people with diffuse parenchymal lung disease (one study, N=120) Coutinho 2008[18].

Low quality evidence showed specificity of American thoracic society (ATS) clinical criteria to be 75% (47-92%) in people with suspected ILD (one study, N=26) Peckham 2004[98]

Moderate quality evidence showed the specificity of histological diagnosis of UIP to be 74% in people with idiopathic interstitial pneumonia (one study, N=64) Aalokken 2012[4].

Low quality evidence showed the specificity of CT to be 67% (39-86%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

PPV was highest for histological diagnosis as shown by Aalokken 2012[4].

Moderate quality evidence showed the PPV of histological diagnosis of UIP to be 83% in people with idiopathic interstitial pneumonia (one study, N=64) Aalokken 2012[4].

Low quality evidence showed the PPV of ATS clinical criteria to be 77% (50-92%) in one study with 26 people with suspected ILD (one study, N=26) Peckham 2004[98]

Low quality evidence showed the PPV of a clinical diagnosis to be 76% (67-84%) in people with diffuse parenchymal lung disease (one study, N=120) Coutinho 2008[18].

Low quality evidence showed the PPV of CT to be 71% (51-92%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

NPV was highest for clinical diagnosis as reported by Coutinho 2008[18].

Low quality evidence showed the PPV of a clinical diagnosis to be 85% (80-89%) in people with diffuse parenchymal lung disease (one study, N=120) Coutinho 2008[18].

Low quality evidence showed NPV of CT to be 76% (39-86%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

Low quality evidence showed the NPV of ATS clinical criteria to be 73% (54-86%) in people with suspected ILD (one study, N=26) Peckham 2004[98].

Moderate quality evidence showed the NPV of histological diagnosis of UIP to be 61% in 64 people with idiopathic interstitial pneumonia (one study, N=64) Aalokken 2012[4].

Yield

In the papers that provided diagnostic yield of IPF/ not IPF, it was not possible to pool results as the diagnoses were not consistently categorised and differed between IPF, UIP and IIP (very low quality evidence).

MDT

Due to the varied nature of reporting of MDT papers, it was not possible to pool results and report on sensitivity or specificity separately as the outcome measures were not consistently reported, however, a narrative summary of each paper is provided below.

Hunninghake et al 2001[49]

Moderate quality evidence showed that using biopsy as the gold standard, clinical and radiological data give a high level of specificity and sensitivity when diagnosed by a chest physician and radiologist with extensive experience in the care of people with ILD. The clinical core made up of 3 chest physicians had a sensitivity of 72%, specificity of 84% and PPV of 87%. The radiology core made up of 4 radiologists had a sensitivity of 77%, specificity of 72% and PPV of 85% (one study, N=91).

Flaherty et al 2003[32]

Low quality evidence showed that patients with histological UIP diagnosis who were diagnosed with IPF by a UIP on biopsy, 37% of CT findings would have strongly disagreed and a further 27% of CT findings would have moderate disagreement. Where biopsy ruled out IPF, CT findings moderately disagreed with 22% of cases. However, in cases where the specialities did not agree, a different trajectory of survival could be observed (one study, N=73).

Flaherty et al 2007[34]

Low quality evidence evaluated the agreement in classification of people with suspected IIP in community and academic settings. They found that a significantly higher level of disagreement exists between physicians in the community setting compared to those in an academic setting. The inter observer agreement (K score) was higher in all clinical groups in the academic setting. K scores of the academic centre; Clinical: 0.71 (± 0.03 SE) Radiological: 0.55 (± 0.08 SE) Pathology: 0.57 (± 0.05 SE). K score of the community centre Clinical: 0.44 (±0.07 SE), Radiological: 0.32 (±0.11 SE), Pathology: 0.41 (±0.13 SE). Ranges for sensitivity and specificity were calculated by using data reported by Flaherty et al[34]. The data shows that sensitivity and specificity in both academic and community settings increases with MDT consensus (see extraction table in Appendix C) (one study, N=39).

Lynch et al 2005[72]

Very low quality evidence showed that using data derived from a prospective multinational trial, CT interpretations of IPF by study site radiologists (using predefined criteria) was confirmed by core radiologists (expert group) in 90% of cases. This indicates that study site radiologists have adequate expertise to diagnose IPF based on CT data when compared with expert opinion (one study, N=315).

Spencer 2011[119]

Low quality evidence showed that in patients diagnosed as having 'definite IPF' by a referring centre, the diagnosis was changed in 27 cases and in 40 cases it was confirmed when assessed by an MDT in a tertiary centre (one study, N=67).

Sumikawa et al 2008[121]

Low quality evidence showed that in patients diagnosed with UIP, radiological diagnosis did not concur with pathological diagnosis in 30% of cases. Radiologists' classification of UIP in patients diagnosed with UIP pathologically was definite UIP in 33/112(34%), consistent with UIP in 36/112 suggestive of alternative diagnosis in 21/112 (21%) and unclassified in 8/112 (8%). The inter-observer agreement of CT diagnosis was consistent with UIP (definite or probable) or suggestive of alternate diagnosis (suggestive of NSIP or indeterminate) was moderate (k 5 0.60) between radiologists (one study, N=112).

Sverzellati et al 2010[122]

Low quality evidence showed that radiological diagnosis alone is not sufficient to correctly diagnose 100% of patients when compared to histopathological diagnosis as the gold standard. The combined observations of IPF probability by 3 radiologists were high in 15/55, intermediate in 6/55 and low in 34/55. The inter-observer agreement between radiologist for first choice diagnosis was moderate: (k = 0.45 (95% CI: 0.32, 0.58)) in people with biopsy proven IPF (one study, N=55).

Thomeer et al 2008[126]

Low quality evidence showed that the diagnosis of IPF proposed by a respiratory specialist was rejected in 12.8% of cases after review of histology and CT by expert committee. The mean level of agreement between 3 different CT reviewers was 0.40 (mean weighted K) and 2 pathology reviewers 0.30 (one study, N=not clearly reported).

Raghu et al 1999[103]

Moderate quality evidence showed that in a cohort of patients suspected of IPF the specificity of diagnosing IPF through clinical assessment or CT features alone is high (97% and 90% respectively) but the sensitivity is low (62% and 78.5%). This shows that the diagnosis can be missed in up to 30% of new-onset IPF cases (one study, N=59).

6.4.5 Economic evidence statements

No published economic evaluations were identified to aid consideration of cost effectiveness.

It is likely that involvement of a multidisciplinary team at each stage of the diagnostic pathway for IPF patients is cost effective when compared to no involvement. This is based on evidence with direct applicability but with potentially serious limitations.

It is likely that with the involvement of a multidisciplinary team at each stage of the diagnostic pathway a diagnosis using clinical and radiological findings alone is more cost effective than a diagnosis using clinical and radiological findings with biopsy. This is based on evidence with direct applicability but with potentially serious limitations.

6.5 Recommendations and link to evidence

Recommendations	1. **Be aware of idiopathic pulmonary fibrosis when assessing a patient with the clinical features listed below and when considering requesting a chest X-ray or referring to a specialist:** • **age over 45 years** • **persistent breathlessness on exertion** • **persistent cough** • **bilateral inspiratory crackles when listening to the chest** • **clubbing of the fingers** • **normal spirometry or impaired spirometry usually with a restrictive pattern but sometimes with an obstructive pattern.**
Relative values of different outcomes	This is a scene setting recommendation and was based on GDG consensus. The initial identification and assessment of possible ILD in primary care was considered a key aspect in the early diagnosis and clinical care pathway of people with IPF. People with suspected IPF can then be referred to secondary care to establish diagnosis and to enable initiation of appropriate clinical management. The GDG considered that including this consensus recommendation would increase awareness in primary care.
Trade-off between clinical benefits and harms	This recommendation was based on GDG consensus. The GDG discussed the importance of improving the initial assessment of people with suspected IPF. The GDG considered that there would be a risk of delays in the diagnosis and initiation of appropriate clinical management and best supportive care in people with IPF if awareness of the signs and symptoms of IPF were not highlighted for healthcare professionals in primary care, which can then be followed up by specific specialist investigations in secondary care. The GDG agreed that this should enable referrals in a more timely fashion. The GDG considered age to be a clinical feature of IPF. They also discussed that the incidence of the disease increases with older age and that presentation with IPF tends to occur in the range of 60-75 years of age. The GDG agreed that cases of IPF below 45 years of age are very rare. The GDG acknowledged that the diagnosis of IPF may be delayed in a significant proportion of patients as the symptoms and signs can be attributed to more common conditions, such as heart failure or COPD. People with IPF may have co-existing COPD. This can result in inappropriate treatments, deprive the patient of appropriate advice and support, and delay the identification of reversible causes of fibrosis The GDG discussed that in addition to the symptoms and signs outlined in the recommendation, patients suspected of IPF may present with episodes of increased sputum and breathlessness and that oxygen saturation may be normal at rest but fall on exertion.
Economic considerations	No published economic evidence was identified to inform this recommendation.

	The GDG discussed the economic implications of the criteria which should be used in primary care for referral for X ray, noting that if the criteria were too broad there would be inappropriate over-referral and cost to the NHS. In determining the criteria, the GDG also considered other conditions for which a chest X ray would be indicated and felt confident that any patient with persistent cough or crackles and persistent breathlessness on exertion would benefit from a chest X ray for timely diagnosis of IPF or other conditions. For people with IPF, an earlier diagnosis would allow for best supportive care and monitoring to be put in place, and potentially may increase the proportion of patients eligible for transplant (as an earlier diagnosis gives opportunity for an earlier referral with associated health benefits). In addition, the GDG considered the potential cost to the NHS of inappropriately treating people with IPF with treatments for other conditions, such as asthma or COPD, whilst IPF remains misdiagnosed. Given the potential health benefits for people with IPF and people with other respiratory conditions of an early diagnosis, and the need for an accurate diagnosis so that only cost effective interventions are offered, the GDG considered that the likely increase in referrals to chest X ray according to the criteria listed was highly likely to be cost effective.
Quality of evidence	This recommendation was based on GDG consensus.
Other considerations	The GDG discussed the importance of the appropriate implementation of the guideline to raise awareness of IPF with GPs.

Recommendations	**2. The consultant respiratory physician or interstitial lung disease specialist nurse should provide accurate and clear information (verbal and written) to people with idiopathic pulmonary fibrosis, and their families and carers with the person's consent. This should include information about investigations, diagnosis and management.**
Relative values of different outcomes	This recommendation was based on GDG consensus. The importance of effective communication between health professionals and people with IPF and caregivers, was identified by the GDG as an important consideration to facilitate good practice when informing patients of diagnostic information.
Trade-off between clinical benefits and harms	This recommendation was based on GDG consensus. The GDG discussed the importance of people with suspected IPF being given appropriate information to allow them to understand the risks and benefits of each intervention in relation to the accuracy of the intervention. As diagnosis is a terminal illness, some people may prefer not to go through an invasive procedure for that level of accuracy given no treatments are available.
Economic considerations	No published economic evidence was identified to inform this recommendation.

Quality of evidence	This recommendation was based on GDG consensus.
Other considerations	The GDG regarded patient communication to be an important consideration for these recommendations. Communication included information at all stages of disease progression for patients and carers regarding: life expectancy; expectations of future symptoms and management; treatment options; and functional ability; as well as provision of wider welfare and lifestyle issues. GDG discussions centred on the importance of clear and tailored patient and carer information according to the patients' individual requirements, whilst acknowledging that requirements will differ throughout the progression of the disease. The expertise of the health professional and healthcare setting in which information is being provided was also considered important, with tertiary specialist care staff and facilities providing increased confidence and reassurance to patients regarding their care.

Recommendations	**3. Assess everyone with suspected idiopathic pulmonary fibrosis by:** • **taking a detailed history, carrying out a clinical examination (see recommendation 1 for clinical features) and performing blood tests to help exclude alternative diagnoses, including lung diseases associated with environmental and occupational exposure, with connective tissue diseases and with drugs, and** • **performing lung function testing (spirometry and gas transfer) and** • **reviewing results of chest X-ray and** • **performing CT of the thorax (including high-resolution images).**
Relative values of different outcomes	The GDG agreed that the critical outcomes to inform decision making were mortality, survival, sensitivity and specificity.. The GDG recognised that sensitivity and specificity would difficult to interpret, because studies choose different interventions for the gold standard test for comparison. The GDG considered routine practice, inter-observer agreement and clinical experience to be important outcomes to inform this recommendation. Outcomes identified from studies included in the diagnostic evidence review where used to inform this recommendation.
Trade-off between clinical benefits and harms	The GDG discussed the trade-off between the value of obtaining an accurate diagnosis based on baseline tests (clinical evaluation, lung function tests and CT) against the accuracy of achieving a confident diagnosis using more invasive procedures, which are associated with adverse events.
Economic considerations	No published economic evaluation was identified to inform this recommendation. The GDG considered the unit cost of baseline diagnostic interventions (clinical evaluation, lung function tests and CT) alongside the findings of the clinical review. It was noted that the baseline investigations are not invasive and do not carry the risk of adverse events or complications, and it is unlikely any downstream costs are associated with the interventions themselves. It was also recognised that other diagnostic interventions, such as BAL and biopsy, may not be appropriate for a proportion of people with suspected IPF due to patient safety or patient preference. The level of diagnostic accuracy was thought to be sufficient to help determine

	whether further diagnostic tests would be required to ascertain a level of confidence in the diagnosis that was desired according to patient preference and to initiate appropriate clinical management, and on this basis the cost of these baseline interventions were thought justified. This assumption was supported by placing the evidence reported by the clinical review into an economic framework as detailed in Appendix J. This analysis shows that ending diagnostic investigation when a confident diagnosis is achieved through clinical and radiological findings is very likely to be cost effective when compared to further diagnostic investigation for these patients. The cost effectiveness of further investigation for patients without a confident diagnosis is less clear and discussed below. The unit cost was sourced from national reference costs and deemed to be reflective of the intervention in this population.
Quality of evidence	This recommendation was based on GDG consensus, as the evidence was of low to very low quality due to the limitations in study design and inconsistency across populations and diagnostic procedures.
Other considerations	It was recognised that baseline lung function tests and CT not only inform a person's diagnosis, but also have an additional benefit to predict the prognosis of people with IPF. Blood tests for precipitins and antibodies should be carried out as these are essential in patients with yet undeclared connective tissue disease associated ILD.

Recommendations	**4. Diagnose idiopathic pulmonary fibrosis only with the consensus of the multidisciplinary team (listed in table 19), based on:** • **the clinical features, lung function and radiological findings (see recommendation 3)** • **pathology when indicated (see recommendation 6).**
Relative values of different outcomes	The GDG agreed that the critical outcomes to inform decision making were mortality, survival, sensitivity and specificity. The GDG agreed that performing a diagnostic procedure such as biopsy may also increase mortality, due to the invasiveness of the procedure. The GDG recognised that sensitivity and specificity would difficult to interpret, because studies choose different interventions for the gold standard test for comparison. Discussion focused on the value of ascertaining a true positive and true negative with confidence, as well as avoidance of false negatives and false positives. The value of ascertaining an unconfirmed diagnosis (i.e. where it was recognised a patient may have UIP or NSIP) was discussed and acknowledged to be a potential benefit as the average disease progression may differ from those people with a classical presentation. The GDG also considered inter-observer agreement to be an important outcome.
Trade-off between clinical benefits and	The GDG came to a consensus that the inclusion of the MDT, compared to no MDT involvement, was likely to result in a greater diagnostic yield and accuracy in

harms	diagnosis of people with IPF and other ILD patients who may have missed treatment opportunities (e.g. involvement in a clinical trial) if incorrectly diagnosed with IPF. Unlike other diagnostic interventions (i.e. biopsy), the involvement of an MDT does not carry risk of further complications or adverse events. It was considered that an MDT may decrease the potential health risk to patients if a confident diagnosis could be achieved without the need for tissue sampling (i.e. BAL, TBB and surgical lung biopsy). Diagnostic accuracy and precision is increased when there is discussion between clinicians, radiologists and histopathologists. The benefit of reduced anxiety for the patient in knowing a diagnosis was noted to be very important.
Economic considerations	No published economic evidence was identified to inform this recommendation. The GDG considered how the incremental health benefit of MDT involvement could be achieved. An important driver of cost effectiveness of a diagnostic strategy with MDT involvement is the reduced need for further more expensive and invasive procedures (i.e. surgical lung biopsy) due to the increased certainty of diagnosis achieved with specialist input at an early stage of the diagnostic pathway. The GDG agreed that the incremental cost of MDT involvement in the diagnostic pathway for an ILD patient is comparable to other diagnostic interventions. This was based on a costing where every suspected IPF diagnosis was confirmed at a specialist MDT. The GDG considered the clinical evidence presented in an economic decision analytic framework. The GDG discussed the implications of the uncertainty surrounding the downstream benefits and costs associated with different diagnostic outcomes, including cases where agreement between MDT members could not be ascertained. The analysis showed that diagnostic scenarios without MDT involvement were likely to be dominated (i.e. less effective and more costly) by a diagnostic scenario with MDT. The need for confidence and certainty in a diagnosis was discussed. It was noted that a patient's quality of life may be decreased through increased anxiety or potentially depression if their diagnosis remained uncertain or if they had little confidence that the diagnosis achieved was correct. The GDG discussed that such patients may continue to seek a more confident diagnosis with further GP contacts and secondary care consultations, which would be at a cost to the NHS. The potential for the MDT to increase the number of people with a diagnosis which was agreed across the specialities involved in the diagnostic pathway could provide further benefit that was not captured in the analysis presented. As such, the GDG considered the results of the sensitivity analysis presented where the potential QALY gain was higher than that of the base case, and interpreted the results with care given that no downstream cost was associated with an uncertain diagnosis. The opportunity cost of staff time for an MDT was based on nationally available estimates from an NHS perspective. This included the cost of training to take into account the need for specialist staff. The assumptions made in the costing and subsequent analysis presented in Appendix J were agreed and their implications discussed. Given the number of assumptions made and the quality of the clinical evidence used, it was noted the results of the analysis needed to be interpreted with caution. It was recognised that the cost effectiveness of the addition of MDT involvement

	may depend on the cost effectiveness of the management plan that follows, as the MDT is likely to improve accuracy and the number of correct diagnoses. The cost effectiveness of MDT involvement could increase if emerging IPF management plans are costly (as fewer false positive cases will have inappropriate costly treatment) and/or bring substantial health benefit for people with IPF (as more true positives will be able to benefit from this treatment).
Quality of evidence	Nine studies investigating the role of the MDT in diagnosing people with IPF informed this recommendation. Only one study provided data on diagnostic accuracy. All studies reported inter-observer agreement between health professionals of various specialities and expertise, from various locations. The quality of the evidence ranged from low to moderate quality due to limitations in study design.
Other considerations	Raising the index of suspicion of possible ILD in primary care and timely referral to a respiratory specialist was considered an important factor for diagnosing people with IPF at an earlier stage in their disease.

Recommendations

5. At each stage of the diagnostic care pathway the multidisciplinary team should consist of a minimum of the healthcare professionals listed in table 19, all of whom should have expertise in interstitial lung disease.

Table 18: Minimum composition of multidisciplinary team involved in diagnosing idiopathic pulmonary fibrosis

Stage of diagnostic care pathway	Multidisciplinary team composition (all healthcare professionals should have expertise in interstitial lung disease)[e]
After clinical evaluation, baseline lung function and CT	• Consultant respiratory physician • Consultant radiologist • Interstitial lung disease specialist nurse • Multidisciplinary team coordinator
When considering performing bronchoalveolar lavage, and/or transbronchial biopsy or surgical lung biopsy Only some patients will have bronchoalveolar lavage or transbronchial biopsy but they may be being considered	• Consultant respiratory physician • Consultant radiologist • Consultant histopathologist • Thoracic surgeon as appropriate • Interstitial lung disease specialist nurse • Multidisciplinary team coordinator

[e] See chapter 6.5 (Multidisciplinary Team) for more information on the expertise of the multidisciplinary team.

	for surgical lung biopsy	
	When considering results of bronchoalveolar lavage, transbronchial biopsy or surgical lung biopsy	• **Consultant respiratory physician** • **Consultant radiologist** • **Consultant histopathologist** • **Interstitial lung disease specialist nurse** • **Multidisciplinary team coordinator**
Relative values of different outcomes	The GDG agreed that the critical outcomes to inform decision making were mortality, survival, sensitivity and specificity. The GDG agreed that performing a diagnostic procedure such as biopsy may also increase mortality, due to the invasiveness of the procedure. The GDG recognised that sensitivity and specificity would be difficult to interpret, because studies choose different interventions for the gold standard test for comparison. Discussions focused on the value of ascertaining a true positive and true negative with confidence, as well as avoidance of false negatives and false positives. The value of ascertaining an unconfirmed diagnosis (i.e. where it was recognised a patient may have UIP or NSIP) was discussed and acknowledged to be a potential benefit as the average disease progression may differ from those people with a classical presentation. The GDG considered inter-observer agreement to be an important outcome. The relevant expertise of health professionals was also discussed.	
Trade-off between clinical benefits and harms	The GDG acknowledged the increased risk of adverse events associated with biopsy (BAL, TBB and surgical lung biopsy). No direct harms to a person suspected of IPF was attributed to MDT discussions, but the GDG did recognise that harms associated with an incorrect diagnosis made by an MDT and subsequent inappropriate management plan may result in reduced quality of life. However, the GDG felt that the diagnostic accuracy and precision is dependent on the expertise of the clinicians, radiologists and histopathologists. The GDG acknowledged that typically a thoracic surgeon would be present at an MDT to aid surgical planning and further reduce the risk of an incorrect diagnosis being made.	
Economic considerations	No published economic evidence was identified to inform this recommendation. The GDG considered the addition of an MDT for diagnosing IPF in the context of an ILD network. The discussion was informed by a detailed costing which considered the composition, role and setting of an MDT at each of the different time points in the diagnostic care pathway of a patient with IPF. The following assumptions for the purpose of the costing were agreed; the population served by each MDT; the composition of the MDT and the level of expertise required; the resources required; and opportunity cost of staffing and the number of diagnostic reviews. The level and expertise of the thoracic surgeon was not costed into the analysis, as the GDG considered that typically little of the thoracic surgeon's time would be spent discussing a surgical plan for people with IPF. The GDG discussed the evidence regarding the setting of an MDT and concluded that the requirement for MDT members to have specialist expertise in ILD could be a means to further improve the	

	accuracy and yield of MDT involvement without significantly increasing cost from an NHS perspective. The GDG expected the NHS cost of staff with a specialist expertise to be similar to the cost of staff of the same cadre without a specialist expertise, however as the recommendation indicates a potential change in skill mix qualification costs were incorporated into all cost calculations. The GDG discussed that an example of a chest physician with expertise in ILD may be someone who runs a service seeing at least 500 ILD patients per year or has done an MD/PHD in ILD or a clinical fellowship in ILD for at least 6 months. The opportunity cost of staff time for an MDT was based on nationally available estimates from an NHS perspective. This included the cost of qualification to take into account the need for specialist staff. The assumptions made in the costing and subsequent analysis presented in Appendix I were agreed and their implications discussed. Given the number of assumptions made and the quality of the clinical evidence used, it was noted the results of the analysis needed to be interpreted with caution. The benefit of a confident diagnosis agreed by the different clinical specialists involved in the care pathway was discussed as a potential driver of cost effectiveness of an MDT. If confidence in the diagnosis is increased by staff members having expertise in ILD, this could also be an important consideration for cost effectiveness. The incremental cost of the involvement of an MDT in the diagnostic pathway of an ILD patient is likely to be comparable or lower to other diagnostic tests. However, the actual incremental cost of the involvement of an MDT in the diagnostic pathway is likely to be influenced by use of clinical network arrangements already in place, local need and commissioning arrangements. The GDG acknowledged that local expertise would influence the number of cases being referred to, and time requirement of, the specialist MDT. Therefore, the most cost effective arrangement is potentially highly influenced by local factors. The GDG considered the additional benefit of an ILD MDT network in management of patients to be an important consideration. It was noted that for an MDT to fulfil this additional role, the MDT composition would need to also include other cadres of health professionals, such as pharmacists, who were not considered in the costing for the diagnostic element of the MDT.
Quality of evidence	Nine studies investigating the role of the MDT in diagnosing people with IPF informed this recommendation and ranged from low to moderate quality due to limitations in study design and inconsistency across populations and diagnostic procedures. These studies did not provide data on diagnostic accuracy or yield, but reported inter-observer agreement between health professionals of various specialities and expertise, from various locations.
Other considerations	The GDG discussed examples of the level of expertise in terms of the composition an MDT: an ILD specialist nurse or respiratory nurse with expertise in ILD, would be ideally someone involved in a service seeing at least 500 ILD patients per year or has completed specialist training in ILD for at least 6 months; a specialist chest physician with expertise in ILD who may be someone who runs a service seeing at least 500 ILD patients per year or has done an MD/PHD in ILD or a clinical fellowship in ILD for at least 6 months; a radiologist who for example may be someone who interprets at least 750 thoracic CT studies, attends at least 50% of the local ILD multidisciplinary meetings, provides a substantial contribution to ILD regional service, and has undertaken a fellowship in thoracic imaging including ILD for at least 6 months and a

	thoracic surgeon. An ILD specialist nurse would likely work autonomously, but be required at MDTs to effectively capture and assess care needs of people with ILD and their families from referral though to treatment and management, including providing relevant support and information. A greater proportion of time (estimated 1.5 hours a week) from a thoracic transplant surgeon would also extend to a lung transplant assessment meeting (compared to diagnosing a patient with IPF), where they would participate in the decision to accept patients onto the waiting list and give advice regarding any surgical technical issues as well as whether the patient is listed for bilateral or single lung transplantation as part of the multidisciplinary team. The early identification of possible ILD in primary care and referral was considered an important aspect in diagnosing a patient with IPF.

Recommendations	**6. If the multidisciplinary team cannot make a confident diagnosis from clinical features, lung function and radiological findings, consider:** **• bronchoalveolar lavage or transbronchial biopsy and/or** **• surgical lung biopsy, with the agreement of the thoracic surgeon.** **7. Discuss with the person who may have idiopathic pulmonary fibrosis:** **• the potential benefits of having a confident diagnosis compared with the uncertainty of not having a confident diagnosis and** **• the increased likelihood of obtaining a confident diagnosis with surgical biopsy compared with bronchoalveolar lavage or transbronchial biopsy and** **• the increased risks of surgical biopsy compared with bronchoalveolar lavage or transbronchial biopsy.** **8. When considering bronchoalveolar lavage, transbronchial biopsy or surgical lung biopsy take into account:** **• the likely differential diagnoses and** **• the person's clinical condition, including any comorbidities.**
Relative values of different outcomes	The GDG agreed that the critical outcomes to inform decision making were mortality, survival, sensitivity and specificity. The GDG agreed that performing a diagnostic procedure such as biopsy may also increase mortality, due to the invasiveness of the procedure. Sensitivity, specificity and adverse events were considered to be critical outcomes to determine the added value of conducting a bronchoalveolar lavage (BAL) and/or bronchoscopic/transbronchial biopsy, or surgical biopsy when baseline clinical history, PFTs and CT have been performed. The GDG recognised that sensitivity and specificity would be difficult to interpret, because studies choose different interventions for the gold standard test for comparison. In the absence of these outcomes the GDG considered length of hospital stay, routine practice and clinical experience to be important.
Trade-off between clinical benefits and harms	The GDG considered the value of obtaining an accurate diagnosis based on baseline tests (clinical evaluation, lung function tests and CT) against the accuracy of achieving a confident diagnosis using more invasive procedures, which are

	associated with adverse events. The GDG considered the value of excluding diagnoses other than IPF using bronchoalveolar lavage against adverse outcomes and the clinical limitations associated with bronchoscopic/transbronchial biopsy. BAL may have low sensitivity in confirming the diagnosis of IPF but may be helpful in pointing towards other diagnoses such as hypersensitivity pneumonitis or sarcoidosis if a significant lymphocytosis is present. The GDG considered BAL an additional investigation which can be considered on an individual basis particularly as it is less invasive than surgical lung biopsy. It was acknowledged that bronchoscopic/transbronchial biopsy would not be appropriate for a proportion of patients suspected with IPF due to safety concerns or patient preference, and for these patients BAL may be an appropriate alternative in achieving more confidence in a diagnosis. The GDG considered the incremental benefit of conducting a surgical biopsy against biopsy sampling error and adverse events such as risk of infection, haemoptysis and pneumothorax. In a proportion of the people with possible IPF, the risks of performing a surgical lung biopsy outweigh the benefits of confirming diagnosis.
Economic considerations	No economic evidence of sufficient quality and applicability was identified to inform this recommendation. The GDG considered the unit cost and value of performing a BAL against the risks and unit costs associated with conducting a bronchoscopic/ transbronchial biopsy, along with potential length of hospital stay. The potential health risk and cost of adverse events associated with the procedures were considered. The GDG also acknowledged that surgical biopsy was the most expensive of the diagnostic interventions for IPF and had the greatest potential to generate downstream cost and health risk. VATS was considered by the GDG to result in fewer complications and lower morbidity than open surgical lung biopsy and for this reason is likely to be less costly from an NHS perspective. The economic benefit of BAL as a means to reduce the number of ILD patients being referred on to biopsy was discussed in relation to the confidence of radiological and clinical findings. Where there is a confident diagnosis of IPF from radiological and clinical findings, it is less likely BAL would be a cost effective strategy as it is a specific rather than sensitive test, that is to say it is a potentially useful investigation for identifying people with diagnoses other than IPF, such as hypersensitivity pneumonitis and sarcoidosis. However, the inflammatory cell counts in BAL from people with IPF are relatively non-specific so a surgical biopsy would still be required to confirm the diagnosis. The additional cost of undertaking BAL for every patient in this group would outweigh any diagnostic benefit and the majority of patients in this group should be referred directly for biopsy when appropriate. It was recognised that in cases where CT findings were less characteristic of IPF, there would be a greater likelihood of conditions other than IPF. In such cases, BAL may be a useful intervention in the diagnostic pathway to identify people with other diagnoses. Additionally, if BAL can successfully exclude people without IPF, then the prevalence of IPF in the group of patients referred for biopsy will rise and the positive predictive power of biopsy would be improved. The potential downstream benefit (and cost) of a confident diagnosis was discussed. The GDG acknowledged that the treatment pathways for people with IPF are still emerging and uncertain, and therefore the potential health benefit associated with a correct diagnosis is also uncertain. It was acknowledged that on a patient level, the

	utility associated with a more certain diagnosis and prognosis will differ on a case by case basis. As such clinical qualitative judgement should be used in assessing whether the benefit of having a more confident diagnosis offsets the higher costs and health risk of a more invasive procedure. Increased accuracy of an intervention through MDT discussion at an earlier stage of the diagnostic pathway reduces the incremental benefit of offering all patients a biopsy at a later stage of the pathway. In the analysis presented to the GDG, a diagnostic strategy with biopsy never presented as optimal in terms of cost effectiveness using a threshold of £20,000 per Quality Adjusted Life Year. However, the GDG noted the limitations of the analysis, including the uncertainty surrounding the health benefit gained and the potential of reduced downstream cost through an accurate diagnosis (which was not incorporated). With increased health benefit (through emerging management plans for both IPF and people without IPF) and consideration of the cost in correcting an inaccurate diagnosis, it was considered that a scenario with biopsy could be a cost effective means to improve confidence in a diagnosis in a subgroup of patients where this was deemed appropriate (i.e. the patient was fit for biopsy, did not have a confident diagnosis, the patient preferences, risks and benefits had been taken into account). The unit cost was sourced from national reference costs and deemed to be reflective of the intervention in this population. The assumptions made in the costing and subsequent analysis presented in Appendix J were agreed and their implications discussed. Given the number of assumptions made and the quality of the clinical evidence used, it was noted the results of the analysis needed to be interpreted with caution. Overall, the GDG considered that they were unable to make any firm conclusions regarding the cost effectiveness of biopsy, and came to a consensus that it should only be conducted when appropriate, which in part would rely on clinical expert judgement regarding its added value to the confidence and accuracy in the diagnosis. There was consensus that a thoracic surgeon should be involved in MDT discussions regarding the appropriateness of a lung biopsy given the potential health risks and benefits involved. It was recognised that a thoracic surgeon's time had not been included in the MDT costing as presented. It was acknowledged that the thoracic surgeon was unlikely to be required throughout the MDT and the time commitment required at the MDT for diagnostic biopsy would not be substantial. It is likely the additional opportunity cost of their time would be offset by the benefit realised by the appropriate prevention of biopsy realised by expertise consideration of the costs and risks of biopsy, and the decision to utilise VATS over open surgical lung biopsy where possible.
Quality of evidence	The evidence for BAL consisted of one retrospective study with 74 patients diagnosed with IPF. All patients in the study received BAL, regardless of the initial results from CT findings and the results showed that 8% of patients diagnosed with IPF on CT had an alternative diagnosis to IPF on BAL. The GDG acknowledged that the small sample size, expertise of the interpreters and lack of established cut-off for BAL lymphocytes were limitations for this study and was of very low quality. The GDG discussed the evidence from 1 study which investigated, transbronchial biopsy and 16 studies which investigated surgical lung biopsies for the diagnosis of people with IPF. Two studies provided enough data to calculate diagnostic accuracy (sensitivity, specificity, positive predictive and negative predictive values). The other

	studies presented diagnostic yield figures. The GDG acknowledged the very low quality of these studies due to the limitations in study designs. These studies did not always clearly report the index tests and reference standards. The terminology reported by studies on the final diagnosis was a serious limitation, as this varied between IPF, interstitial fibrosis, UIP and IIP in these studies.
Other considerations	Patient preferences for diagnostic interventions and quality of life were also considered important factors by the GDG in formulating this recommendation. It was recognised that a patient may prefer to trade the benefit of having a confident diagnosis against the risk of further tests associated with adverse events. There was consensus that the involvement of an ILD specialist nurse could aid patient level decision making and MDT knowledge of patient preference in this regard. The GDG also considered the age range of the populations included in the studies to have important implications as a diagnostic factor, as well as for prognosis. Research recommendation The GDG agreed that the lack of evidence and very low quality evidence for BAL and surgical biopsy justified making a research recommendation to question the value of bronchoalveolar lavage in patients in whom IPF is considered the most likely diagnosis when clinical and/or CT findings are insufficient to attain a confident diagnosis. For further information on research recommendations see Appendix P.

Recommendations	**9. If a confident diagnosis cannot be made continue to review the person under specialist care.**
Relative values of different outcomes	This recommendation was based on GDG consensus.
Trade-off between clinical benefits and harms	This recommendation was based on GDG consensus. The GDG discussed the importance of having a recommendation to identify what should be done if a healthcare professional is unable to make a diagnosis once all the necessary investigations have been completed. Through clinical experience, the GDG estimated that approximately one quarter to a half of all people suspected of IPF may remain undiagnosed. The GDG discussed that in practice and in the best interest of the patients, having expert review until their disease phenotype, particularly progressive versus non progressive, becomes apparent seems best course of action.
Economic considerations	No published economic evidence was identified to inform this recommendation.
Quality of evidence	This recommendation was based on GDG consensus.
Other considerations	The GDG discussed the importance of the appropriate implementation of the guideline to raise awareness of IPF with GPs.

7 Prognosis

7.1 Review introduction

Studies have consistently reported that the median survival of patients diagnosed with IPF is approximately 3 years. However, it is also recognised that disease progression amongst individual patients is highly variable; in some the disease progresses rapidly, whilst others exhibit very little change over many years. In part, this spectrum of disease progression may be explained by the way in which IPF is defined. There are a number of fibrotic lung conditions that share the clinical features of IPF, yet are pathologically distinct and have different, often better, prognoses. Securing a confident diagnosis of IPF, through multidisciplinary integration of clinical, radiological and, where available histological data, helps clinicians and patients to better anticipate the likely prognosis. However, even when IPF is confidently diagnosed in this way, there remains marked variability in disease progression.

The uncertainty in estimating how quickly the disease will progress is troubling for patients and their carers. As a result, several studies have attempted to describe disease characteristics in IPF that can be used to better predict survival. Ideally, these characteristics should be easy to measure at the time of diagnosis and accurately predict the rate of progression in an individual patient. However, baseline measurements alone may not be sufficiently powerful to estimate risk of progression; hence some studies have investigated the utility of changes in variables such as lung function, exercise tests and CT scanning as predictors of disease progression. In this context, the value of repeating a clinical investigation in order to determine prognosis must be balanced against any potential risks of performing the test and its cost.

7.2 Clinical questions and review methodology

The following clinical questions were included in this chapter.

For full details see review protocols in Appendix C.

7.2.1 Do serial pulmonary function tests (PFTs) (resting spirometric, gas transfer measurement and oxygen saturation) predict prognosis of IPF?

Table 19: PICO characteristics for PFTs predicting prognosis of IPF

Population:	Adults with IPF
Prognostic Factors:	• FVC <5% change> • TLCO or DLCO <15% change> • Oxygen saturation <92%> (Risk factors - age, sex, smoking status, baseline lung function, previous hospitalisations)
Outcomes:	Critical outcomes • Mortality or survival (time to event) Other outcomes • Progression free survival • Acute exacerbation (time to event) • Respiratory hospitalisations (surrogate outcome for acute exacerbation) • Eligibility for lung transplant

Study design:	Cohort studies

7.2.2 Does baseline sub-maximal exercise testing predict prognosis of IPF?

Table 20: PICO characteristics for sub-maximal exercise testing

Population:	Adults with IPF
Prognostic Factors:	Sub-maximal exercise testing (threshold unknown) (Risk factors - ge, sex, smoking status, baseline lung function)
Outcomes:	Critical outcomes • Mortality or survival (time to event) Other outcomes • Progression free survival • Acute exacerbation (time to event) • Respiratory hospitalisations (surrogate outcome for acute exacerbation) • Eligibility for lung transplant
Study design:	Cohort studies

7.2.3 Does baseline echocardiography predict prognosis of IPF?

Table 21: PICO characteristics for echocardiography

Population:	Adults with IPF
Prognostic Factors:	Pulmonary arterial systolic pressure (threshold unknown) (Risk factors - age, sex, smoking status, baseline lung function)
Outcomes:	Critical outcomes • Mortality or survival (time to event) Other outcomes • Progression free survival • Acute exacerbation (time to event) • Respiratory hospitalisations (surrogate outcome for acute exacerbation) • Eligibility for lung transplant
Study design:	Cohort studies

7.2.4 Do baseline CT scores predict prognosis of IPF?

Table 22: PICO characteristics for CT scores

Population:	Adults with IPF
Prognostic Factors:	CT features/patterns (Risk factors - age, sex, smoking status, baseline lung function)
Outcomes:	Critical outcomes • Mortality or survival (time to event) Other outcomes • Progression free survival

	• Acute exacerbation (time to event) • Respiratory hospitalisations (surrogate outcome for acute exacerbation) • Eligibility for lung transplant
Study design:	Cohort studies

The objectives of the clinical questions were to determine whether:

- resting spirometric, gas transfer measurements and oxygen saturation predict prognosis of IPF
- sub-maximal exercise testing predicts prognosis of IPF (the GDG agreed to limit sub- maximal exercise testing to the 6 minute walk distance (6MWD) as this is the most common submaximal exercise testing used routinely in the U.K.)
- echocardiography predicts prognosis of IPF
- CT predicts prognosis of IPF.

The literature was searched for all years for studies assessing whether PFTs, 6MWD, echocardiography and CT predict prognosis of IPF.

Inclusion criteria were as follows:

- any duration of follow-up
- any sample size
- population ≥18 years
- study design: diagnostic cohorts, (prospective and retrospective)
- studies published post 1994 (studies that span inclusion of subjects pre 1994 are also included).

Note: A modified version of GRADE has been used and a narrative summary provided in this evidence review. The statistics used for this prognostic review differ from those used in intervention reviews.

7.3 Clinical evidence

7.3.1 Summary of included studies

Eighteen studies in total were identified; some reported on more than 1 prognostic factor and therefore were included in more than one section of this evidence review. Of these, 16 studies reported on PFTs, 2 studies on six minute walk test (6MWT), 1 study on echocardiography and 4 studies reported on CT.

Authors of 2 studies[26,28] were contacted to provide extra analysis of data upon advice of the GDG. This unpublished data has been used by the GDG in their decision-making and referred to as DuBois 2013 throughout the guideline[2].

Survival, including mortality, was identified as an outcome in all studies; however no papers reported eligibility for lung transplant. One study on PFTs had progression free survival as an outcome. One study[79] reported on the effect of baseline and 6 month DLCO on acute exacerbations.

The protocol states that multivariable analysis will be used. Therefore, studies only reporting univariable analysis were excluded. Where studies have reported both univariable and multivariable analysis, only the results of the multivariable analysis have been reported in evidence tables and included in the final analysis.

The minimum set of confounding factors that were identified by the GDG consisted of: age, sex, smoking status, previous hospitalisations and, in the PFT section only, baseline PFTS.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F study and selection flow chart in Appendix Q and exclusion list in Appendix R.

Table 23: Summary of all studies included in the review

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
Best 2008[8]	167	IPF	USA Hospital	Median 1.5 years	Baseline CT	Mortality prediction	Confounding factors adjusted for unclear.
Caminati 2009[13]	44	IPF	Italy	1 year	Baseline and 12 month change in resting room air arterial oxygen saturation FVC & DLCO as continuous variables 6MWD, per unit/ continuous variable	Survival	35 patients received drug therapy during the study period. Adjusted for co-variables, which were clinically and statistically significant; these were age and sex only.
DuBois2011A[26]	1099	IPF	UK	1 year	24 week change in percent –predicted FVC </= - 10%, - 5% to – 9.9%, > - 5% Change in percent-predicted FVC </= 50%, 51%-65%, 66%-79%, >/=80%	1 year risk of death	Confounding factors adjusted for: age, oxygen use, surgical lung biopsy, history of respiratory hospitalisation, drug treatment, physiologic % predicted FVC, 24 week change in % predicted FVC, % predicted DLCO, 24 week change in % predicted DLCO, dyspnoea and HRQL UCSD SOBQ and 24 week change in UCSD SOBQ.
DuBois 2011B[28]	1156	IPF	UK Unclear	1 year to 72 weeks	24 week absolute change in percent –predicted FVC </= - 10%, - 5% to – 9.9%, > - 5%	1 year risk of death	Patients receiving active drug treatment during were adjusted for in the analysis, but study did not

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
					Change in percent-predicted FVC </= 50%, 51%-65%, 66%-79%, >/=80%		report adjusting of other confounders as identified in the protocol.
DuBois2013[27]	748	IPF	UK unclear	1 year	Baseline and serial FVC Baseline and serial 6MWD	1 year IPF related mortality All-cause mortality	Adjusted for age, respiratory hospitalisations, PFTs, 6MWD.
Hamada 2007[40]	78	IPF	Japan Secondary care	Unclear	Baseline DLCO <40%, - dichotomous variable	5 year survival risk	None of the patients were receiving pharmacological interventions or immunosuppressants. Confounding factors adjusted for: age, gender, mean pulmonary arterial pressure, $Pa0_2$, $P0_2$ in mixed venous blood, FVC % predicted, DLCO% predicted and cardiac index. Numbers of patients included in the analysis unclear.
Hallstrand 2005[39]	28	IPF transpl-ant centre	USA	Median (range) 5.4 years (4.3-6.2)	Baseline resting room air arterial oxygen saturation Walk distance 30-metre	Units to mortality	Survival time was measured in days from enrolment until death or censoring. Patients were censored at the end of the

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
					units to mortality, continuous variable		follow-up period or if they underwent lung transplantation. The multivariable model included age, sex, FVC % predicted, time from the onset of symptoms and supplemental oxygen administration during the test as confounding factors. Patient population was taken from a transplant centre; this is a potential bias.
Jeon 2006[57]	88	Patholo-gically confirmed UIP and IPF	South Korea Hospital	Unclear (>1 year)	Baseline FVC and DLCO as continuous variables	Mortality prediction	Adjusted for age, sex, severity of dyspnoea, FVC and DLCO and treatment, multivariable survival analysis.
Kurashima 2010[67]	439	CT diagnosed UIP, with or without emphysema	Japan Hospital	0	Baseline FVC and DLCO as continuous variables	Risk of death	Confounding factors adjusted for in multivariable survival analysis not clearly reported. Unclear if patients receiving treatment were adjusted for in analysis.
Lynch 2005[72]	315	Mild to moderate IPF	Multinatio nal	Unclear	CT consistent/ not consistent with IPF	Survival	Patients were enrolled in a trial of Interferon. Confounding factors

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
			Academic and community centres		DLCO		adjusted for included: overall disease extent score on CT, reticulation pattern score, honeycomb pattern score, predominant pattern reticulation, % predicted DLCO, A-a gradient and current O_2 use.
Manali 2008[73]	25	IPF	Greece Respiratory outpatient clinic	0	Baseline FVC as continuous variables	Mortality	Confounding factors adjusted for in multivariable survival analysis not reported.
Mejia 2009[74]	110	IPF (ATS/ERS 2000 criteria, with or without emphysema	Mexico Institute of Respiratory diseases	Unclear	Estimated systolic pulmonary artery pressure FVC <50% Predicted	Mortality Survival	Some patients had co-existing emphysema. Confounding factors adjusted for in multivariable survival analysis: male gender, emphysema, CT and fibrotic score.
Mogulkoc 2001A[75]	115	Mild to moderate IPF Age <65 years	UK Research centre	Unclear	CT fibrosis score CT ground glass score DLCO % predicted	Mortality	Lung transplantation patients. All patients had been treated with corticosteroids and various chemotherapeutic regimens before and after referral to the centre.

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
							Confounders adjusted for: FEV1, FVC, TLC, DLCO, KCO and CT ground glass appearance.
Mura 2012[79]	70	Newly diagnosed IPF	Italy Research centre	3 years	DLCO	Survival Acute exacerbation	Confounders adjusted for: BMI, MRC dyspnoea score, 6MWD % predicted, desaturation@ 6MWD, PaO_2, FVC % predicted, DLCO % predicted, CPI, CT fibrosis score, BAL cell count.
Richeldi 2012A[108]	142	IPF	USA	12 months	Decline in % predicted FVC at -5%, -10% and -15%	Mortality	Relative and absolute change data. Confounders adjusted for: gender, baseline age, O_2 use, FVC and DLCO.
Schmidt 2011[111]	N=211 (6month change) N=144 (12month change)	IPF	USA University Hospital	15 months	Decline in percent-predicted FVC at -5%, -10%, -15% & -20% Decline in percent-predicted DLCO at -10%, -15%, -20% & -25%	Mortality risk	Adjusted for age at diagnosis, sex and smoking status in multivariable survival analysis. Study did not evaluate the potential impact of treatment on outcome.
Sumikawa 2008[121]	98	IPF on biopsy and clinical findings	Japan	79 months (mean) 63 months (median)	CT findings.	Survival	Confounding factors adjusted for were: each one of the following CT findings: presence of

Reference	Number of patients	Patient group	Location	Follow-up period	Prognostic factor	Outcomes	Notes
							ground-glass attenuation; airspace consolidation; nodules; interlobular septal thickening; thickening of bronchovascular bundles; intralobular reticular opacities; irregular interlobular septal thickening; non-septal linear or plate-like opacities; presence of honeycombing, cysts, emphysema, architectural distortion, or traction bronchiectasis; fibrosis score; the extent of disease close to the hilum; and upper, lower, peripheral, dependent, peribronchovascular, and asymmetric predominant distribution.
Zappala 2010[133]	84	IPF	UK Secondary care	6 months +/- 2	Serial PFT trends at 6(±2)months expressed as percentages of baseline values	Mortality Progression free survival	PFT trends analysed using proportional hazards analysis and multivariable analysis adjusting for age, sex, smoking status and baseline disease severity.

7.4 Study Quality

For all prognostic interventions, quality was assessed using a checklist. Domains that were assessed for quality included: the population sample used, loss to follow-up, measurement of the prognostic factor, measurement of outcomes, accounting for confounders and the statistical analysis used.

7.4.1 Summary of study quality for all studies included in the review

The studies were all of moderate to low quality. In several cases loss to follow-up was unclear and in some cases the method of assessing the prognostic factor was unclear.

Table 24: Study quality checklist of all studies included in the review

Reference	Representative population sample	Loss to follow up described	Prognostic factor measured appropriately	Outcomes adequately measured	Confounders accounted for	Appropriate statistical analysis	Quality
Best 2008 [8]	Yes	Unclear (a)	Yes	Yes	Unclear (b)	Yes	Low
Caminati 2009 [13]	Yes	Unclear (a)	Yes	Yes	Yes (d)	Yes	Low
DuBois2011A [26]	Yes	Unclear (a)	Yes	Yes	Yes	Yes	Moder-ate
DuBois2011B [28]	Yes	Yes	Unclear (c)	Yes	Yes (d)	Yes	Moder-ate
DuBois2013[27]	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder-ate
Hallstrand 2005 [39]	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder-ate
Hamada 2007 [40]	Yes	Unclear (a)	Unclear (c)	Yes	Unclear	Yes	Low
Jeon2006 [57]	Yes	Unclear (a)	Yes	Yes	Yes (d)	Yes	Low
Kurashima2010 [67]	Yes	Yes	Unclear (c)	Yes	Unclear	Yes	Low
Lynch 2005 [72]	Yes	Yes	Yes	Yes	No (e)	Yes	Moder-ate
ManaliI2008 [73]	Yes	Yes	Yes	Yes	Not reported	Unclear(b)	Low
Mejia 2009 [74]	Yes	Unclear (a)	Yes	Yes	Yes (b)	Yes	Moder-ate
Mogulkoc 2001A [75]	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder-ate
Mura 2012[77]	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder-ate
Richeldi 2012A[108]	Yes	Unclear (a)	Unclear (c)	Yes	Yes (d)	Yes	Moder-ate
Schmidt2011[106]	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder-ate
Sumikawa	Yes	Yes	Yes	Yes	Yes (d)	Yes	Moder

Reference	Representative population sample	Loss to follow up described	Prognostic factor measured appropriately	Outcomes adequately measured	Confounders accounted for	Appropriate statistical analysis	Quality
2008 [121]							-ate

(a) Dropouts not reported.
(b) No confounding factors were identified or included in the analysis.
(c) No detail provided on how prognostic factors were measured.
(d) Some confounding factors in protocol adjusted for.
(e) Did not adjust for any confounding factors in protocol.

7.5 Do serial pulmonary function tests (PFTs) (resting spirometric, gas transfer measurement and oxygen saturation) predict prognosis of IPF?

7.5.1 Overview

Sixteen studies were relevant to the clinical question and included in the review: DuBois 2013[27], Caminati 2009[13], DuBois 2011A[26], DuBois 2011B[28], Hallstrand 2005[39], Hamada 2007[40], Jeon 2006[57], Kurashima 2010[67], Lynch 2005[72], Manali 2008[73], Mejia2009[74], Mogulkoc 2001A[75], Mura 2012[79], Richeldi 2012A[108], Schmidt2011[111], Zappala2010[133].

- Sixteen studies looked at people with IPF. Two studies investigated survival in patients with UIP and IPF and did not distinguish between these groups in their analysis [57, 67]
- In two studies [67,74] the population also included emphysema in some cases.
- Two studies investigated the value of oxygen saturation on prognosis[13,39]
- One study investigated progression free survival[133]
- Eight studies looked at baseline PFTs[2,40, 57, 67, 72, 73,75]
- Five studies looked at serial PFTs [2, 4; 75, 111, 133]

A summary of the characteristics of included studies is given in Table 22 and study quality is presented in Table 23.

See forest plots in Appendix E, evidence tables in Appendix F, and unit costs in Table 29.

7.5.2 Results

Table 25: Baseline value of PFTs in predicting mortality/ survival - Clinical summary of findings

Reference	Prognostic Factor	Confounders adjusted for	Effect size
Caminati 2009[13] (n=44)	Baseline resting room air arterial oxygen saturation Baseline FVC (L) Baseline DLCO (mL/min/mmHg)	Age and sex only	Sat. O_2 rest: HR 0.816 (95% CI: 0.537-1.241), p value: 0.3416 FVC: HR 0.365(95% CI 0.124-1.078) p value 0.0681 DLCO: HR 0.723 (95% CI 0.548-0.954) p value 0.0219
DuBois2011A[26] (n = 1099)	Change in percent-predicted FVC </= 50%, 51%-65%, 66%-79%, >/=80%	Confounding factors adjusted for: age, oxygen use, surgical lung biopsy, history of respiratory hospitalisation, drug treatment, physiologic % predicted FVC, 24 week change in % predicted FVC, % predicted DLCO, 24 week change in % predicted DLCO, dyspnoea and HRQL UCSD SOBQ and 24 week change in UCSD SOBQ	</=50% vs. >/=80%: HR 5.79 (95% CI:2.55-13.15) p value <0.001 51% - 65% vs. >/=80%: HR 3.54 (95% CI: 1.95-6.44) p value <0.001 66%-79% vs. >/=80%: HR 2.20 (95% CI:1.19-4.09) p value <0.001
DuBois2011B[28] (n=1156)	Change in percent-predicted FVC </= 50%, 51%-65%, 66%-79%, >/=80%	Patients receiving active drug treatment during were adjusted for in the analysis, but study did not report adjusting of other confounders as identified in the protocol	</=50% vs. >/=80%: HR 7.44 (95% CI:3.28-16.87) p value <0.001 51% - 65% vs. >/=80%: HR 4.09 (95% CI: 1.87-8.98) p value <0.001 66%-79% vs. >/=80%: HR 1.97 (95% CI:0.85-4.55) p value 0.111

Reference	Prognostic Factor	Confounders adjusted for	Effect size
DuBois2013[27] (n=748)	Baseline percent-predicted FVC </=50% vs. >/=80%: 51% - 65% vs. >/=80%: 66%-79% vs. >/=80%:	Age, respiratory hospitalisations, change in FVC, 6MWD, and change in 6MWD.	All-cause mortality </=50% vs. >/=80%: HR: 6.86 (95% CI:1.99-23.60), p value <0.01 51% - 65% vs. >/=80%: HR: 2.92 (95% CI: 1.39-6.13), p value <0.01 66%-79% vs. >/=80%: HR: 2.17 (95% CI:1.02-4.63), p value 0.05
Hallstrand 2005[39] (n=28)	Baseline resting room air arterial oxygen saturation	The multivariable model included age, sex, FVC % predicted, time from the onset of symptoms and supplemental oxygen administration during the test as confounding factors	Arterial oxygen saturation Relative hazard (95% CI): 1.06(0.83–1.37) p value: 0.637
Hamada 2007[40] (n=25)	Baseline % DLCO <40	Stepwise regression model, adjusting for age, gender, PaO_2, PvO_2, mPAP, cardiac index and %VC	Low DLCO <40% (n=25) RR 2.70 (95% CI: 1.46 to 4.99)
Jeon 2006[57] (n=88)	Baseline FVC, % predicted per 10% decrease Baseline DLCO, % predicted per 10% decrease	Age, sex, severity of dyspnoea, FVC, DLCO and treatment	FVC: HR 1.7 (95% CI: 1.2-2.3) p value 0.004 DLCO: HR 1.5 (95% CI: 1.1-2.1) p value 0.033
Kurashima 2010[67] (n=660)	Baseline FVC, % predicted per 1 % (n=362) Baseline DLCO, % predicted per 1 % (n=251)	Not clearly reported	%FVC predicted per 1% (n=362) HR 0.988 (95% CI: 0.967-1.010) p value: 0.27 %DLCO predicted per 1% (n=251): HR 0.987 (95%CI: 0.971-1.002) p value: 0.21
Lynch 2005[72] (n=315)	Baseline % predicted DLCO	Cox proportional hazard model stratified by smoking status Overall disease extent score on CT,	HR 0.94 (95% CI: 0.90- 0.98) p value: 0.004

Reference	Prognostic Factor	Confounders adjusted for	Effect size
		reticulation pattern score, honeycomb pattern score, predominant pattern reticulation A _a gradient and current O_2 use.	
Manali 2008[73] (n=25)	Baseline FVC, % predicted	Not reported	FVC: RR 1.045 (95% CI: 0.956-1.142) p value: 0.033
Mejia2009 (n=110)	Baseline FVC <50% predicted	Male gender, emphysema, CT fibrotic score	FVC<50% predicted: HR 2.6 (95% CI 1.19-5.68) p value 0.016
Mogulkoc 2001A[75] (n=115)	Baseline DLCO, % predicted per 1% decrease (n=85)	FEV1, FVC, TLC, DLCO, KCO, CT ground glass appearance	HR/OR 0.957 (95% CI 0.928-0.987) p value 0.005
Mura 2012[79]	Baseline DLCO % predicted	BMI, MRC dyspnoea score, 6MWD % predicted, desaturation@ 6MWD, PaO_2, FVC % predicted, DLCO % predicted, CPI, CT fibrosis score, BAL cell count	HR 0.93 (0.89- 0.97) p value 0.008

Table 26: 'Serial' value of PFTs in predicting mortality/ survival/ progression free survival - Clinical summary of findings

Reference	Prognostic Factor	Confounders adjusted for	Effect size
DuBois2011A[26] (n = 1099)	24 week absolute change in percent – predicted FVC </= - 10%, - 5% to – 9.9%, > -5%	Confounding factors adjusted for: age, oxygen use, surgical lung biopsy, history of respiratory hospitalisation, drug treatment, physiologic % predicted FVC, 24 week change in % predicted FVC, % predicted DLCO, 24 week change in % predicted DLCO, dyspnoea and HRQL UCSD SOBQ and 24 week change in UCSD SOBQ	</=-10% vs. >-5%: HR 7.99 (95% CI: 5.26-12.14) p value: <0.001 -5 to -9.9% vs. >-5% HR 2.60 (95% CI: 1.75-3.85) p value: <0.001
DuBois2011B[28] (n=1156)	24 week absolute change percentage predicted FVC </= - 10%, - 5% to – 10%, > -5%	Patients receiving active drug treatment during were adjusted for in the analysis, but study did not report adjusting of other confounders as identified in the protocol	</=-10% vs. >-5%: HR 4.78 (95% CI: 3.12-7.33) p value: <0.001 -5 to -10% vs. >-5% HR 2.14 (95% CI: 1.43-3.20)

Reference	Prognostic Factor	Confounders adjusted for	Effect size
			p value:0.012
Caminati 2009[13] (n=44)	Change in oxygen saturation over 12 months follow up compared to baseline Change in FVC at 12 months Change in DLCO at 12 months	Age and sex only	Change in oxygen saturation HR 0.25 (95% CI: 0.075-0.837) p value: 0.02 Change in FVC HR 0.142 (95% CI: 0.018-1.1) p value: 0.06 Change in DLCO HR 0.49 (95% CI: 0.232-1.036) p value: 0.06
DuBois2013[27]	24 week change in percent-predicted FVC	Age, respiratory hospitalisations, FVC, 6MWD, and change in 6MWD.	All-cause mortality 24 week change in percent-predicted FVC </=10% vs. >-5%: HR: 5.86 (95% CI:3.33-10.31), p value <0.01 -5% - -9.9% vs. >-5%: HR: 2.74 (95% CI: 1.61-4.68), p value <0.01
Mogulkoc 2001A[75] (n=115)	DLCO % predicted per 1% decrease, at 2 years (n=70)	FEV1, FVC, TLC, DLCO, KCO, CT ground glass appearance	HR/OR 0.923 (95% CI 0.863-0.98) p value 0.021
Richeldi 2012A[108]	12 month absolute and relative change in % predicted FVC	Gender, baseline age, O_2 use, FVC and DLCO	Death at 2 years (time to event) ≥5% decline in % predicted FVC at 12 months (adjusted OR/HR) 1.61 (0.89-2.92) relative change 2.89 (1.53-5.46) absolute change Death at 2 years (time to event) ≥10% decline in % predicted FVC at 12 months (adjusted OR/HR) 2.75 (1.46-5.17) relative change

Reference	Prognostic Factor	Confounders adjusted for	Effect size
			2.41 (1.15-5.05) absolute change Death at 2 years (time to event) ≥15% decline in % predicted FVC at 12 months (adjusted OR/HR) 3.18 (1.16-6.26) relative change 2.49 (1.02-6.06) absolute change
Schmidt2011[111] (n=321)	Change in FVC over 6 months (n=211)	Adjusted for age at diagnosis, sex and smoking status	% FVC predicted: 5: HR 1.8(95% CI: 1.2-2.7), p value 0.002 10: HR 1.4(95% CI: 0.9-2.1), p value 0.122 15: HR 1.1(95% CI: 0.6-1.8), p value 0.857 20: HR 2.0(95% CI: 1.0-4.0), p value 0.051
	Change in DLCO over 6 months (n=211)		% DLCO predicted: 10: HR 1.7(95% CI: 1.1-2.5), p value 0.011 15: HR 1.6(95% CI: 1.1-2.5), p value 0.029 20: HR 1.8(95% CI: 1.1-3.0), p value 0.030 25: HR 2.3(95% CI: 1.2-4.2), p value 0.010
	Change in FVC over 12 months (n=144)		% FVC predicted: 5: HR 1.8(95% CI: 1.2-2.9), p value 0.012 10: HR 2.4(95% CI: 1.5-3.8), p value <0.001 15: HR 2.6(95% CI: 1.6-4.5), p value <0.001 20: HR 3.6(95% CI: 1.9-6.9), p value <0.001
	Change in DLCO over 12 months (n=144)		% DLCO predicted: 10: HR 2.2(95% CI: 1.4-3.5), p value 0.001 15: HR 2.3(95% CI: 1.5-3.7), p value <0.001 20: HR 3.0(95% CI: 1.8-4.9), p value <0.001 25: HR 3.5(95% CI: 2.0-6.1), p value <0.001
Zappala 2010 [133]	Decline in FVC at 6 months -adjusted for DLCO in IPF (n=84)	Age, sex, smoking status and baseline disease severity	5-10% decline in FVC: HR 3.33 (1.61-6.88), p value <0.001
	Progression free survival patients with 5-10%	Age, sex, smoking status and baseline disease	5-10% decline in FVC compared with stable disease: HR 1.82 (0.97-3.40), p value 0.06

Reference	Prognostic Factor	Confounders adjusted for	Effect size
	decline in FVC (n=84)	severity	5-10% decline in FVC compared with stable disease- adjusted for baseline DLCO: HR 2.56 (1.17-4.38), p value 0.02

7.6 Does baseline sub-maximal exercise testing predict prognosis of IPF?

7.6.1 Overview

The 3 papers which were included in the review were; DuBois 2013[27]Hallstrand 2005[39] and Caminati 2009.[13]

- One study took their patient group from a transplant centre, which introduces another level of bias as patients in this group are of a younger age and in better health than the general IPF population[39].
- Only mortality data was reported, no paper reported any of the other outcomes of interest.
- One study[39] looked at baseline 6MWD measurement and another investigated the change in 6MWD and mortality.[13]
- One study[2] looked at baseline and serial 6MWD.
- All 3 studies adjusted for some of the confounders identified in the protocol, but not all.

A summary of the characteristics of included studies is given in Table 22 and study quality is presented in Table 23.

See forest plots in Appendix E, evidence tables in Appendix F and unit costs in Table 29.

7.6.2 Results

Table 27: 'Baseline' value of sub-maximal exercise testing in predicting survival - clinical summary of findings

Reference	Prognostic Factor	Confounders adjusted for	Effect size
DuBois2012[29]	Baseline 6MWD	Age, respiratory hospitalisations, FVC, change in FVC, and change in 6MWD.	All-cause mortality <250m vs. >/=350m: HR: 2.12 (95% CI: 1.15 to 3.92), p value 0.02 250-349m vs. >/=350m: HR: 1.28 (95% CI: 0.74-2.21), p value 0.38
Hallstrand 2005 [39]	6MWD as a continuous variable, 30-metre units to mortality	The multivariable model included age, sex, FVC % predicted, time from the onset of symptoms and supplemental oxygen administration during the test as confounding factors	Relative hazard (95% CI): 0.91 (0.81–1.02) p value: 0.098
Caminati 2009 [13]	6MWD as a continuous variable, to mortality	Age and sex only	Hazard ratio (95% CI): 0.995 (0.990-0.999) p value: 0.0308

Table 10: 'Serial' value of sub maximal exercise testing in predicting mortality/ survival - clinical summary of findings

Reference	Prognostic Factor	Confounders adjusted for	Effect size
Caminati 2009 [13]	Change in 6MWD as a continuous variable, at 12 months follow up to mortality (change at 12 months – basal value)	Age and sex only	HR 0.994 (95% CI: 0.988-1) p value: 0.05
DuBois2012[29]	Serial 6MWD, at 24 weeks	Age, respiratory hospitalisations, FVC, change in FVC, and change in 6MWD.	All-cause mortality <-50m vs. >/=-25m: HR: 2.73 (95% CI: 1.60-4.66), p value <0.01 -50 to -26m vs. >/=-25m:

Reference	Prognostic Factor	Confounders adjusted for	Effect size
			HR: 2.94 (95% CI: 1.56-5.53) p value <0.01

7.7 Does baseline echocardiography predict prognosis of IPF?

7.7.1 Overview

One retrospective study was relevant to the clinical question and was included in the review: Mejia 2009[74]. The population included patients with co-existing IPF and emphysema; however it is unclear if the final analysis included this group.

Only mortality was provided as an outcome; no other outcomes in the protocol were identified.

Confounding factors adjusted for were: male gender, emphysema and CT fibrotic score; other factors identified in the protocol were not adjusted for.

A summary of the characteristics of included studies is given in Table 22 and study quality is presented in Table 23.

See forest plots in Appendix E, evidence tables in Appendix F, and unit costs in Table 29.

7.7.2 Results

Table 28: Baseline value of echocardiography in predicting survival – Clinical summary of findings

Reference	Prognostic Factor	Confounding factors adjusted for	Effect size	Interpretation
Mejia 2009 [74]	Estimated systolic pulmonary artery pressure (ESPAP) > 75 mmHg on echocardiography	Male gender, emphysema, CT fibrotic score	ESPAP HR:2.25 95% CI: 1.12-4.54 p value: 0.022	ESPAP >75mmHg at baseline was a predictor of worse prognosis (p value 0.022)

7.8 Do baseline CT scores predict prognosis of IPF?

7.8.1 Overview

Four studies were relevant to the clinical question and were included in the review: Best 2008[8], Lynch 2005,[72] Mogulkoc 2001A[75] and Sumikawa 2008.[121]

One study was prospective[75] and the 3 other studies were retrospective.[8, 72,121] One study was an analysis of a cohort from an RCT.[72]

- all studies looked at people with IPF
- all studies provided data on mortality
- no studies provided data on the other outcomes in the protocol
- all studies adjusted for some, but not all of the confounders identified in the protocol.

A summary of the characteristics of included studies is given in Table 22 and study quality is presented in Table 23.

See forest plots in Appendix E, evidence tables in Appendix F and unit costs in Table 29.

7.8.2 Results

Table 29: Baseline value of CT scores for predicting survival – clinical summary of findings

Reference	Prognostic Factor	Confounders adjusted for	Effect size
Best 2008[8]	CT findings	Unclear Multivariable analysis - "other possible predictors were taken into account"	Fibrosis (n=33) OR estimate 1.104 (95% CI: 1.018-1.198) p value: 0.017
Lynch 2005[72]	CT FVC DLCO	Adjusted for: overall disease extent on CT, reticulation pattern score, honeycomb pattern score, predominant pattern= reticulation, % predicted DLCO, A-a gradient, current O_2 use	Overall extent of fibrosis score: HR 2.71 (95% CI: 1.61- 4.55) p value: <0.0001
Mogulkoc 2001A[75]	PFTs and CT findings	Adjusted for: FEV1, FVC, TLC, DLCO, KCO and CT ground glass appearance	CT fibrosis score- baseline: HR/ OR 0.957 (95% CI: 1.726-3.914) p value: 0.026 CT fibrosis score-at 2 year follow-up: HR/ OR 6.274 (95% CI: 1.317-29.897) p value: 0.021
Sumikawa 2008[121]	CT findings	On multivariate analysis, the variables were selected using a stepwise procedure including each one of the following CT findings: presence of ground-glass attenuation; airspace consolidation; nodules; interlobular septal thickening; thickening of bronchovascular bundles; intralobular reticular opacities; irregular interlobular septal thickening; non-septal linear or plate-like opacities; presence of honeycombing, cysts, emphysema, architectural distortion, or traction bronchiectasis; fibrosis score; the extent of disease close to the hilum; and upper, lower, peripheral, dependent, peribronchovascular, and asymmetric predominant distribution. Findings were retained if they contributed to the power of the regression equation ($P < 0.10$)	Traction bronchiectasis: HR 1.30 (95% CI 1.18-1.43) no p value Fibrosis score: HR 1.10 (95% CI 1.03-1.19) no p value

7.9 Economic evidence

7.9.1 Published literature

No health economic literature assessing an intervention for a prognostic purpose in an IPF population was identified. No studies were selectively excluded.

7.9.2 Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided below to aid consideration of cost effectiveness.

Table 30: Unit costs of prognostic interventions

Intervention	Mean unit cost (interquartile range)	Notes
Outpatient appointment	£162 (£136 to £231)	Consultant led, face to face, outpatient procedure code: 340. Most likely to be conducted as part of the diagnostic pathway.
Lung Volume Studies	£187 (£122 to £298)	Outpatient procedure; HRG code DZ45Z. Baseline conducted as part of the diagnostic pathway.
Simple airflow study	£168 (£135 to £195)	Outpatient procedure; HRG code DZ44Z. Note that this procedure is likely to be within the same episode as the lung function study and would be included under the cost of the lung volume study Baseline conducted as part of the diagnostic pathway.
Simple Gas Exchange Studies	£146 (£124 to £183)	Outpatient procedure; HRG code DZ40Z Note that this procedure is likely to be within the same episode as the lung function study, and would be included under the cost of the lung volume study. Baseline conducted as part of the diagnostic pathway.
Simple Lung Function Exercise Testing e.g. six minute walk, shuttle walk	£269 (£188 to £263)	Outpatient procedure; HRG code DZ32Z. This intervention may or may not be included in the diagnostic work up and is likely to occur in a separate episode to that where lung volume, airflow and gas exchange is studied.
Simple echocardiogram	Outpatient: £84 (£52 to £92) Direct Access: £91 (£55 to £88)	Outpatient procedure; HRG code RA60Z. Note that this is an unbundled cost so the cost would be in addition to another procedure or consultation. This intervention is not normally undertaken as part of the diagnostic pathway.

Abbreviations: HRG = Health Resource Group

Source: NHS Reference costs 2010-2011[25]

7.10 Evidence statements

Note:

- *Only the results of the multivariable analysis have been reported in evidence tables and included in the final analysis.*
- *Hazard ratios for declines of PFT measures/predicted values are stated below (in some instances hazard ratios were inversed in order to present all hazard ratios according to declines in PFTs).*
- *Hazard ratios presented below were also calculated per 5% decline for FVC and per 10% decline for DLCO*

PFTs:

Baseline FVC

Moderate quality evidence suggests that a baseline FVC (% predicted) in people with IPF (mean age not reported) is associated with an increased risk of all-cause mortality (<=50 vs. >=80: HR 6.86, 95% CI 1.99-23.60, p value <0.01; 51 to 65 vs. >=80: HR 2.92, 95% CI 1.39-6.13, p value <0.01; 66 to 79 vs. >=80: HR 2.17, 95% CI 1.02-4.63, p value 0.05) (one study, N=748)[27].

Low quality evidence suggests that low FVC (L) at baseline in people with IPF (mean age 61.9 years) is associated with an increased risk in mortality (HR 2.74, p value 0.0681)[13].

Moderate quality evidence suggests that a decline in FVC (% predicted) in people with IPF (mean age not reported) enrolled in a clinical trial at baseline (</=50% vs. >/=80%) is associated with an increased risk of death at 1 year (HR 5.79, p value <0.001) (one study, N=1156)[28].

Moderate quality evidence in people with IPF (mean age not reported) enrolled in a clinical trial suggests that a change in FVC (% predicted) from baseline (</=50% vs. >/=80%) is associated with an increased risk of death at 1 year (HR 7.44, p value <0.001) (one study, N=1156)[28].

Low quality evidence in people with IPF and UIP enrolled from a hospital suggests that a 5% predicted decrease in FVC between patients at baseline is associated with an increased risk of mortality (HR 1.30, p value 0.004) (one study, N=39)[57].

Low quality evidence in people with UIP (mean age 72.9 years) enrolled from a hospital suggests that a 5 % predicted decrease in baseline FVC is associated with a decreased risk of mortality (HR 0.988, p value 0.27) (one study, N=439)[67].

Low quality evidence in people with IPF (mean age 64 years) enrolled from a respiratory outpatient clinic suggests that baseline FVC per 5% predicted decrease, is associated with a decreased risk of mortality (RR 0.978, p value 0.033) (one study, N=25)[73].

Low quality evidence in people with IPF (mean age unclear) enrolled from a national institute of respiratory diseases suggests that a low FVC (<50% predicted) at baseline is associated with an increased risk of mortality (HR 2.6, p value 0.016) (one study, N=110)[74].

Baseline DLCO

Low quality evidence in people with IPF (mean age 61.9 years) suggests that low DLCO (mL/min/mmHg) at baseline is associated with an increased risk in mortality (HR 1.38, p value 0.0219)[13].

Low quality evidence in people with IPF (mean age 62 years) enrolled at a university hospital suggests that low DLCO (<40% predicted) at baseline is associated with an increased risk of mortality (RR 2.70, no p value given) (one study, N=78)[40].

Low quality evidence in people with IPF and UIP enrolled from a hospital suggests that baseline DLCO per 10% predicted decrease is associated with an increased risk of mortality (HR 1.5, p value 0.033) (one study, N=39)[57].

Low quality evidence in people with UIP (mean age 72.9 years) enrolled from a hospital suggests that baseline DLCO per 10% predicted decrease is associated with a decreased risk of mortality (HR 0.987, p value 0.21) (one study, N=439)[67].

Moderate quality evidence in people with IPF (mean age not reported) enrolled in a multinational study suggests that baseline DLCO per 10% predicted decrease is associated with an increased risk of mortality (HR 1.86, p value 0.004). The study was of moderate quality (one study, N=315)[72].

Low quality evidence in people with IPF (mean age 56±8 years) enrolled from a transplant centre suggests that baseline DLCO per 10% predicted decrease, is associated with an increased risk of mortality (HR/OR 1.55, p value 0.005) (one study, N=115)[75].

Low quality evidence in people with IPF (mean age 67 years) enrolled in a prospective cohort suggests that a low baseline DLCO Is associated with an increased risk of mortality (HR 0.93, no p value given) (one study, N=70)[79]

Serial FVC

Moderate quality evidence in people with IPF suggests that at 24-weeks, change in FVC (% predicted) is a significant independent predictor of all-cause mortality (<=-10 vs. >-5: HR 5.86, 95% CI 3.33-10.31, p value <0.01; -5 to -9.9 vs. >-5: HR 2.74, 95% CI 1.61-4.68, p value <0.01) (one study, N=748)[27].

Low quality evidence in people with IPF (mean age 61.9 years) suggests that a decline in FVC over 12 months is associated with a decreased risk in mortality (HR: 0.142, p value 0.06) (one study, N=44)[13].

Moderate quality evidence in people with IPF (mean age not reported) enrolled in a clinical trial suggests that a decline in FVC of <10%, compared to <5% over 24 weeks is associated with an increased risk of death at 1 year (HR: 7.99, p value <0.001) (one study, N=1099)[26].

Moderate quality evidence in people with IPF (mean age not reported) enrolled in a clinical trial suggests that a decline in FVC of <10%, compared to <5%, over 24 weeks is associated with an increased risk of death at 1 year (HR: 4.78, p value <0.001) (one study, N=1156)[28].

Moderate quality evidence in people with IPF (mean age 67.0 years) enrolled in a prospective cohort suggests that a decline in FVC of 5%, 10% and 15% is associated with a higher risk of death (HR 1.62, 2.75, 3.18, respectively, no p value given). The study was of moderate quality (one study, N=142)[108].

Moderate quality evidence in people with IPF (mean age 63.2 years) enrolled from secondary care suggests that a decline in FVC of 10%, 15% and 20% (HR: 1.4, 1.1 & 2.0 at 6 months and 2.4, 2.6 & 3.6 at 12 months, respectively) over 6 and 12 months is associated with an increased risk of mortality) (one study, N=321) [111].

Moderate quality evidence in people with IPF (mean age 57.4 years) enrolled from secondary care suggests that a 5-10% decline in FVC over 6 months is associated with an increased risk of mortality when adjusted for DLCO, compared with stable disease (HR: 3.33, p value <0.001) (one study, N=84) [133].

Serial DLCO

Low quality evidence in people with IPF (mean age 61.9 years) suggests that a decline in DLCO over 12 months is associated with a decreased risk in mortality (HR: 0.49, p value 0.06) (one study, N=44) [13].

Low quality evidence in people with IPF (mean age 56±8 years) enrolled from a transplant centre suggests that DLCO per 10% predicted decrease at 2 years is associated with an increased risk of mortality (HR/OR 2.23, p value 0.021)(one study, N=115)[75].

Moderate quality evidence in people with IPF (mean age 63.2 years) enrolled from secondary care suggests that a decline in DLCO of 15%, 20% and 25% (HR: 1.6, 1.8 & 2.3 at 6 months and 2.3, 3.0 & 3.5 at 12 months, respectively) over 6 and 12 months is associated with an increased risk of mortality (one study, N=321) [111]

Oxygen saturation:

Baseline

Low quality evidence in people with IPF (mean age 61.9 years) suggests that resting baseline oxygen saturation is associated with a decreased risk in mortality (HR: 0.816) (one study, N=44) [13].

Moderate quality evidence in people with IPF enrolled from a transplant centre suggests that resting baseline oxygen saturation is associated with an increased risk of mortality (HR: 1.06). There is uncertainty in this effect and the study was moderate quality (one study, N=28)[39].

Change over 12 months

Low quality evidence in people with IPF (mean age 61.9 years) suggests that a decline in resting oxygen saturation over 12 months is associated with a decreased risk in mortality (HR: 0.25). There is uncertainty in this effect and the study was low quality (one study, N=44) [13].

Progression Free Survival

Moderate quality evidence in people with IPF (mean age 57.4 years) enrolled from secondary care suggests that a 5-10% decline in FVC over 6 months is associated with a decline in progression free survival when adjusted for DLCO, compared with stable disease (HR: 2.56). There is uncertainty in this effect and the (one study, N=84) [133].

Sub-maximal exercise testing:

6MWD at baseline

Moderate quality evidence in people with IPF (mean age not reported) suggests that 6MWD (meters) at baseline is associated with an increased risk of all-cause mortality (<250 vs. >=350: HR 2.12, 95% CI 1.15-3.92, p value = 0.02; 250 to 349 vs. >=350: HR 1.28, 95% CI 0.74-2.21, p value = 0.38) (one study, N=748)[27].

Low quality evidence in people with IPF (mean age 62.7 years) suggests that for every 30 metre increase in distance walked there is association with a decreased risk of mortality (HR=0.91). There is uncertainty in the effect (one study, N=28)[39].

Low quality evidence in people with IPF (mean age 61.9 years) suggests that for every unit increase in distance walked there is an association with a decreased risk of mortality (HR=0.995) (one study, N=44)[13].

Serial 6MWD

Moderate quality evidence in people with IPF (mean age not reported) suggests that 24-week change in 6MWD (meters) is associated with an increased risk of all-cause mortality (<-50 vs. >=-25: HR 2.73, 95% CI 1.60-4.66, p value <0.01; -50 to -26 vs. >=-25: HR 2.94, 95% CI: 1.56-5.53, p value <0.01) (one study, N=748)[27].

Low quality evidence in people with IPF (mean age 61.9 years) suggests that for every unit increase in distance walked there is an association with a decreased risk for mortality (HR=0.994) from baseline to 12 months follow up (one study, N=44)[13].

Echocardiography

Low quality evidence in people with IPF (mean age unclear) enrolled from a national institute of respiratory diseases suggests that an estimated systolic pulmonary arterial pressure on baseline echocardiography >75mmHg is associated with an increased risk of mortality (HR: 2.25; p value 0.022) (one study, N=110)[74].

CT

Low quality evidence in people with IPF (mean age 63 years) enrolled in a clinical trial suggests that baseline fibrosis on CT is associated with an increased risk of mortality (HR: 1.10)(one study, N=167)[8].

Moderate quality evidence in people with IPF (mean age not given) with IPF suggests that the overall extent of fibrosis score on baseline CT is associated with an increased risk of mortality (HR 2.71, p value <0.0001) (one study, N=315)[72].

Low quality evidence in people with UIP suggests that a higher CT fibrosis score is associated with an increased risk of mortality (HR 2.067, p value 0.026) (one study, N=85)[75].

Low quality evidence in people with IPF/UIP suggests that the presence of traction bronchiectasis on CT is suggestive of an increased risk of mortality (HR 1.30, p value not given) (one study, N=98) [121].

Low quality evidence in people with IPF/UIP suggests that fibrotic score on CT is suggestive of an increased risk of mortality (HR 1.10, p value not given) (one study, N=98)[121].

Economic

- No relevant economic evaluations were identified that compared interventions with a purpose of achieving a prognosis in an IPF population.

7.11 Recommendations and link to evidence

Recommendations	**10. Measure the initial rate of decline in the person's condition, which may predict subsequent prognosis, by using lung function test results (spirometry and gas transfer) at:** • **diagnosis and** • **6 months and 12 months after diagnosis. Repeat the lung function tests at shorter intervals if there is concern that the person's condition is deteriorating rapidly.**

Relative values of different outcomes	The GDG considered time to event outcomes, mortality or survival, to be the critical outcomes for predicting prognosis. One study investigated whether lung function tests predict progression free survival. No studies were retrieved which investigated lung function tests and time to lung transplantation, time to acute exacerbation or hospitalisations.
Trade-off between clinical benefits and harms	The GDG did not consider there to be any harms related to patients undergoing spirometry or measurements of gas transfer. The GDG acknowledged the difficulties of predicting prognosis using spirometry and gas transfer when considering the different rates of disease progression (stable versus severe) in people with IPF. Also, obtaining reasonable PFT measures (including reproducible measures) are not always possible as these tests are dependent on the patients effort, as cough and dyspnoea may interfere.
Economic considerations	No published economic evidence was identified to inform this recommendation. When taking into account the potential cost effectiveness of the prognostic interventions, the GDG considered: the unit cost of each intervention, whether the intervention would be undertaken for a purpose other than prognosis in the care pathway for an IPF patient, as well as the clinical benefit. Baseline PFTs have been recommended to be routinely performed alongside CT scans as part of the diagnostic pathway. The GDG thought there would be negligible additional cost in using the same results for prognostic purposes. The unit costs of PFTs were taken from NHS reference costs. However, these were thought to be an overestimation of the cost of the intervention (potentially as a result of the unit cost being an average cost for a group of lung function tests, of which spirometry is the cheapest) to greatly overestimate the cost of the intervention. The GDG advised that current referral charges for spirometry were approximately £40, so they recognised that the cited NHS unit cost of £154 (IQR: £94-£183) was unlikely to be reflective of the cost incurred by the NHS. They advised that on average a gas transfer test would take between 30 and 60 minutes per patient and be conducted by a staff member from Band 5 to 8A, whilst spirometry would be conducted by a staff member from band 3 to 7, and would take between 5 to 15 minutes per patient. As such the gas transfer would be more costly than spirometry. However, serial gas transfer, in addition to spirometry was agreed to be justifiable due to the additional information it provides, especially in a subgroup of IPF patients who also have emphysema (approximately one third of patients). The GDG considered the additional resource use required to undertake serial PFTs as well as follow up appointments that may be required to explain the findings to patients. They considered that the NHS unit costs incurred to undertake PFTs was low in comparison to other interventions (i.e. CT) and additional resource use was worthwhile to obtain a more accurate prognosis, especially as it may influence clinical management (that is, initiating discussions regarding end of life care for people with a poorer prognosis). The optimal time interval between tests, the duration of time the course of serial tests run for, and timing of the serial tests in relation to disease progression were discussed

	as important economic considerations. It was noted that prognosis could be reviewed at follow up, and that the optimal timing of monitoring and review has relevance for this recommendation. It was noted that there was sufficient value in establishing the rate of disease progression to justify the cost of prognostic review at diagnosis and at 6 months post diagnosis, as the serial change was likely to suggest the rate of disease progression thereafter, assuming a linear rate of decline. However, it was also recognised that substantial gaps in knowledge existed on how the results of serial PFTs may predict a change in rate of disease progression, the likelihood of acute exacerbation and mortality given a previously stable and slow rate of decline. A key driver of the cost effectiveness of a prognostic intervention is the improvement of the management that follows a certain prognostic result. Determining whether or not a patient was likely to have rapid deterioration influences the decision to refer to palliative services and lung transplant. For this reason, there was consensus that it would be important to have a prognostic review of a subgroup of patients suspected of rapid deterioration at 3 months. It was felt that a clinical history and a patient's own feeling of disease progression (that is, acute worsening of breathlessness) would be valuable in identification of patients who had a high probability of rapid progression within the first 6 months of diagnosis. However, given that the majority of patients were likely to have a more stable course of disease progression and given the limited evidence available, routine prognostic review at three months could not be justified. Given the infrequency of the recommended prognostic review, there was a strong consensus that the prognostic interventions should be conducted in a secondary care setting with appropriate equipment and expertise of staff to maximise the accuracy of the results. In current practice, prognostic review offers an opportunity for patient contact with the specialist centre and as such may currently serve a dual function. The effectiveness and cost effectiveness of interventions to provide for patient contact, review and support is considered in other chapters of this guideline.
Quality of evidence	Evidence comprised of seventeen studies, (low to moderate quality) and the effect sizes were generally conclusive of FVC and DLCO as prognostic factors, both when measured at baseline and when measured serially.
Other considerations	The GDG regarded patient communication to be an extremely important consideration for these recommendations and this is reflected in recommendation 1.3.1. Communication included information at all stages of disease progression for patients and carers regarding: life expectancy, expectations of future symptoms and management, treatment options and functional ability. There is an advantage of continuing to monitor patients to assess whether patients are stable or deteriorating and to provide reassurance and support. Patients that are rapidly deteriorating (acute exacerbation) may not benefit from further prognostic tests. However, follow-up of these patients remain important for monitoring purposes. The GDG acknowledged that the American Thoracic Society (ATS) identified baseline FVC as an unclear prognostic predictor, whereas DLCO was found to be more reliable predictor and a decline in FVC over 6 or 12 months was reliably associated with decreased survival. Less consistently, a decline in DLCO has also been associated with decreased survival. Research recommendation The GDG agreed that the lack of evidence for echocardiography and CT scores justified

	developing a research recommendation to address the prognostic value of echocardiography and CT scoring in people with IPF. For further information on research recommendations see Appendix P.

Recommendations	**11. Do not use the 6-minute walk distance at diagnosis to estimate prognosis. (The 6-minute walk test may be useful for other purposes, see recommendation 14).**
Relative values of different outcomes	The GDG considered time to event outcomes, mortality or survival, to be the most important for predicting prognosis. No studies were retrieved which investigated 6MWD and progression free survival, time to lung transplantation or previous hospitalisations.
Trade-off between clinical benefits and harms	No evidence for sub-maximal exercise tests other than the 6MWT was identified. However, the GDG considered the 6MWT to be the most widely used, reliable and validated tool compared to other sub-maximal tests. The 6MWT is not directly considered a prognostic test, but is required to monitor patients with confirmed IPF in order to assess oxygen requirements. It is confounded by disease progression and time of diagnosis. The GDG discussed the small risks associated with 'exertional tests', such as fainting, but did not consider that people with IPF undergoing the 6MWT would be pushed to this extreme. No subgroups of patients that would be able to undertake the 6MWT and would not be able to undertake a PFT were identified. The GDG consensus was that a 6MWT does not offer incremental benefit to baseline and serial PFT in determining a prognosis in people with IPF.
Economic considerations	No economic evidence was identified for the use of the 6MWT to inform prognosis. There is currently wide variation in the use of the 6MWT and it is not conducted for the purpose of diagnosis, and most often performed for a primary purpose other than for prognosis i.e. as part of ambulatory oxygen assessment or for pulmonary rehabilitation. The GDG considered the NHS reference costs for exercise testing, which is calculated as an average of the cost of several different types of exercise test. The GDG considered the unit cost to be an overestimate for the cost of a 6 minute walk test as a single intervention. This is because the 6MWT can be performed in little time and as part of a consultation with an ILD nurse. AS such. The 6MWT was thought to use less healthcare resource than other tests categorised within the same healthcare resource group on which the reference unit cost of exercise testing is derived. Nonetheless, given that the clinical evidence suggested a 6MWT did not offer additional value in determining a prognosis to other interventions, the additional cost for performing this test for prognostic purposes alone could not be justified. In current practice, prognostic review offers an opportunity for patient contact with the specialist centre and as such may currently serve a dual function. The effectiveness and

	cost effectiveness of interventions to provide for patient contact, review and support is reviewed in other chapters of this guideline.
Quality of evidence	Evidence comprised of two studies (low to moderate quality) and the results showed that the distance walked during a 6MWT at baseline did not add significantly to estimation of patient prognosis when added to other routinely obtained tests. Serial change in distance walked probably improved estimates of survival to a small extent when added to other measures, principally change in FVC. The GDG acknowledged that the ATS guideline concludes that the prognostic value of the 6MWT is limited due to the lack of standardisation of the test in people with IPF. Desaturation during 6MWT, as well as shorter walk distance and delayed heart rate recovery after walk testing have been associated with an increased risk of subsequent mortality. One study defined a threshold of <72% predicted 6MWT, which was not one known to the GDG or used in the UK. Therefore, the applicability of the results was limited and the GDG decided to exclude this part of the study on that basis. In the absence of convincing evidence to predict prognosis for 6MWT, the GDG did not consider the 6MWT to add extra value over other prognostic tests, but did discuss that value of the test for other patient management purposes.
Other considerations	The GDG regarded patient communication to be an extremely important consideration for these recommendations and this is reflected in recommendation 1.3.1. Communication included information at all stages of disease progression for patients and carers, where appropriate regarding: life expectancy; expectations of future symptoms and management; treatment options; and functional ability. There is an advantage of continuing to monitor patients to assess whether patients are stable or deteriorating and to provide reassurance and support. Patients that are rapidly deteriorating (acute exacerbation) may not benefit from further prognostic tests. However, follow-up of these patients remains important for monitoring purposes. Consideration of when to discharge a patient and sharing of patient care across specialities and healthcare settings (primary, secondary and tertiary) in order to monitor co-morbidities and reassure patients about their healthcare was also discussed. The GDG acknowledged that the ATS guideline concludes that the prognostic value of the 6MWT is limited due to the lack of standardisation of the test in people with IPF. Desaturation during 6MWT, as well as shorter walk distance and delayed heart rate recovery after walk testing have been associated with an increased risk of subsequent mortality. Research recommendations The prognostic evidence review also questioned the prognostic value of echocardiography and CT soring. The GDG agreed that the lack of evidence in these clinical areas justified developing a research recommendation to address the prognostic value of echocardiography and CT scoring in people with IPF. For further information on research recommendations see Appendix P.

Recommendations	**12. The consultant respiratory physician or interstitial lung disease specialist nurse should provide accurate and clear information (verbal and written) to people with idiopathic pulmonary fibrosis, and their families and carers with the person's consent. This should include information about investigations, diagnosis and management.** **13. Discuss prognosis with people with idiopathic pulmonary fibrosis in a sensitive manner and include information on:** • **the severity of the person's disease and average life expectancy** • **the varying courses of disease and range of survival** • **management options available.**
Relative values of different outcomes	These recommendations were agreed using informal GDG consensus methods. The importance of effective communication between healthcare professionals and people with IPF and their caregivers was identified by the GDG as an important consideration to facilitate good practice when informing patients of prognostic information.
Trade-off between clinical benefits and harms	These recommendations were agreed using informal GDG consensus methods.
Economic considerations	No economic evidence was identified to inform this recommendation.
Quality of evidence	These recommendations were based on informal GDG consensus.
Other considerations	The GDG regarded patient communication to be an extremely important consideration for these recommendations. Communication included information at all stages of disease progression for patients and carers regarding: life expectancy, expectations of future symptoms and management, treatment options and functional ability; as well as provision of wider welfare and lifestyle issues. GDG discussions centred on the importance of clear and tailored patient and carer information according to the patient's individual requirements, whilst acknowledging that requirements will differ throughout the progression of the disease. The expertise of the health professional and healthcare setting in which information is being provided was also considered important, with tertiary specialist care facilities providing increased confidence and reassurance to patients regarding their care. The GDG also acknowledged that some people would not want to know their prognosis.

8 Pulmonary rehabilitation

8.1 Review introduction

Shortness of breath, fatigue and reduced exercise tolerance are symptoms frequently experienced by people with IPF. The systemic consequences of COPD have been well characterised, particularly in respect of skeletal muscle dysfunction, but there is limited evidence of the impact that other lung diseases including IPF have upon musculoskeletal function.

Pulmonary rehabilitation (PR) is conventionally offered as a package of supervised exercise and education over a 6 week period by a multidisciplinary team. There is limited availability of PR for those with IPF.

8.2 Clinical questions and review methodology

The following clinical questions are included in this chapter:

8.2.1 What are the benefits of pulmonary rehabilitation programmes for people with confirmed IPF?

For full details see review protocol in Appendix C.

Table 31: PICO characteristics of review question

Population	Adults with confirmed IPF
Intervention/s	Pulmonary rehabilitation
Comparison/s	Best usual care/usual medical management Self-management
Outcomes	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates Other outcomes • Dyspnoea • Hospitalisations due to IPF complications (including IPF exacerbations) • Improvement in cough and breathlessness • Improvement in health-related quality of life • Performance on sub-maximal walk test (distance walked and lowest SaO_2) • Improvement in psychosocial health (including depression)
Study design	RCTs, systematic reviews of RCTs and cohort studies

8.2.2 What is the optimal course content, setting and duration for people referred for pulmonary rehabilitation programmes?

The protocol for this review question was the same as above, see Table 31.

The objectives of this review were to determine the benefits or harms of pulmonary rehabilitation and the requirements of a pulmonary rehabilitation programme to provide optimal symptomatic relief people with IPF. No restrictions were used for sample size, publication date, and the population was extended to include people with ILD as the GDG indicated that there would be limited literature

available on IPF people alone. Studies in abstract form were also included in order to capture all relevant data. Studies with an indirect population such as COPD were not included as the GDG considered that people with COPD have different disease trajectories and needs and are thus not comparable with people who have IPF.

8.3 Clinical evidence

We searched for systematic reviews, randomised controlled trials and cohort studies comparing pulmonary rehabilitation versus no treatment/usual care for people with ILD.

Thirteen studies[5,30,36,41,42,44,55,65,80,88,96,104,124] were identified which gave information on the benefits of pulmonary rehabilitation programmes for people with ILD. One abstract [36] was also retrieved. No studies were identified addressing the review question regarding the optimal course content, setting and duration for people referred for pulmonary rehabilitation programmes.

One Cochrane review[42] was identified, which included data from two randomised controlled trials[41,88] identified in the search on the use of physical training for ILD people. Additional data was extracted from this Cochrane review[42] which analysed IPF patient data separately. Holland et al were contacted to provide QoL data. This unpublished data on QoL domains for SF36 has been analysed and used by the GDG in their decision-making.

All the PR programmes lasted between 6-12 weeks, consisting of a mixture of educational lectures, supervised and unsupervised exercise, psychosocial support and self-management training. Two papers included looked at home based PR programmes with telephone support[29, 30]. Full details of the interventions can be seen in Table 2 and study evidence tables in Appendix F.

Nine observational studies were also identified, seven were prospective studies[55,65,80,96,104,124, 44] and two were retrospective studies.[5 30] Seven observational studies did not have control groups,[5,30,44,55,80,96,104] two studies[65,124] had control groups composed of people with COPD which have not been reported in this report, as ideally a control group would be made up of people with IPF receiving usual care, and indirect populations such as COPD were not considered.

Evidence in this review has been separated into randomised controlled trials and observational studies. Due to the lack of a control group /direct comparison with observational studies the data is shown in this review as reported in the study.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

8.3.1 Summary of included studies

Table 32: Summary of studies included in this review

Study	Intervention/ comparison	Population	Outcomes	Comments
Almoamary 2012[5]	8 weeks (18 sessions) pulmonary rehabilitation programme which comprised of education, exercise	ILD n= 21	6MWD (m) Distance on treadmill (m) Distance on bicycle (m) Distance on ergometer (m) Emergency department	Retrospective design bias. Does not account for confounding. Small sample size

Study	Intervention/ comparison	Population	Outcomes	Comments
	and psychosocial support		visits (no.) Outpatient department visits (days).	and single centre study – lacks generalisability.
Gaunaurd 2011 [36]	12 week PR programme, constituted of educational lectures and supervised exercise/ details of control group not specified	IPF n=6	6MWD Change in VO_2.	Reports on people who have completed the intervention portion of the study to date. Small sample size and single centre study – lacks generalisability Abstract- lack of detail. Results were calculated by the NCGC.
Ferreira 2009[30]	Data taken from 3 centres of a 6-8 weeks PR programme consisting of exercise and educational activities and psychosocial support	ILD n=99	Dyspnoea (Borg score) Dyspnoea (UCSD questionnaire) 6MWTD Depression(CES-D score) 6MWD, % change.	No control group. Confounding factors were not accounted for. Variation in practice with the use of oxygen during PR between the centres. Important differences between participating centres could be present that were missed due to inadequate numbers. Non-randomised.
Holland 2008[41] (including unpublished data received from the authors), Holland 2008[42] (Cochrane review from which additional data was extracted)	8 weeks supervised exercise programme/weekly telephone support	ILD n=57, IPF n=34	Change in 6MWT immediately following training Change in 6MWT at long-term follow-up Change in dyspnoea score immediately following training Change in dyspnoea score at long-term follow-up Change in quality of life immediately following training Change in quality of life at long-term follow-up Six month survival QoL-SF36 domains (unpublished data).	Large number of drop outs. The effect of disease aetiology and severity on response to exercise training– the study was not powered to adequately assess this outcome. Small sample size and single centre study – lacks generalisability.

Study	Intervention/ comparison	Population	Outcomes	Comments
Holland 2012[44]	Twice weekly supervised exercise program for eight weeks supplemented with an unsupervised home exercise program. Participants also attended an education and self-management program / no control group	IPF n=25 (only reported IPF data)	Dyspnoea: Change in CRQ dyspnoea domain (at 8 weeks & 6 months) Change in 6MWD (at 8 weeks & 6 months) Number of people achieving gains exceeding the MID for 6MWD (at 8 weeks & 6 months) Number of people achieving gains exceeding the MID for CRQ dyspnoea (at 8 weeks & 6 months).	Confounding factors were not accounted for. Small sample size. No control group. Non-randomised.
Jastrzebski 2006[55]	4 weeks hospital-based rehabilitation continued later at home/ no control group	ILD n=38, IPF n=13	Dyspnoea (MRC scale, baseline dyspnoea index, Borg scale) QoL (SGRQ domains).	No baseline data provided. Confounding factors were not accounted for. No control. Small sample size and single centre study – lacks generalisability.
Kozu 2011[65]	8 weeks outpatient programme comprising 2 classes per week/ same for COPD group	IPF n=45	Dyspnoea (MRC scale) Exercise capacity (6MWD) QoL (SF-36)	Large number of drop outs. Inconsistencies in reporting some data. Control group composed of COPD. Does not account for all confounding factors for example pulmonary hypertension. Small sample size and single centre study – lacks generalisability.
Naji 2006[80]	People initially admitted to hospital for 3 days for baseline assessments and to commence on the programme. The programme consisted of exercise and education was continued post discharge 2 times per	ILD n=19	Shuttle test (m) CRDQ (dyspnoea) QoL (SGRQ)	Some figures related to dropouts and survival data doesn't add up correctly. Not clearly reported. Small sample size and single centre study – lacks generalisability High dropout rate.

Study	Intervention/ comparison	Population	Outcomes	Comments
	week over a period of 8 weeks.			
Nishiyama 2008[88] and Holland 2008[42] (Cochrane review from which additional data was extracted)	9 weeks supervised exercise programme/control group not specified	IPF n=28	Change in 6MWT immediately following training Change in dyspnoea score immediately following training Change in quality of life immediately following training QoL (SGRQ domains)	Blinding of investigators not reported. Sequence generation unclear. Selective reporting may be a problem, due to insufficient data it is not possible to determine if all data was made available.
Ozalevli 2010[96]	Home based pulmonary rehabilitation programme lasting 12 weeks/no control group	IPF n=17	6MWD Dyspnoea (MRC scale) QoL (SF-36)	Did not account for confounding factors. No control group. Small sample size and single centre study – lacks generalisability.
Rammaert 2011[104]	Home-based pulmonary rehabilitation for 8 weeks lasting 30-45 minutes per day/no control group	IPF n-17	6MWT Dyspnoea (MRC scale, Borg scale) QoL(Visual Analogue Scale)	Confounding factors were not accounted for. Large number of drop outs – 41%. No comparison group. Small sample size and single centre study – lacks generalisability.
Swigris 2011[124]	6-8 week pulmonary rehabilitation programme consisting of 18 sessions/no control group	IPF n=21	6MWD Anxiety (General anxiety questionnaire) Depression (Patient health questionnaire)	Small sample size. Substantial proportion of drop outs. PR was paid for through people's insurance therefore may be a highly motivated group. COPD control group.

8.3.2 Study quality and summary of findings

Table 33: Clinical evidence profile: pulmonary rehabilitation versus no pulmonary rehabilitation – randomised controlled trials

Quality assessment							No of people		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	No pulmonary rehab.	Relative (95% CI)	Absolute	
Six month survival, Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	Very serious[6]	None	2/20 (10%)	2/14 (14.3%)	RR 0.7 (0.09 to 3.31)	43 fewer per 1000 (from 130 fewer to 330 more)	Very low
Change in 6MWD immediately following training (m) (better indicated by higher values), Gaunaurd2011,[36] Holland 2008,[41,42] Nishiyama 2008[88]											
3	Randomised trials	Very serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Serious [5,10]	None	36	32	Not applicable	MD 30.19 higher (7.25 to 53.12 higher)	Very low
Change in 6MWD at long-term follow-up. Mean change from baseline (m) (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9]	None	20	14	Not applicable	MD 23.08 lower (70.59 lower to 24.43 higher)	Moderate
Change in dyspnoea score immediately following training (better indicated by lower values), Holland 2008,[41,42] Nishiyama 2008[88]											
2	Randomised trials	Serious[1,2,3]	No serious inconsistency	No serious indirectness	Very serious[12]	None	33	29	Not applicable	SMD 0.43 lower (0.94 lower to	Very low

Quality assessment							No of people		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	No pulmonary rehab.	Relative (95% CI)	Absolute	
										0.08 higher)	
Change in dyspnoea score at long-term follow-up (better indicated by lower values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	Very serious[6,11]	None	20	14	Not applicable	MD 0.01 higher (0.79 lower to 0.81 higher)	Very low
Change in quality of life immediately following training (better indicated by higher values), Holland 2008,[41,42] Nishiyama 2008[88]											
2	Randomised trials	Serious[1,2,3]	No serious inconsistency	No serious indirectness	Very serious[12]	None	33	29	Not applicable	SMD 0.57 higher (0.06 to 1.09 higher)	Very low
Change in quality of life at long-term follow-up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	Very serious [6,8,11]	None	20	14	Not applicable	MD 7.05 higher (8.29 lower to 22.39 higher)	Very low
QoL: SF36 domain: physical functioning immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 8.89 higher (2.74 lower to 20.52 higher)	Moderate

Quality assessment							No of people		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Pulmonary rehab.**	**No pulmonary rehab.**	**Relative (95% CI)**	**Absolute**	
QoL: SF36 domain: bodily pain immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 7.29 higher (7.93 lower to 22.51 higher)	Moderate
QoL: SF36 domain: physical role functioning immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 0.93 lower (20.72 lower to 18.86 higher)	Moderate
QoL: SF36 domain: general health perceptions immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 2.25 higher (7.48 lower to 11.98 higher)	Moderate
QoL: SF36 domain: vitality immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 10.27 higher (0.12 lower to 20.66 higher)	Moderate
QoL: SF36 domain: social role functioning immediately following training (better indicated by higher values), Holland 2008[41,42]											

Quality assessment							No of people		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Pulmonary rehab.**	**No pulmonary rehab.**	**Relative (95% CI)**	**Absolute**	
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 3.24 higher (10.98 lower to 17.46 higher)	Moderate
QoL: SF36 domain: emotional role functioning immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 0 higher (22.58 lower to 22.58 higher)	Moderate
QoL: SF36 domain: mental health immediately following training (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 13.96 higher (3.88 to 24.04 higher)	Moderate
QoL: SF36 domain: physical functioning at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 7.59 higher (4.11 lower to 19.29 higher)	Moderate
QoL: SF36 domain: physical role functioning at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 5.56 lower	Moderate

Quality assessment							No of people		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	No pulmonary rehab.	Relative (95% CI)	Absolute	
										(22.09 lower to 10.97 higher)	
QoL: SF36 domain: bodily pain at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 8 higher (8.53 lower to 24.53 higher)	Moderate
QoL: SF36 domain: mental health at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 11.29 higher (1.46 to 21.12 higher)	Moderate
QoL: SF36 domain: vitality at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 3.9 higher (7.14 lower to 14.94 higher)	Moderate
QoL: SF36 domain: general health perceptions at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 4.81 higher (7.07 lower to	Moderate

Quality assessment							No of people		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Pulmonary rehab.**	**No pulmonary rehab.**	**Relative (95% CI)**	**Absolute**	
										16.69 higher)	
QoL: SF36 domain: social role functioning at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 0.93 lower (16 lower to 14.14 higher)	Moderate
QoL: SF36 domain: emotional role functioning at long term follow up (better indicated by higher values), Holland 2008[41,42]											
1	Randomised trial	Serious[2,3]	Not applicable	No serious indirectness	No serious imprecision[9,10]	None	30	27	Not applicable	MD 11.11 lower (34.33 lower to 12.11 higher)	Moderate

[1] Nishiyama 2008[88] did not report adequate sequence generation or blinding
[2] High dropout rate
[3] Small sample size
[4] Abstract
[5] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID
[6] Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID
[8] Imprecision for this QoL outcome was assessed using the default MID as no established MID was found for a total SF36 score.
[9] Imprecision for this outcome was assessed using the established MID, as only one study contributed data the standardised mean difference was used to assess the imprecision however only the mean difference is reported here.
[10] Imprecision for these QoL: SF36 domains and 6MWT distance were assessed using established MIDs (see the methodology chapter for further details).
[11] Imprecision for this outcomes was assessed using the default MID, as only one study contributed data the standardised mean difference was used to assess the imprecision however only the mean difference is reported here.
[12]Imprecision could not be calculated because different scales for dyspnoea and QoL scores were used by the studies.

Table 34: Clinical evidence profile: pulmonary rehabilitation versus no pulmonary rehabilitation – observational studies

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
Almoamary 2012[5]										
6MWD (m) (better indicated by higher values)										
1	Observational study	Serious[1,2]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	179 ±74	293 ±97	Very low
Distance on treadmill (m) (better indicated by higher values)										
1	Observational study	Serious[1,2,3]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	114±66	371±199	Very low
Distance on bicycle (m) (better indicated by higher values)										
1	Observational study	Serious[1,2,3]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	1031 ± 358	2532± 1120	Very low
Distance on ergometer (m) (better indicated by higher values)										
1	Observational study	Serious[1,2,3]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	555±136	1238 ±522	Very low
Emergency department visits (no.) (better indicated by lower values)										
1	Observational study	Serious[1,2,3]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	1.3 ±1.9	0.6 ±0.9	Very low
Outpatient department visits (days) (better indicated by lower values)										
1	Observational study	Serious[1,2,3]	Not applicable	No serious indirectness	Could not be calculated	None	21 ILD	4.7 ± 2.7	2.7±0.6	Very low
Ferreira 2009[30]										
Dyspnoea (Borg score) (better indicated by lower values)										

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
1	Observational study	Serious[1,2,3,6]	Not applicable	No serious indirectness	Could not be calculated	None	99 ILD	3.6 ±2.0	2.7 ±1.7	Very low
Dyspnoea (UCSD questionnaire) (better indicated by lower values)										
1	Observational study	Very serious[1,2,3,6,16]	Not applicable	No serious indirectness	Could not be calculated	None	29 ILD	57.4±25	49.1 ±25	Very low
6MWD (m) (better indicated by higher values)										
1	Observational study	Very serious[1,2,3,6,17]	Not applicable	No serious indirectness	Could not be calculated	None	99 ILD	335 ±131	391 ±118	Very low
Depression (CES-D score) (better indicated by lower values)										
1	Observational study	Very serious[1,2,3,6,16]	Not applicable	No serious indirectness	Could not be calculated	None	27 ILD	15.7 ±8	13.6 ±8	Very low
6MWD (m) (% change)										
1	Observational study	Very serious[1,2,3,6,17]	Not applicable	No serious indirectness	Could not be calculated	None	99 ILD	Median (25th percentile, 75th percentile): 14 (2, 33)		Very low
Holland 2012[44]										
Dyspnoea: Change in CRQ dyspnoea domain at 8 weeks										
1	Observational study	Serious[2,3,5]	Not applicable	No serious indirectness	Could not be calculated	None	25 IPF	NR	2.7 ±5.6	Very low
Dyspnoea: Change in CRQ dyspnoea domain at 6 months										
1	Observational study	Serious[2,3,5]	Not applicable	No serious indirectness	Could not be calculated	None	25 IPF	NR	Reported: "Non-significant change from	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
									baseline"	
Change in 6MWD (m) at 8 weeks										
1	Observational study	Serious[2,3,5]	Not applicable	No serious indirectness	Could not be calculated	None	25 IPF	NR	21 ±58	Very low
Change in 6MWD (m) at 6 months										
1	Observational study	Serious[2,3,5]	Not applicable	No serious indirectness	Could not be calculated	None	25 IPF	NR	Reported: "Non-significant change from baseline"	Very low
Jastrzebski 2006[55]										
Dyspnoea (MRC scale) (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	2.3±0.8	2.0±0.9	Very low
Dyspnoea (oxygen cost diagram) (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	72.2± 14.6	77.2±15.9	Very low
Dyspnoea (BDI) (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	6.3±2.8	6.8±3.3	Very low
Dyspnoea (Borg scale) (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	3.0±1.4	2.5±1.4	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
QoL: SF36 domain: physical functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	55	65	Very low
QoL: SF36 domain: physical role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	40	55	Very low
QoL: SF36 domain: vitality (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	53	58	Very low
QoL: SF36 domain: bodily pain (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	69	67	Very low
QoL: SF36 domain: general health perceptions (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	38	41	Very low
QoL: SF36 domain: social role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	58	70	Very low
QoL: SF36 domain: emotional role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	69	80	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
QoL: SF36 domain: mental health (better indicated by higher values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	62	68	Very low
QoL: SGRQ domains: symptoms (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	45	46	Very low
QoL: SGRQ domains: activity (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	52	45	Very low
QoL: SGRQ domains: influence (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	47	37	Very low
QoL: SGRQ total domains (better indicated by lower values)										
1	Observational study	Serious[2,3,4,5,6,14]	Not applicable	No serious indirectness	Could not be calculated	None	38 ILD	47	42	Very low
Kozu 2011[65]										
Dyspnoea (MRC scale) (better indicated by lower values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	3.0±0.8	2.5±1.1 6 months: 2.9±1	Very low
6MWD (m) (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be	None	45 IPF	323±109	340±122 6 months:	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
					calculated				320±106	
QoL: SF36 domain: physical functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	38.6±19	40.6±22.6 6 months: 37.8±23	Very low
QoL: SF36 domain: physical role functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	34.9±21.5	35.9±20.7 6 months: 30.4±23.7	Very low
QoL: SF36 domain: vitality (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	43.1±20	43.9±21 6 months: 42.1±23.6	Very low
QoL: SF36 domain: bodily pain (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	66.1±30	63.4±28.1 6 months: 62.5±30.3	Very low
QoL: SF36 domain: general health perceptions (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	37.1±20	36.9±21.1 6 months: 34.4±21.5	Very low
QoL: SF36 domain: social role functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	51±23.8	50.3±25.3 6 months: 45.8±26.9	Very low
QoL: SF36 domain: emotional role functioning (better indicated by higher values)										
1	Observational	Very	Not	No serious	Could not	None	45 IPF	39.6±30.7	38.7±31.3	Very

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
	study	serious[2,3,5,7,8,9]	applicable	indirectness	be calculated				6 months: 35.8±29.8	low
QoL: SF36 domain: mental health (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,7,8,9]	Not applicable	No serious indirectness	Could not be calculated	None	45 IPF	50.7±18.7	52.6±-20.5 6 months 47.5±21.8	Very low
Naji 2006[80]										
Shuttle test (m) (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,6,15]	Not applicable	No serious indirectness	Could not be calculated	None	19 ILD	171±102	232±118	Very low
Dyspnoea (CRDQ) (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,6,15]	Not applicable	No serious indirectness	Could not be calculated	None	19 ILD	Median (ranges):15.6 (9.7, 22.6)	Median (ranges): 17.2(14.6, 27.1)	Very low
QoL: SGRQ total (better indicated by lower values)										
1	Observational study	Very serious[2,3,5,6,15]	Not applicable	No serious indirectness	Could not be calculated	None	19 ILD	Median (ranges):48.1 (23, 82)	Median (ranges): 26.4(17.4, 69.4)	Very low
Ozalevli 2010[96]										
6MWD (m) (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	390.3	430.5	Very low
Dyspnoea (MRC scale) (better indicated by lower values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	2.3±1.2	1.4±1.3	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
QoL: SF36 domain: physical functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	57.00±5.7	58.7±7.3	Very low
QoL: SF36 domain: physical role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	56.00±1.7	68.3±1.6	Very low
QoL: SF36 domain: vitality (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	52.00±4.9	55±4.2	Very low
QoL: SF36 domain: bodily pain (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	25.00±2.6	72±2.2	Very low
QoL: SF36 domain: general health perceptions (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	67.30±4.6	74±4.7	Very low
QoL: SF36 domain: social role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	75.80±2.7	89.1±1.8	Very low
QoL: SF36 domain: emotional role functioning (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	29.00±1.3	65±1.4	Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
QoL: SF36 domain: mental health (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	49.90±6.7	56.8±5.4	Very low
Rammaert 2011[104]										
6MWD (m) (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6,10]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	383±115	375±101	Very low
Dyspnoea (MRC scale) (better indicated by lower values)										
1	Observational study	Serious[2,3,5,6,10]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	Median (range):1.5 (1-3)	Median (range):2 (1-3)	Very low
Dyspnoea (Borg scale) (better indicated by lower values)										
1	Observational study	Serious[2,3,5,6,10]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	Median (range):4 (2-8)	Median (range):3 (2-9)	Very low
QoL: Visual Analogue Scale (total) (better indicated by higher values)										
1	Observational study	Serious[2,3,5,6,10]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	38±8	42±12	Very low
QoL (SF-36, SGRQ & HAD)										
1	Observational study	Serious[2,3,5,6,10]	Not applicable	No serious indirectness	Could not be calculated	None	17 IPF	Reported: “Perceived physical limitation during exercise as described in the SF-36 decreased after PR (P=0.047). No significant differences were observed for the other SF-36 parameters, the SGRQ or the hospital anxiety and		Very low

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
								depression (HAD) scale.		
Swigris 2011[124]										
6MWD (feet) (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	906±111	1108±164	Very low
Anxiety (general anxiety disorder 7) (better indicated by lower values)										
1	Observational study	Very serious[2,3,5,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	2.7±0.8	1.3±0.5	Very low
QoL: Patient Health Questionnaire 8(better indicated by lower values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	3.4±0.0	2.5±0.7	Very low
QoL: SF36 domain: physical functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	31.9±2.4	33.1±2.8	Very low
QoL: SF36 domain: physical role functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	36.4±2.3	38±2.8	Very low
QoL: SF36 domain: vitality (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	47.2±2.2	50.8±2.6	Very low
QoL: SF36 domain: bodily pain (better indicated by higher values)										
1	Observational	Very	Not	No serious	Could not	None	21 IPF	45±2.2	47.6±2.7	Very

Quality assessment							No of people	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Pulmonary rehab.	Baseline	Post treatment	
	study	serious[2,3,5,9,11,12]	applicable	indirectness	be calculated					low
QoL: SF36 domain: general health perceptions (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	38.3±1.7	39.8±2.9	Very low
QoL: SF36 domain: social role functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	45.1±2	47.1±3	Very low
QoL: SF36 domain: emotional role functioning (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	45.7±2.6	43.8±4	Very low
QoL: SF36 domain: mental health (better indicated by higher values)										
1	Observational study	Very serious[2,3,5,9,11,12]	Not applicable	No serious indirectness	Could not be calculated	None	21 IPF	51.8±2	53.3±1.4	Very low

[1] *Retrospective design- biases.*
[2] *Does not account for confounding factors.*
[3] *No blinding of investigators and non-randomised.*
[4] *No baseline data provided.*
[5] *Small sample size and concerns over generalisability.*
[6] *No control/comparison group.*
[7] *Large number of drop outs 20% drop out rate in IPF group (not including follow up period).*
[8] *Inconsistencies in reporting some data (when comparing IPF performance with COPD).*
[9] *Control group composed of COPD people not IPF people receiving usual care.*
[10] *Large number of drop outs - 41%.*
[11] *Large number of drop outs - 33%.*
[12] *PR was paid for through people's insurance therefore were a highly motivated group.*
[14] *Data taken from a graph.*

[15] Large number of drop outs – 46% and unclear reporting of dropout data at different follow up points.
[16] Single centre contributed data for this outcome, therefore there are concerns over generalisability.
[17] Differences between the participating centres with the amount of oxygen given to people during the 6MWT.
Note: where imprecision could not be calculated, this is because observational studies only provided individual domain scores and not total scores on which MIDs had been agreed.

Table 35: Summary of findings of cohort studies

Reference	Outcome	Baseline (mean ± SD)	After pulmonary rehabilitation (mean ± SD)	After 6 months (mean ± SD)
Almoamary 2012[5]	6MWD (m)	179 ±74	293 ±97	NR
	Distance on treadmill (m)	114±66	371±199	NR
	Distance on bicycle (m)	1031 ± 358	2532±1120	NR
	Distance on ergometer (m)	555±136	1238 ±522	NR
	Emergency department visits (no.)	1.3 ±1.9	0.6 ±0.9	NR
	Outpatient department visits (days)	4.7 ± 2.7	2.7±0.6	NR
Ferreira 2009[30]	Dyspnoea (Borg score)	3.6± 2.0	2.7 ±1.7	NR
	Dyspnoea (UCSD questionnaire)	57.4 ±25	49.1 ±25	NR
	6MWT distance (m)	335 ±131	391 ±118	NR
	Depression (CES-D score)	15.7 ±8	13.6 ±8	NR
	6MWT distance, % change (n =99) Median (25th percentile, 75th percentile).	NR	Change: 14 (2, 33) P: 0.002	NR
Holland 2012[44]	Dyspnoea: Change in CRQ dyspnoea domain	NR	2.7 ±5.6 Reported: "significantly improved from baseline p<0.5"	Reported: "Non-significant change from baseline"
	6MWD (m)	NR	21 ±58 Reported: "significantly improved from baseline p<0.5"	Reported: "Non-significant change from baseline"
Jastrzebski 2006[55]	Dyspnoea (MRC scale)	2.3±0.8	2.0±0.9	NR
	Dyspnoea (oxygen cost diagram)	72.2±14.6	77.2±15.9	NR
	Dyspnoea (BDI)	6.3±2.8	6.8±3.3	NR
	Dyspnoea (Borg scale)	3.0±1.4	2.5±1.4	NR
	QoL: SF36 domain: physical	31.9±2.4	33.1±2.8	NR

Reference	Outcome	Baseline (mean ± SD)	After pulmonary rehabilitation (mean ± SD)	After 6 months (mean ± SD)
	functioning			
	QoL: SF36 domain: physical role functioning	36.4±2.3	38±2.8	NR
	QoL: SF36 domain: vitality	47.2±2.2	50.8±2.6	NR
	QoL: SF36 domain: bodily pain	45±2.2	47.6±2.7	NR
	QoL: SF36 domain: general health perceptions	38.3±1.7	39.8±2.9	NR
	QoL: SF36 domain: social role functioning	45.1±2	47.1±3	NR
	QoL: SF36 domain: emotional role functioning	45.7±2.6	43.8±4	NR
	QoL: SF36 domain: mental health	51.8±2	53.3±1.4	NR
Kozu 2011[65]	Dyspnoea (MRC scale)	3.0±0.8	2.5±1.1	2.9±1
	6MWD (m)	323±109	340±122	320±106
	QoL: SF36 domain: physical functioning	38.6±19	40.6±22.6	37.8±23
	QoL: SF36 domain: physical role functioning	34.9±21.5	35.9±20.7	30.4±23.7
	QoL: SF36 domain: vitality	43.1±20	43.9±21	42.1±23.6
	QoL: SF36 domain: bodily pain	66.1±30	63.4±28.1	62.5±30.3
	QoL: SF36 domain: general health perceptions	37.1±20	36.9±21.1	34.4±21.5
	QoL: SF36 domain: social role functioning	51±23.8	50.3±25.3	45.8±26.9
	QoL: SF36 domain: emotional role functioning	39.6±30.7	38.7±31.3	35.8±29.8
	QoL: SF36 domain: mental health	50.7±18.7	52.6±-20.5	47.5±21.8
Naji 2006[80]	Shuttle test (m)	171±102	232±118	NR
	Dyspnoea (CRDQ) Median (ranges)	15.6 (9.7, 22.6)	17.2(14.6, 27.1)	NR

Reference	Outcome	Baseline (mean ± SD)	After pulmonary rehabilitation (mean ± SD)	After 6 months (mean ± SD)
	QoL (SGRQ) Median (ranges)	48.1 (23, 82)	26.4(17.4, 69.4)	NR
Ozalevli 2010[96]	6MWD (m)	390.3	430.5	NR
	Dyspnoea (MRC scale)	2.3±1.2	1.4±1.3	NR
	QoL: SF36 domain: physical functioning	57.00±5.7	58.7±7.3	NR
	QoL: SF36 domain: physical role functioning	56.00±1.7	68.3±1.6	NR
	QoL: SF36 domain: vitality	52.00±4.9	55±4.2	NR
	QoL: SF36 domain: bodily pain	25.00±2.6	72±2.2	NR
	QoL: SF36 domain: general health perceptions	67.30±4.6	74±4.7	NR
	QoL: SF36 domain: social role functioning	75.80±2.7	89.1±1.8	NR
	QoL: SF36 domain: emotional role functioning	29.00±1.3	65±1.4	NR
	QoL: SF36 domain: mental health	49.90±6.7	56.8±5.4	NR
Rammaert 2011[104]	6MWD (m)	383±115	375±101	NR
	Dyspnoea (MRC scale)	1.5 (1-3)	2 (1-3)	NR
	Dyspnoea (Borg scale)	4 (2-8)	3 (2-9)	NR
	QoL: VAS (total)	38±8	42±12	NR
	QoL (SF-36, SGRQ & HAD)	NR	Reported: “perceived physical limitation during exercise as described in the SF-36 decreased after PR (P=0.047). No significant differences were observed for the other SF-36 parameters, the SGRQ or the HAD scale”.	NR
Swigris 2011[124]	6MWD (feet)	906±111	1108±164	NR
	General anxiety disorder 7	2.7±0.8	1.3±0.5	NR

Reference	Outcome	Baseline (mean ± SD)	After pulmonary rehabilitation (mean ± SD)	After 6 months (mean ± SD)
	Patient Health Questionnaire 8	3.4±0.0	2.5±0.7	NR
	QoL: SF36 domain: physical functioning	31.9±2.4	33.1±2.8	NR
	QoL: SF36 domain: physical role functioning	36.4±2.3	38±2.8	NR
	QoL: SF36 domain: vitality	47.2±2.2	50.8±2.6	NR
	QoL: SF36 domain: bodily pain	45±2.2	47.6±2.7	NR
	QoL: SF36 domain: general health perceptions	38.3±1.7	39.8±2.9	NR
	QoL: SF36 domain: social role functioning	45.1±2	47.1±3	NR
	QoL: SF36 domain: emotional role functioning	45.7±2.6	43.8±4	NR
	QoL: SF36 domain: mental health	51.8±2	53.3±1.4	NR

8.4 Economic evidence

Published literature

No relevant economic evaluations that assessed pulmonary rehabilitation in an IPF population were identified. However, due to variation in practice and uncertainty surrounding the cost effectiveness of pulmonary rehabilitation programmes for people with IPF, an economic evaluation was conducted. The results of which are summarised in the below economic evidence profile and in Appendix L, and the full report is detailed in Appendix L. No studies were selectively excluded.

Table 36: Economic evidence profile: no rehabilitation versus rehabilitation with exercise component only versus rehabilitation with exercise and educational components.

Study	Applicability	Limitations	Other comments	Total cost per patient	Total effects (QALY)	Cost effectiveness NMB at £20,000 (£30,000)	Uncertainty
NCGC economic model	Directly applicable (a)	Potentially serious limitations (b)	Markov decision analytical model.	1.No rehabilitation: £0 2.Rehabilitation with exercise component only: £678 3.Rehabilitation with exercise and educational components: £770	1.No rehabilitation: 2.474 2.Rehabilitation with exercise component only: 2.573 3.Rehabilitation with exercise and educational components: 2.559	1.No rehabilitation: £49,480 (£74,220) 2.Rehabilitation with exercise component only: £50,785 (£76,453) 3.Rehabilitation with exercise and educational components: £50,413 (£75,933) (dominated by rehabilitation with exercise) Incremental cost effectiveness ratio of rehabilitation with exercise only compared to no rehabilitation = £6841	In the PSA the strategies had the following probabilities of being optimal at the £20,000 (£30,000) threshold: 1.No rehabilitation: 5% (4%) 2.Rehabilitation with exercise component only: 48% (48%) 3. Rehabilitation with exercise and educational components: 47% (48%) These results show that when uncertainty around the mean of model inputs is taken into account, both types of programmes have similar probability of being cost effective. This result is driven by the wider uncertainty surrounding the potential QoL gain of the exercise and educational programme. A variety of deterministic sensitivity analyses showed the results to be robust under a number of different input and structural assumptions. A 3 way sensitivity analysis showed that optimal time period between programmes was sensitive to assumptions regarding treatment effect duration and magnitude of effect on repeated offers. Under plausible assumptions, pulmonary rehabilitation was shown to be cost effective if repeated every 6 to 12 months.

(e) From UK perspective with use of NHS published costs.

(f) Treatment effect from two published RCTs; EQ5D values mapped from the SF36; FVC% predicted as a marker for disease progression in the IPF population, and as a proxy for ability to participate and benefit from the pulmonary rehabilitation programme; rate of disease progression assumed linear; limited sources informing transition probabilities between health states; no account of possible reduction of healthcare contacts (due to no evidence to inform this parameter). Treatment effect taken from moderate to very low quality evidence (Holland 2008[41,42], Nishiyama 2008[88]). PSA undertaken to explore uncertainty.

8.5 Evidence statements

Clinical

RCT data

Performance on sub-maximal walk test

Very low quality evidence showed that PR may be clinically effective at increasing 6 minute walk distance immediately following training compared to those who did not undertake PR [3 studies N = 68].

Moderate quality evidence showed that there was no clinically effective difference in the 6 minute walk distance at long term follow up between people who undertook a PR programme compared to those who didn't [1 study n=34].

Dyspnoea

Very low quality evidence showed that PR may be clinically effective at reducing levels of dyspnoea immediately following training compared to those who did not undertake PR, however imprecision could not be calculated for this outcome [2 studies n=62].

Very low quality evidence showed that PR may be clinically effective at reducing levels of dyspnoea at long term follow up compared to those who did not undertake PR, but the direction of the estimate of the effect could favour either intervention [1 study n=34].

Survival rate

Very low quality evidence showed that PR may be clinically effective at improving six month survival rates compared to those who did not undertake PR , but the direction of the estimate of the effect could favour either intervention [1 study n=34].

Health-related quality of life

Very low quality evidence showed that PR may be clinically effective at improving quality of life scores on the SF36 and SGRQ immediately following training compared to those who did not undertake PR, however imprecision could not be calculated for this outcome [2 studies n=62].

Very low quality evidence showed that PR may be clinically effective at improving quality of life scores on the SF36 at long term follow up compared to those who did not undertake PR, but the direction of the estimate of the effect could favour either intervention [1 study n=34].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain physical functioning immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain bodily pain immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain physical role functioning immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain general health perceptions immediately following training, between people who undertook a PR programme compared to those who did not [1 studies =57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain vitality immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain social role functioning immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain emotional role functioning immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain mental health immediately following training, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain physical functioning at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain physical role functioning at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain bodily pain at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain mental health at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain vitality at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain general health perceptions at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain social role functioning at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Moderate quality evidence showed that there was no clinically effective difference in the SF-36 quality of life score for the domain emotional role functioning at long term follow up, between people who undertook a PR programme compared to those who did not [1 studies n=57].

Observational data

Psychosocial health

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of depression (CED-D score) [1 retrospective study n=27 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of anxiety (general anxiety disorder 7) [1 prospective study n=21 IPF people].

Dyspnoea

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (Borg score and UCSDQ) [1 retrospective study n=99 (Borg score) and 29 (UCSDQ) ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (change in CRQ domain) at 8 weeks following PR but is not sustained at 6 months follow up[1 retrospective study n=25 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen levels of dyspnoea (BDI score) [1 prospective study n=38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen levels of dyspnoea (MRC scale score) and does not improve at 6 months follow up [1 prospective study n=45 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (CRDQ) [1 retrospective study n=19 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (oxygen cost diagram score, MRC scale and Borg scale) [1 prospective study =38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (MRC scale) [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen levels of dyspnoea (MRC scale) [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce levels of dyspnoea (Borg scale) [1 prospective study n=17 IPF people].

Performance on sub-maximal walk test

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve distance walked (m) in sub maximal exercise testing (6MWT, treadmill, bicycle and ergometer) in people undertaking PR [1 retrospective study n=21 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve 6MWD (m) [1 retrospective study n=99 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve 6MWD (m) 8 weeks following PR but is not sustained at 6 months follow up [1 retrospective study n=25 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve 6MWD which is not maintained at 6 month follow up [1 prospective study n=45 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve walking distance (m) (shuttle walk test) [1 retrospective study n=19 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve 6MWD [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen 6MWD [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve 6MWD [1 prospective study n=21 IPF people].

Health-related quality of life

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores in the following SF-36 domains: physical functioning, physical role functioning, vitality, general health perceptions, social role functioning, emotional role functioning and mental health [1 prospective study n=38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen the QoL scores in the SF-36 domain bodily pain [1 prospective study n=38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores in all SGRQ domains; activity, influence, and total domains [1 prospective study n=38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen the QoL score in the SGRQ domain symptoms [1 prospective study n=38 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores in the following SF-36 domains; physical functioning, physical role functioning, vitality and mental health however this is not maintained at 6 months follow up for any domain listed [1 prospective study n=45 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen the QoL scores in the SF-36 domain bodily pain, general health perceptions social role functioning and emotional role functioning and does not improve at 6 months follow up [1 prospective study n=45 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL score in the SGRQ total score [1 retrospective study n=19 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores in all the SF-36 domains including; physical functioning, physical role functioning, bodily pain, vitality, general health perceptions, social role functioning, emotional role functioning and mental health [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL (visual analogue scale) [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores related to physical limitations in the SF-36 but showed no difference in QoL scores in other SF-36 domains, SGRQ and HAD scale [1 prospective study n=17 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve QoL scores (patient health questionnaire) [1 prospective study n=21 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may improve the QoL scores in the SF-36 domains; physical functioning, physical role functioning, bodily pain, vitality, general health perceptions, social role functioning, and mental health [1 prospective study n=21 IPF people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may worsen the QoL scores in the SF-36 emotional role functioning [1 prospective study n=21 IPF people].

Resource use

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce the number of emergency department visits [1 retrospective study n=21 ILD people].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that PR may reduce the number of outpatient department visits (days) [1 retrospective study n=21 ILD people].

Optimal course content, setting and duration

No clinical evidence was found addressing what is the optimal course content, setting and duration for people referred for pulmonary rehab programmes are.

Economic

No published health economic studies were identified to aid consideration of cost effectiveness.

It is highly likely that pulmonary rehabilitation is cost effective as a means to improve quality of life for people with IPF. This is based on evidence of direct applicability and with potentially serious limitations.

It is uncertain whether pulmonary rehabilitation with exercise alone is cost effective when compared to a programme with an educational component. Both programmes are highly likely to be cost effective when compared to no rehabilitation. This is based on evidence of direct applicability and with potentially serious limitations.

Pulmonary rehabilitation could be cost effective if offered at 6 to 12 month intervals to people with IPF, given appropriate assessment of the patient prior to the programme. If the duration of long term effect is shorter in the exercise programme than the educational programme, it is likely it is more cost effective to repeat this component of pulmonary rehabilitation in shorter time intervals (i.e. 6 months) than an educational component (that is, 12 months or more). This is based on evidence of direct applicability and with potentially serious limitations.

8.6 Recommendations and link to evidence

Recommendations	**14. Assess people with idiopathic pulmonary fibrosis for pulmonary rehabilitation at the time of diagnosis. Assessment may include a 6-minute walk test (distance walked and oxygen saturation measured by pulse oximetry) and a quality-of-life assessment.** **15. Repeat the assessment for pulmonary rehabilitation for people with idiopathic pulmonary fibrosis at 6-month or 12-month intervals.** **16. If appropriate after each assessment, offer pulmonary rehabilitation including exercise and educational components tailored to the needs of people with idiopathic pulmonary fibrosis in general.** **17. Pulmonary rehabilitation should be tailored to the individual needs of each person with idiopathic pulmonary fibrosis. Sessions should be held somewhere that is easy for people with idiopathic pulmonary fibrosis to get to and has good access for people with disabilities.**
Relative values of different outcomes	Mortality and survival where considered to be the critical outcomes. As none of these outcomes were reported by studies included in this review, the GDG considered capacity measured through 6MWD, dyspnoea and QOL to be the most important outcomes to inform decision making. The GDG noted that quality of life scored encompassed a variety of domains, and expected the impact of other important outcomes such as fatigue to manifest in changes in quality of life scores. The GDG recognised that muscle capacity is a limitation associated with these outcomes, as muscle capacity varies between individuals due to exercise capacity and body mass.
Trade-off between clinical benefits and harms	The GDG discussed the potential benefits of pulmonary rehabilitation for improving all health related quality measures in people with IPF and stressed the importance of early and repeated assessments for pulmonary rehabilitation. The GDG considered exercise exertion beyond a person's normal capacity to be the only harm associated with pulmonary rehabilitation, but that this risk was unlikely due to the programmes being conducted by trained health professionals (physiotherapists and nurses). When assessing a person's suitability for pulmonary rehabilitation, their oxygen saturation profile considered so that severe desaturation is avoided in during the pulmonary rehabilitation class. Submaximal and endurance test may be appropriate to ascertain the patient's needs when undertaking the course. It was agreed that the assessment should occur in a hospital setting so that in the unlikelihood of over exertion, appropriate emergency care would be available.
Economic considerations	No health economic evidence was identified to inform this question. Currently in the UK there is much variation in practice in the offer and the content of pulmonary rehabilitation programmes for IPF people. Pulmonary rehabilitation designed and provided specifically for the IPF population is not known to exist in the current UK setting. Either people are not offered pulmonary rehabilitation, or are offered places on pulmonary rehabilitation courses designed for people with Chronic Obstructive Pulmonary Disease (COPD). Content in programmes designed for COPD may be inappropriate for an IPF population, and a greater health benefit may be

realised with courses tailored to the IPF population. Stakeholders and the GDG thought it likely that pulmonary rehabilitation is underutilised as a means of improving quality of life in people who live with IPF, including both people and carers. An educational component of pulmonary rehabilitation educates people how to self-manage symptoms of IPF and could prevent unnecessary contact with the National Health Service and therefore could reduce costs. However, as pulmonary rehabilitation is not widely offered to people with IPF, a recommendation to offer pulmonary rehabilitation routinely would come at additional cost, especially if the course were tailored specifically to the IPF population and offered on a frequent basis.

As no published economic evidence was identified to assess the cost effectiveness of such programmes in the IPF population specifically, the GDG considered it was appropriate to prioritise this topic area for an economic model. The economic model compared three strategies of no rehabilitation, rehabilitation with only an exercise component and rehabilitation with an exercise and educational component. The content and associated resource use of the rehabilitation components, as specified by two RCTs included in the evidence review, informed the model. The implementation of rehabilitation programmes may require additional staff or a change in skills, so qualification costs were included in the unit cost used to estimate the cost of staff time.

In the context of limited evidence, the GDG wished to explore thresholds and relationships between the trade-offs that exist in the decision problem, especially those relating to non-participation or ability to participate due to hospitalisation, disease progression and/or death, treatment effect in terms of duration and magnitude; and the cost effectiveness of repeating the programme more than once. The model was therefore designed to explore these factors in relation to incremental differences of cost and effect for decision making rather than to produce an accurate tally of the lifetime QALYs and costs that may accrue across the lifetime of a person with IPF. The results were interpreted with this limitation in mind. Other key limitations identified and acknowledged to potentially influence results were simplifications: made in modelling the natural course of IPF progression and use of prediction scores derived from a clinical trial cohort that would not necessarily be reflective of the UK population with IPF, the limited data sources available for predicting respiratory hospitalisation, limitations of quality of life score measures and mapping functions and the use of FVC % predicted as a marker for disease progression and proxy to determine who could participate and benefit from pulmonary rehabilitation. For the base-case, the GDG decided the most conservative assumptions in favour of no rehabilitation should be made when examining whether rehabilitation programmes were cost effective in the IPF population, and as far as possible all assumptions should be explored in a sensitivity analysis.

There is currently no evidence examining the potential decrease in the number of healthcare contacts made by a patient undergoing rehabilitation, and therefore this parameter was not explored in the model. However, the GDG acknowledged that if the number of healthcare contacts did decrease with rehabilitation to a significant extent, the cost effectiveness of rehabilitation would improve further, and may even become a cost saving intervention.

It was noted that as assessment is undertaken as a one to one consultation, the cost of staff time per patient is higher for this component of rehabilitation than it is for a rehabilitation class, whereby two members of staff will be responsible for class sizes from 10 to 30 patients. The assessment cost therefore comprises on nearly half of the total cost per patient undertaking a pulmonary rehabilitation course. However, the assessment for pulmonary rehabilitation was considered to be important not only to determine potential safety concerns for patients who may undertake rehabilitation in

	the community, but also in determining which people are most likely to benefit from pulmonary rehabilitation. An important factor to consider at assessment is whether a patient is likely to be able to benefit from the rehabilitation programme after the programme ends, and whether there is likely to be added value if the patient is being offered a programme for a second or third time, as both of these factors were found to influence the cost effectiveness of offering a programme on a repeated basis. The GDG formed a consensus that patients are more likely to benefit from repeated exercise components of rehabilitation, but the incremental value of offering a repeated educational component may be minimal as the effect could be sustained for longer. In conclusion, it is likely that a repeated offer of rehabilitation of exercise maintenance is more cost effective than a repeated offer of educational classes on symptom management. The GDG noted that in order to undertake pulmonary rehabilitation assessment, lung function tests would have already been recently performed and oxygen requirements are already established and catered for. It was acknowledged oxygen reassessment may be required within a short timeframe prior to the rehabilitation assessment. Overall, the GDG thought that the model results indicated pulmonary rehabilitation to be highly cost effective using conservative assumptions, and in consideration of the limitations of the model likely to underestimate the cost effectiveness of these programmes. The probabilistic analysis showed it is not certain which type of rehabilitation (exercise alone or with an educational component) is most cost effective. A three way analysis showed that if the same treatment effect is observed on repeated offers, unless duration of treatment effect is very long (i.e. 24 months), it is most cost effective to repeat the programme every 6 months. If it is expected that each repeated programme is at least 80% as effective as the one previously undertaken, it is likely that repeating the programme every 12 months will be cost effective. This is with the exception when the treatment effect is likely to be less than 18 months. Once the magnitude of effect started to decrease by 60% on each subsequent programme the optimal time interval between programmes extends to 18 months or more. If the effectiveness of programmes more than halve on each offer, it is increasingly likely that the programme should not be repeated. The model highlighted that further research was required to help inform which components, length, duration and frequency of pulmonary rehabilitation course was optimal for this patient group.
Quality of evidence	Evidence was retrieved from 13 studies (this included 1 Cochrane review was which provided data on 2 RCTs, 9 observational studies and 1 abstract). Quality of life outcomes ranged from moderate to very low quality due to small sample size, lack of blinding and no allocation concealment. Studies which reported on the use of physical training for people with ILD compared to no treatment/usual care for people with IPF/ILD, informed this review question. The GDG discussed the potential benefit of pulmonary rehabilitation for improving all health related quality measures, but acknowledged that the quality and study type of evidence received showed conflicting effects in domain scores for SF36 and SGRQ at certain time points (immediately after training and after long term follow-up of 6 months). Differences in baseline characteristics between patient groups in the trials may explain the conflicting results seen at certain points, as differences in lung function would not have been accounted for. The GDG acknowledged that difference in 'change' scores from baseline showed an overall improvement in health related quality of life domain scores. The GDG discussed that the lack of IPF tailored pulmonary rehabilitation programmes may reflect variation in practice in the UK and therefore explain why there is no data

	examining the effect of different components of pulmonary rehabilitation programmes for people with IPF. The effect of any individual component of the programme, for example optimisation of oxygen, remains unclear. Currently, people with IPF are either offered pulmonary rehabilitation designed for people with COPD or they are not offered pulmonary rehabilitation due to the lack of access and availability of programmes. As no evidence was retrieved that investigated the optimal course content or duration of pulmonary rehabilitation programmes, the GDG considered the personal experiences of the patient members of the guideline group. Knowledge regarding pulmonary rehabilitation and typical courses of pulmonary rehabilitation informed the following discussions: Availability and duration of IPF specific pulmonary rehabilitation, as currently availability of programmes varies according to region and is largely tailored for people with COPD. Experiences of components of pulmonary rehabilitation, such as psychosocial support and education regarding: diet; exercise; social support; and benefits. The GDG agreed that it was appropriate to include abstracts and observational studies with no comparison group, as evidence to inform this review question, because pulmonary rehabilitation was identified as a high priority area for health economic modelling, and due to the poor quality and lack of evidence found. (It should be noted that across the guideline, relevant abstracts and observational studies were not always retrieved for areas where no evidence was found or where published studies were considered low quality. In some instances, abstracts and observational studies were not included as evidence to inform a review question, because they were not deemed by the GDG to add additional value over published studies to inform decision making).
Other considerations	The GDG considered patient access to pulmonary rehabilitation programmes to be important. The patient members highlighted the importance of the education component of the PR programme of particular importance when learning how to live with IPF. The GDG acknowledged that components of rehabilitation programmes designed for COPD may be inappropriate for an IPF population, and a greater health benefit may be realised with courses tailored to the IPF population. The GDG considered the following guidance when making recommendations for pulmonary rehabilitation: Chronic obstructive pulmonary disease. NICE clinical guideline 101 (2010). Available from www.nice.org.uk/guidance/CG101 Research recommendations The GDG agreed that the lack of evidence for pulmonary rehabilitation tailored specifically to people with IPF justified developing a research recommendation to address whether pulmonary rehabilitation improves outcomes for people with IPF. For further information on the research recommendations see Appendix P.

9 Best supportive care

9.1 Review Introduction

Best supportive care aims to help people with IPF and their carers cope with their condition. This help should run in parallel with the diagnostic process, continue during the course of disease through death. Supportive care also helps optimise quality of life at different stages of what is often a progressive illness. There is considerable variation in practice with the delivery and components of best supportive care.

A large component of best supportive care is advice and management of symptoms, particularly breathlessness, cough and fatigue. Palliation of these symptoms requires a multidisciplinary approach, including input from primary care, ILD specialist nurses and specialists in palliative care. Oxygen therapy is a particularly important intervention. People with IPF typically experience breathlessness on exertion, which is often associated with hypoxia. Exercise-induced hypoxaemia in people with IPF may be more dramatic and unpredictable than in patients with other lung diseases, and higher flow rates of oxygen are frequently required to correct this hypoxaemia. Furthermore, resting oxygen is not a good indicator of oxygen desaturation on exercise. Domiciliary oxygen can be delivered by an oxygen concentrator for long-term use or in a portable form for ambulatory use. Portable oxygen cylinders weigh about 2-3kg (6-7lb) and come with a carrying case. Some of the various components of best supportive care are listed below.

- Accurate diagnosis and explanation of management options
- Complementary therapies
- End of life and bereavement care
- Management of co-morbidities
- Oxygen therapy
- Providing feedback on disease progression including test results and prognosis
- Psychological support
- Pulmonary rehabilitation
- Social support
- Spiritual support
- Symptom control
- Teaching self-management strategies
- Withdrawal of ineffective therapies

9.2 Clinical question and review methodology

The following clinical question was included in this chapter:

9.2.1 What is the clinical and cost effectiveness of best supportive care (palliation of cough, breathlessness and fatigue, and oxygen management) in the symptomatic relief of people with IPF?

For full details see review protocol in Appendix C.

Table 37: PICO characteristics of review question

Population	Adults with confirmed IPF and/ or ILD
Intervention/s	• Oxygen management • Palliation of cough

	• Palliation of breathlessness • Palliation of fatigue
Comparison/s	No treatment/usual care
Outcomes	Critical outcomes 1. Improvement in health-related quality of life Other outcomes 2. All cause and IPF related mortality 3. Hospitalisations due to IPF complications (including IPF exacerbations) 4. Improvement in cough and breathlessness 5. Improvement in psychosocial health (including depression) 6. Performance on sub-maximal walk test (distance walked and lowest SaO_2) 7. Symptom relief
Study design	RCTs, systematic reviews of RCTs, cohort studies

The objectives of this review were to determine the most clinically and cost effective best supportive care for people with IPF. No restrictions were used for sample size or publication date and the population was extended to include ILD patients and studies in abstract form in order to capture all relevant data. However this excludes studies which stated they did not have any IPF patients included in the ILD group investigated. Studies with an indirect population such as COPD were not included as the GDG considered that people with COPD have different disease trajectories and needs and are thus not comparable with people who have IPF.

9.3 Clinical evidence

We searched for randomised trials and cohort studies comparing different strategies for best supportive care versus no treatment/usual care for people with IPF/ILD. Best supportive care comprised of; oxygen management, palliation of cough, palliation of breathlessness, and palliation of fatigue.

Nine studies were identified which covered the following areas of best supportive care; oxygen management[19,134,125,90], palliation of cough[46,47,109,48] and palliation of breathlessness[20]. No studies were identified which studied interventions aimed at palliation of fatigue.

Two systematic reviews were identified for oxygen management. A Cochrane review[19] and a systematic review [134] both provided evidence from an unpublished study by Braghiroli et al[11]. Only the study data relevant to the review question was extracted from both reviews. One RCT was identified by Swinburn et al [125] investigating the use of oxygen versus air in a double blind cross over study and one observational study in abstract form by Obi et al[90]. The study by Obi et al[90] looked at the use of oxygen therapy in a population of patients with advanced chronic lung diseases including IPF patients, there was no comparison/control group so a meta-analysis could not be carried out and the data is presented in this report as described in the study.

Hopegill et al[46] investigated the use of prednisolone for the palliation of cough. The study's primary aim was to assess the responsiveness of IPF patients to cough inducing agents. A small sub set of patients were treated with prednisolone to investigate how the cough response is affected. Horton et al[48] also looked at the treatment of cough in IPF patients but with thalidomide. For both studies due to lack of data, a meta-analysis could not be carried out and the data is presented in this report as described in the study. Two abstracts [47,109] were also identified which have been included looking at thalidomide for the palliation of cough.

Currow et al[20] investigated the use of morphine for the palliation of breathlessness. Data were taken from the phase II arm of a pharmacovigilance study and due to the lack of a direct comparison, the data could not be meta-analysed and is shown in this report as reported in the study.

Presentation of the evidence in this review has been separated into randomised controlled trials and observational studies. Due to the lack of a control group/direct comparison with observational studies the data is shown in this review as reported in the studies.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

9.3.1 Summary of included studies

Table 38: Summary of studies included in the review

Study	Intervention / control	Population	Outcomes	Comments
Oxygen management				
Crockett 2001[19] & Zielinski 2000[134]	Treatment with long-term domiciliary oxygen therapy vs. no oxygen therapy.	Patients diagnosed with interstitial pulmonary fibrosis N= 62	Mortality.	Review of Braghiroli unpublished data (1988). A number of additional outcomes mentioned for which results were not reported. Method of randomisation not stated. The method of blinding was not described. Missing baseline data per group.
Obi 2010[90]	Supplementary O_2 versus nil supplementary O_2. Comparison of 6MWT done with and without O_2 on the same day.	Advanced chronic lung diseases IPF data presented separately N=22	6MWD, lowest SaO_2, dyspnoea.	Abstract only. No baseline characteristics. No blinding. No randomisation. Small sample size. No description of sample given.
Swinburn 1991[125]	Patients received both oxygen (28%) and air through the same face mask using the same source flow rate (4L/min).	ILD including cryptogenic fibrosing alveolitis (8 patients) amiodarone lung toxicity (1 patient) hypersensitivity pneumonitis (1 patient) N=10	SaO_2, % Visual analogue scale (VAS) intensity of dyspnoea (100mm VAS).	Double blind cross over study. Baseline VAS scores not provided. Small sample size. Order effects and carry-over between treatments: unclear if wash out period is adequate, potential for confounding. Method of randomisation not stated.
Palliation of cough				
Prednisolone for the palliation of cough				
Hopegill 2003[46]	Patients received prednisolone 40-60 mg per day for at least 4 weeks.	Patients diagnosed with IPF according to ATS criteria. And a visual analogue scale (VAS)	VAS intensity of cough (10cm)	No baseline data provided. Small sample size. No comparison/control. Method of blinding not

Study	Intervention / control	Population	Outcomes	Comments
		intensity of cough of more than 5 cm N= 6		reported.
Thalidomide for the palliation of cough				
Horton 2008[47]	Thalidomide daily in 100-400mg / no control.	Individuals with chronic cough caused by IPF N=11	Cough score (question 2 of the SGRQ).	Open label phase II trial. Abstract: limited information on methodology and patient characteristics at baseline and post treatment. Small sample size. At 3 months follow up 5 drop outs.
Horton 2012[48]	Thalidomide (50mg increased to 100mg if no improvement in cough after 2 weeks) patients also received sodium docusate 100mg for constipation/ Placebo (12 weeks in each arm). All subjects received vitamin B complex supplements and all prescriptions for cough were discontinued 2 weeks prior to study.	Patients diagnosed with IPF and chronic cough (>8 weeks) N=24	Cough specific QoL measured by the cough quality of life questionnaire. VAS intensity of cough (10cm) SGRQ.	Small sample size. Single centre study. Short duration of study. Treatment cross-over unclear if washout out period is adequate.
Saini 2011[109]	Thalidomide: no further details provided.	Patients who had severe enough cough after 6 weeks of treatment with omeprazole 40mg and prednisolone 10 mg. N=6	Cough score (modified version of the Leicester Cough Questionnaire in conjunction with subjective symptoms).	Abstract. Small sample size. 3 patients stopped thalidomide due to rash, 2 were stable at 50mg daily and 1 was stable at 50 mg alternate daily. Unclear which of these people had IPF. Total number of people with IPF = 4/6. Follow up period not stated.
Palliation of breathlessness				
Morphine for the palliation of breathlessness				
Currow 2011[20]	Patients received 10mg daily of sustained-release morphine sulphate, which was increased in non-responders by	Patients with a palliative diagnosis (only ILD reported on). N= 63 (ILD =10)	VAS intensity of dyspnoea (100mm)	Indirect intervention - phase II of a pharmacovigilance study. No comparison/control group

Study	Intervention / control	Population	Outcomes	Comments
	10mg daily each week to a maximum of 30mg daily. Administered with laxatives (sodium docusate with sennosides).			Small sample size. Method of randomisation and blinding not reported.

9.3.2 Study quality and summary of findings

Table 39: Clinical evidence profile: Best supportive care; oxygen management – randomised controlled trial

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oxygen	Control (Air)	Relative risk (95% CI)	Absolute risk	
Mortality (12 months) Crockett 2001 [19]											
1	Randomised trials	Serious[1,2,3]	No serious inconsistency	No serious indirectness	Very serious[4]	None	7/37 (18.9%)	8/25 (32%)	RR 0.59 (0.25 to 1.42)	131 fewer per 1000 (from 240 fewer to 134 more)	Very low
Mortality (24 months) Crockett 2001 [19]											
1	Randomised trials	Serious[1,2,3]	No serious inconsistency	No serious indirectness	Serious[5]	None	23/37 (62.2%)	12/25 (48%)	RR 1.3 (0.8 to 2.09)	144 more per 1000 (from 96 fewer to 523 more)	Low
Mortality (3 years) Crockett 2001 [19]											
1	Randomised trials	Serious[1,2,3]	No serious inconsistency	No serious indirectness	No serious imprecision	None	34/37 (91.9%)	23/25 (92%)	RR 1 (0.86 to 1.16)	0 fewer per 1000 (from 129 fewer to 147 more)	Moderate
Arterial oxygen saturation (Better indicated by higher values) Swinburn 1991 [125]											
1	Randomised trials	Very serious[1,6,7]	No serious inconsistency	No serious indirectness	Serious[5]	None	10	10	-	MD 9.2 higher (5.43 to 12.97 higher)	Very low
Dyspnoea (VAS) (measured with: Visual analogue scale; Better indicated by lower values) Swinburn 1991 [125]											
1	Randomised trials	Very serious[1,6,7,8]	No serious inconsistency	No serious indirectness	Serious[5]	None	10	10	-	MD 17.9 lower (31.18	Very low

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oxygen	Control (Air)	Relative risk (95% CI)	Absolute risk	
										to 4.62 lower)	

[1] *Method of randomisation and allocation concealment not stated*
[2] *Method of blinding not described*
[3] *Baseline data per group not given*
[4] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*
[5] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID*
[6] *Small sample*
[7] *Order effects and carry-over between treatments- unclear if wash out period is adequate. Potential for confounding*
[8] *Baseline VAS scores not provided*
[9] *Abstract only*
[10] *No baseline characteristics or description of sample given*

Table 40: Clinical evidence profile: Best supportive care; oxygen management – observational study

Quality assessment							No of patients	Effect (Mean change from baseline)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oxygen		
6MWD (m) Obi 2010 [90]									
1	Observational study	Very serious[1,2,6,9,10]	No serious inconsistency	No serious indirectness	Could not be calculated	None	22	19.17	Very low
Lowest SpO_2 (%) Obi 2010 [90]									
1	Observational study	Very serious[1,2,6,9,10]	No serious inconsistency	No serious indirectness	Could not be calculated	None	22	4.83 p<0.05	Very low
Dyspnoea (maximal Borg score) Obi 2010 [90]									

Quality assessment							No of patients	Effect (Mean change from baseline)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Oxygen		
1	Observational study	Very serious[1,2,6,9,10]	No serious inconsistency	No serious indirectness	Could not be calculated	None	22	-1.04 p<0.05	Very low

[1] *Method of randomisation and allocation concealment not stated*
[2] *Method of blinding not described*
[3] *Baseline data per group not given*
[4] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*
[5] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID*
[6] *Small sample*
[7] *Order effects and carry-over between treatments- unclear if wash out period is adequate. Potential for confounding*
[8] *Baseline VAS scores not provided*
[9] *Abstract only*
[10] *No baseline characteristics or description of sample given*

Table 41: Clinical evidence profile: Best supportive care; prednisolone for palliation of cough – observational study

Quality assessment							No of patients	Effect (100 point scale ± SD)		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Prednisolone	Baseline	Post treatment	
Cough (VAS) (follow-up 4 weeks; assessed with: Visual analogue scale) Hopegill 2003 [46]										
1	Observational studies	Very serious[1,2,3]	No serious inconsistency	Serious[4]	Could not be calculated	None	6	7.2±0.8	2.2±2.5	Very low

[1] *No baseline data provided*
[2] *Small sample size*
[3] *observational study biases and no comparison*
[4] *Indirect intervention- prednisolone was used to study the cough reflex to stimulants, there is no direct comparison i.e. A vs. B*

Table 42: Clinical evidence profile: Best supportive care; thalidomide for palliation of cough – randomised controlled trial

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Thalidomide	Placebo	Baseline	Post treatment	Quality
QoL (follow-up 12 weeks; assessed with: Cough quality of life questionnaire) Horton 2012[48]											
1	RCT	Serious	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 60.5 SD:12.0	Mean: 58.7 SD:14.0	Low
QoL (follow-up 12 weeks; assessed with: Visual analogue scale) Horton 2012[48]											
1	RCT	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 64.8 SD:21.4	Mean: 61.9 SD:26.5	Low
QoL (follow-up 12 weeks; assessed with: SGRQ total) Horton 2012[48]											
1	RCT	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 57.4 SD:18.8	Mean: 56.9 SD:17.1	Low
QoL (follow-up 12 weeks; assessed with: SGRQ symptom domain) Horton 2012[48]											
1	RCT	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 67.7 SD:19.7	Mean: 62.0 SD:18.3	Low
QoL (follow-up 12 weeks; assessed with: SGRQ impact domain) Horton 2012[48]											
1	RCT	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 48.1 SD:20.7	Mean: 49.0 SD:19.4	Low
QoL (follow-up 12 weeks; assessed with: SGRQ activity domain) Horton 2012[48]											
1	RCT	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Could not be calculated	None	12	12	Mean: 64.3 SD:22.7	Mean: 65.8 SD:18.7	Low

[1] Treatment crossover from placebo to thalidomide arm – it is unclear if the washout period is adequate, there may be carry over effects therefore potential for confounding

[2]Unclear allocation concealment

[3]Small sample size and single centre study thus there are limitations on the generalisability of the results to other populations of IPF patients

[4]Short duration of study

Table 43: Clinical evidence profile: Best supportive care; thalidomide for palliation of cough – observational studies

Quality assessment							No of patients	Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Thalidomide	Baseline	Post treatment	
Cough score (follow-up 3 months; assessed with: question 2 of the SGRQ) Horton 2008 [47]										
1	Observational study	Very serious [1,2]	No serious inconsistency	No serious indirectness	Could not be calculated	None	11	Mean: 4.9 SD: 0.3	Mean: 2.2 SD: 1.6	Very low
Cough score (follow-up NR; assessed with: modified version of the Leicester Cough Questionnaire) Saini 2011[109]										
1	Observational study	Very serious[1,2,3]	No serious inconsistency	No serious indirectness	Could not be calculated	None	6	Median: 74.5 IQR: 13.25	Median: 51.5 IQR: 49.25	Very low

[1] *Abstract limited information on methodology and patients' characteristic baseline and post treatment data*
[2]*Small sample size- 3 patients stopped thalidomide due to rash, 2 are stable at 50mg daily and 1 is stable at 50 mg alternate daily- of which have IPF?(Saini 2011[109])* And *at 3 months follow up 5 drop outs(Horton 2008[47])*
[3]*Follow up period is not stated*

Table 44: Clinical evidence profile: Best supportive care; morphine for palliation of breathlessness- observational study

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other consideration	Morphine	Difference between the first and last measurements	
Dyspnoea (VAS) (assessed with: Visual Analogue Scale) Currow 2011 [20]									
1	Observational study	Serious[1,2]	No serious inconsistency	Serious [3]	Could not be calculated	None	10	3.2 (SD:32.7, Median: 3.9, Range: -46 to 61)	Very low

[1] *Small sample size*
[2] *No control group and observational study biases*
[3]*Indirect intervention-phase II of a pharmacovigilance study there is no direct comparison i.e. A vs. B, ILD population does not state if there are any IPF patients present*

9.4 Economic evidence

Published literature

No relevant economic evaluations comparing strategies of oxygen management, or palliation of cough, breathlessness or fatigue were identified. One cost minimisation study [86] was selectively excluded on the account that the population in the sample predominantly had obstructive, rather than restrictive lung disease. This is summarised in Appendix R, with reasons for exclusion given.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs of interventions listed in the clinical review and listed as best supportive care options are provided below to aid consideration of cost effectiveness.

Table 45: Unit costs for best supportive care options[3,87]

Item	Unit cost	Notes
Cough		
Prednisolone (40mg per day for at least 4 weeks)	Cost per 5mg 28 tab pack = £0.96 Cost per week for 40mg = £1.92	Corticosteroid monitoring in primary care and vitamin supplements given with long term use.
Thalidomide Celgene® (100mg – 400mg per day)	Cost per 50mg 28 cap pack = £298.48 Cost per week (100mg per day) = £149.24	Non tariff drug, cost sourced via MIMS
Simple Linctus (5 ml three to four times daily)	£0.92 per 200ml £0.64 per week	
Codeine phosphate (15-30mg (twice to four times a day)	Cost per 30mg 100 tab pack = £5.32 Cost per week (30mg four times per day) = £1.49	
Dextromethorphan	Available in over the counter cough syrups at varying costs rather than in prescription products.	Cost would not be incurred by the NHS
Breathlessness		
Morphine modified release capsules (10mg) Zomorph Morphine modified release capsules (30mg) Zomorph	Cost per 10mg 60 cap pack = £3.47 Cost per week = £0.40 Cost per 30mg 60 cap pack = £8.30 Cost per week = £0.97	Dosage reported in clinical review: Patients received 10mg daily of sustained-release morphine sulphate, which was increased in non-responders by 10mg daily each week to a maximum of 30mg daily. Administered with laxatives (sodium docusate with sennosides)
Diazepam Lorazepam	Cost per 10mg 28 tab pack = £0.79 Cost per week (10mg per day) = £0.20 Cost per 2.5 mg 28 tab pack = £4.47 Cost per week (2.5mg per day) = £1.12	
Oxygen Management. Source: Personal communication with the Department of Health (2012); NHS reference costs 2010-2011[25]		
Long term oxygen (home) (a)	Concentrator and back up cylinder =£700/annum	Average price based on single concentrator being piped in, single

Item	Unit cost	Notes
		static cylinder being refilled 12 times in the year, risk assessment and servicing of the concentrator (excludes electricity).
Long term (home) and ambulatory oxygen [(b)]	Concentrator and back up cylinder, and 2 ambulatory cylinders =£1600/annum	Average price based on single concentrator being piped in, single static cylinder being refilled 12 times in the year, risk assessment and servicing of the concentrator and 2 ambulatory cylinders that are refilled 26 times in the year (excludes electricity).
Oxygen assessment and monitoring (DZ38Z: Outpatient)	£181 (Inter quartile range: £137 to £218)	It is probable that ambulatory oxygen and LTOT assessment would be coded together if use was concurrent, which may distort the unit cost reported.
Long-term Oxygen Therapy Test (DA17: Direct Access)	£201 (Inter quartile range: £146 to £284)	

(a) The assumptions underpinning the cost estimate for Long Term Oxygen use were:

1) *The prices are an average across all 10 England regions (based on current SHA boundaries)*
2) *The patient will use the equipment as indicated and not need any additional visits or equipment in the year*
3) *The patient will not need a holiday supply or secondary supply*
4) *Patient will use a concentrator for home use 364 days (electricity not factored in as not possible to calculate)*
5) *Patient will have a backup cylinder that they may use up monthly (albeit they shouldn't)*
6) *The concentrator will be serviced 3 times within first 12 months*

(b) The assumptions underpinning the cost estimate for Ambulatory Oxygen use were:

7) *The patient will have 2 Ambulatory cylinders*
8) *They will not need more than 26 refills in a 12 month period*
9) *No conserving device is in use*

9.5 Evidence statements

Oxygen management

Very low quality evidence suggests that long term domiciliary oxygen therapy is more effective than no oxygen therapy at reducing 12 month mortality in people with IPF (one study, n=62).

Low quality evidence suggests that long term domiciliary oxygen therapy increases mortality at 24 months compared to no oxygen therapy in people with IPF (one study, n=62).

Moderate quality evidence suggests that there is no difference between long term domiciliary oxygen therapy and no oxygen therapy and mortality at 3 years in people with IPF (one study, n=62).

Very low quality evidence suggests that oxygen therapy is more effective than air at increasing arterial oxygen saturation in people with IPF (one study, n=10).

Very low quality evidence suggests that oxygen therapy is more effective than air at reducing levels of dyspnoea in people with IPF (VAS scale) (one study, n=10).

Very low quality evidence suggests that oxygen therapy may be effective at reducing levels of dyspnoea (Borg score) in people with IPF, but imprecision could not be assessed (one study, n=22).

Very low quality evidence suggests that oxygen therapy is effective at improving 6MWD (m) in people with IPF, but imprecision could not be assessed (one study, n=22).

Very low quality evidence suggests that oxygen therapy may be effective at improving lowest level of SpO_2 in people with IPF, but imprecision could not be assessed (one study, n=22).

Palliation of cough

Very low quality evidence suggests that prednisolone therapy may be effective at reducing levels of cough (VAS scale) at 4 weeks follow up in people with IPF, but imprecision could not be assessed (one study, n=6).

Very low quality evidence suggests that thalidomide therapy is effective at reducing levels of cough (using question 2 of the SGRQ and modified version of the Leicester Cough Questionnaire in conjunction with subjective symptoms) in people with IPF/ILD, but imprecision could not be assessed (two studies, n=17).

Palliation of breathlessness

Very low quality evidence suggests that morphine therapy is effective at reducing levels of breathlessness (VAS scale) in people with ILD (one study, n=10).

Economic

- No economic evaluations were identified with the relevant comparators.

9.6 Recommendations and link to evidence

Recommendations	**18. Offer best supportive care to people with idiopathic pulmonary fibrosis from the point of diagnosis. Best supportive care should be tailored to disease severity, rate of progression, and the person's preference, and should include if appropriate:** • **information and support (see recommendation 2)** • **symptom relief** • **management of comorbidities** • **withdrawal of therapies suspected to be ineffective or causing harm** • **end of life care.**
Relative values of different outcomes	The GDG considered the critical outcome for this recommendation to be improvements in health related quality of life. The GDG also considered reductions in breathlessness, cough, fatigue, psychosocial health and symptom relief to be important outcomes to inform this recommendation.
Trade-off between clinical benefits and harms	The GDG discussed the overall harms and benefits associated with the different components of best supportive care. They considered best supportive care to be a care package tailored to the individual requirements (stage and rate of IPF progression) and preferences of people with IPF. Therefore, no appreciable harms were associated with best supportive care. The benefits associated with best supportive care were considered to be improvements with quality of life and empowering people with IPF and their carers to

	feel in control of their well-being through raising awareness of their illness. The GDG discussed the setting and timing of when best supportive care strategies should be implemented in order to achieve maximal health benefit. Due to the unpredictable progression of symptomatic disease, best supportive care measures should be considered as early on in the care pathway as possible and on a case by case basis. Clinical judgement and patient preferences should play an important role when determining the implementation of best supportive care interventions in order to achieve maximal improvement in quality of life. Secondary or tertiary care team members with expertise in ILD (those who run a service seeing at least 500 ILD patients per year or have completed specialist training in ILD for at least 6 months) should be involved throughout implementation of best supportive care strategies. This was considered necessary due to the specialist nature of the care required for IPF and would involve close collaboration with primary care and palliative care services. The patient members of the GDG commented that discharge from a specialist team to palliative care services could negatively impact on psychosocial wellbeing and the continuity of care is an important consideration. The continuity of care and communication between teams is also an important aspect in ensuring that patient preference and history is known by those implementing best supportive care strategies, which in turn is likely to maximise the clinical benefit of these strategies. Furthermore, continuation of care could facilitate patient preference to be incorporated in decision making, which could in turn enhance the benefits associated with best supportive care measures (a sense of increased control over the symptoms of IPF). The GDG considered that an ILD nurse who is involved from the start of a care pathway at diagnosis, through to offering advice in best supportive care, could be one of many possible means of achieving this.
Economic considerations	No published economic evaluations were identified to inform this question. Consideration of cost effectiveness of best supportive care strategies were undertaken with an understanding that the available evidence could not support a recommendation to offer disease modifying pharmacological treatment and a cure for IPF has not yet been established. In this context and in the absence of published economic evidence, the GDG felt strongly that the opportunity to improve the quality of life for IPF patients through a comprehensive best supportive care strategy justified the costs involved. The specific economic considerations given to each best supportive care intervention are shown in each of the relevant links to evidence. The GDG discussed the unit costs of interventions that could be considered part of best supportive care, alongside consideration of the resource implication and cost of adverse effects, appropriate monitoring to ensure maximal health benefits and withdrawal from ineffective treatment (which should ensure appropriate use of healthcare resources). Regular follow up and review was agreed to improve cost effectiveness of best supportive care strategies for people with IPF, especially given the unpredictability of the type and rate of disease progression. It was recognised that formal assessments are required for the more expensive interventions such as pulmonary rehabilitation and oxygen management, which is justified by the increased likelihood of appropriate healthcare resource use. The GDG were unable to make a recommendation on specific service and commissioning arrangements. There was agreement that involvement of an ILD specialist nurse for referral and advice would be beneficial in the best supportive care

	strategy, given the specialist skill set required, the need to incorporate patient preference, knowledge of clinical history and the need to consider best supportive care options from diagnosis to end of life. The recommended involvement of the ILD nurse in the diagnostic MDT could minimise the incremental cost of continued ILD nurse involvement into the best supportive care strategies and ensure best supportive care was considered as early as possible in the care pathway. As part of this discussion, the potential incremental cost to the NHS of the involvement of specialist staff and enhanced communication was discussed. As the majority of IPF patients are already followed up in secondary or tertiary care, the actual cost of staff salary and overheads would not pose a substantial incremental cost, as the specialist interest represented a difference in expertise rather than a need to increase the grade of the staff involved. Overall, it was thought that the benefit of continued specialist care involvement throughout the care pathway (that is,. from diagnosis through to monitoring and advising on best supportive care options) would likely offset the cost of specialist staff involvement.
Quality of evidence	This recommendation was partially based on GDG consensus due to the lack of evidence regarding the key components and timing of a best supportive care package for people with IPF. Overall, nine studies were identified for best supportive care which ranged from very low to moderate quality and which covered oxygen management, palliation of cough and breathlessness. The GDG considered that there was uncertainty in the interpretation of the results from these studies due to the risk of bias. No studies were identified which assessed interventions aimed at palliation of fatigue.
Other considerations	The GDG discussed the importance of using clinical judgement when discussing best supportive care interventions with people with IPF and their carers. In particular, clinical judgement will be needed to assess the likely rate of symptomatic disease progression and the appropriateness of the interventions available. However, the GDG felt strongly that initiation of discussion and consideration of best supportive care interventions should occur when IPF is diagnosed and be followed through to referral to, and working with, palliative care services. The different components of best supportive care were considered to be beneficial in improving all of the following areas (as identified by the SF 36 health status questionnaire): vitality (as a reciprocal indicator of fatigue) physical functioning bodily pain general health perceptions physical role functioning emotional role functioning social role functioning mental health The GDG considered patient preferences for pharmacological intervention to be important considerations, whilst highlighting the potential side effects associated with medication. The GDG also discussed the importance of health professionals recognising patient's individual spiritual and religious beliefs when providing best supportive care.

	The GDG considered other relevant NICE guidance such as the Lung cancer NICE clinical guideline 121 (2011) and the Chronic obstructive pulmonary disease NICE clinical guideline 101 (2010) when making recommendations for best supportive care.

Recommendations	**19.If the person is breathless on exertion consider assessment for:** • **the causes of breathlessness and degree of hypoxia and** • **ambulatory oxygen therapy and long-term oxygen therapy and/or** • **pulmonary rehabilitation.** **20.If the person is breathless at rest consider:** • **assessment for the causes of breathlessness and degree of hypoxia and** • **assessment for additional ambulatory oxygen therapy and long-term oxygen therapy and** • **the person's psychosocial needs and offering referral to relevant services such as palliative care services and** • **pharmacological symptom relief with benzodiazepines and/or opioids.**
Relative values of different outcomes	The GDG considered the critical outcome for this recommendation to be improvements in health related quality of life. The GDG also considered improvements in breathlessness and in psychosocial health to be important outcomes to inform these recommendations.
Trade-off between clinical benefits and harms	The GDG discussed the harms and benefits associated with oxygen therapy as a means of symptom relief for breathlessness. The benefits were considered to be improvements in breathlessness and quality of life. The GDG acknowledged people with IPF may be breathless due to multiple factors that include hypoxia, co-existing COPD, co-existing pulmonary hypertension and deconditioning. A patient may be hypoxic during exercise without marked symptoms. It is not known if oxygen therapy (or other best supportive care measures) will extend life. The potential harms of oxygen therapy are uncertain. Ambulatory oxygen therapy requires the patient to carry portable oxygen, and the benefits of the oxygen need to be balanced against the extra weight being carried. People with IPF may feel inhibited about using ambulatory oxygen, which is easily visible, in public places. The GDG agreed there is currently wide variation in the use of ambulatory oxygen, probably in part due to variation in referral to ambulatory oxygen assessment. The decision to assess for oxygen is frequently delayed until long term oxygen management is considered, at which point in the person's clinical pathway ambulatory oxygen may be of less value in terms of improving quality of life, therefore, currently ambulatory oxygen may be underutilised in the IPF population. The GDG also noted that pulmonary rehabilitation may improve breathlessness and psychosocial health by empowering people with IPF to feel in control of their illness. Exercise exertion beyond a person's normal capacity was considered to be the only harm associated with pulmonary rehabilitation, but that this risk was unlikely due to the programmes being conducted by trained health professionals (physiotherapists

	and nurses). The GDG agreed that there were unlikely to be benefits associated with opiate or benzodiazepine use for the symptomatic relief of breathlessness on exertion, but did acknowledge there were potential benefits at rest. The sedation effects of opiates were also considered to have a potential benefit in reducing anxiety. The GDG acknowledged that use should be based on clinical judgement and patient preferences due to potential side effects of opiates which may include excessive respiratory depression, nausea, vomiting or constipation, and benzodiazepines which may include excessive respiratory depression, drowsiness or dizziness. Evidence from the pharmacological interventions review that provided data on the effect of drugs on cough, breathlessness or fatigue was discussed. Three studies measured the effect on dyspnoea in people with IPF, when treated with sildenafil or bosentan, when compared to placebo, for disease modifying purposes. None of these studies resulted in a clinically important improvement in dyspnoea and therefore were not considered to be viable treatments for breathlessness in people with IPF.
Economic considerations	There was no published and applicable economic evidence regarding symptom relief for breathlessness, including oxygen management, for people with IPF. The cost of pharmacological symptom management with opiates and benzodiazepines is less than that of sildenafil (reviewed as a means to modify disease progression, but also shown to have a potential impact on breathlessness). The GDG thought that management with opiates and benzodiazepines was likely to be more cost effective than the use of sildenafil due to their substantially lower acquisition cost, and these should be recommended as an option for the relief of breathlessness. The GDG considered the estimated cost of oxygen for an average IPF patient, but acknowledged the assumptions underlying the estimate, the regional variation in cost, and the variation in oxygen consumption between people with IPF. Clinical members of the GDG informed the group that a new contract for oxygen services occurred in May 2012. Unlike the old contract where oxygen was charged for on a day by day basis, under the new contract installation of equipment, daily rental and refills (so actual usage) are charged for; the price of which is determined according to regional contracts. The GDG also considered the unit cost of oxygen assessment alongside the specifications of the number of tests involved, the manpower and follow up required for each type of assessment as recommended by the British Thoracic Society (2006).[12]. This comparison gave rise to concern that the actual cost of assessment could be lower than what the NHS reference cost suggested. Given concerns that the average NHS reference cost for assessment may be overestimated, it is likely that oxygen management is the same or less costly than breathlessness management using sildenafil and is likely to provide equal or greater improvement in quality of life. The cost of oxygen management is in part offset by reduced contact with the healthcare system in management of breathlessness (for example it may reduce the number of primary care out of hour calls – with one out of hour call out for a general practitioner costing approximately £121[99]) and one emergency admission, costing £197 (in accident and emergency [type1] with a category 3 investigation and treatment [i.e. CT scan with supplemental oxygen])[25].

	Using NHS reference costs, the cost of an IPF related hospitalisation (without any other complications or co-morbidities) can be approximated at £1174 [24]. This suggests that if one to two IPF related hospitalisations could be avoided per year, oxygen management could be cost neutral or even cost saving. If appropriate oxygen management was offered earlier in the course of disease progression, the cumulative difference in quality of life improvement could be large, and would justify the initial cost of assessment and installation, even if reduced hospitalisation did not occur. However, the GDG exercised caution when using NHS reference costs in their decision making, noting that it was possible the NHS reference cost for a hospital admission for an ILD patient would not accurately reflect the costs incurred by the IPF population group. This is due to the relatively small size of the IPF population in comparison to other patient populations which also contribute to the calculation of the reference cost. They also noted that due to regional pricing, the cost of oxygen may vary from that quoted in the review. Taking the above into account, the GDG considered that oxygen therapy offered after formal assessment was likely to be cost effective as a means to improve quality of life compared to a do nothing approach. However, it still remains unclear what the most cost effective strategy to manage breathlessness is in the IPF population and further research is required. Pulmonary rehabilitation may also be a cost effective means to manage breathlessness and the NCGC model suggests pulmonary rehabilitation is very likely to be extremely cost effective as a means of improving quality of life. Oxygen management, however, should be considered as an adjunct intervention rather than a direct comparator as it enables participation in rehabilitation.
Quality of evidence	Two systematic reviews and one RCT were identified for oxygen management. The quality of evidence ranged from moderate to very low quality. The studies showed that oxygen is more effective than air at improving perceived levels of dyspnoea, arterial oxygen saturation, and improved 12 month mortality rates compared with no oxygen therapy. They also showed that oxygen increased 24 month mortality rates, but showed no difference in mortality at three years compared to no oxygen therapy. However, there was uncertainty in the effect. One study was identified for the palliation of breathlessness. Data was taken from the phase II arm of a pharmacovigilance study, which was investigating the use of morphine for the palliation of breathlessness. Again due to the lack of a direct comparison, the data could not be meta-analysed. The evidence showed that morphine was effective at reducing the perceived levels of breathlessness. However, there was uncertainty in the effect and the study was of very low quality.
Other considerations	The GDG considered patient preferences for pharmacological intervention, safety, access and availability of pulmonary rehabilitation programmes to be important. They also noted that currently patients will attempt to self-medicate by purchasing their own oxygen concentrators at their own expense, however, often these concentrators will not provide the high flow rates required by an IPF patient. Improved oxygen management would hopefully reduce this occurrence. Relaxation via an OT is useful and psychology referral for diagnosis acceptance is often useful for the patients.

	People with IPF may be breathless due to multiple factors that include hypoxia, co-existing COPD, co-existing pulmonary hypertension, anxiety and deconditioning. Appropriate assessment of the breathless patient should identify the cause of breathlessness. The GDG considered other relevant NICE guidance such as the Lung cancer NICE clinical guideline 121 (2011) and the Chronic obstructive pulmonary disease NICE clinical guideline 101 (2010) when making recommendations for best supportive care. Research recommendations The GDG agreed that the lack of evidence for the use of oxygen therapy for people with IPF justified developing a research recommendation to address whether short-burst, ambulatory and nocturnal oxygen therapy improves outcomes for people with IPF. For further information on the research recommendations see Appendix P.

Recommendations	**21. Assess the oxygen needs of people who have been hospitalised with idiopathic pulmonary fibrosis before they are discharged.**
Relative values of different outcomes	The GDG considered the critical outcome for this recommendation to be improvements in health related quality of life. Oxygen management and 6MWD were also considered important outcomes for measuring prognosis at regular intervals in a patients care pathway, in order to determine the optimal times when IPF disease progression should be reviewed.
Trade-off between clinical benefits and harms	Respiratory hospitalisation with IPF is usually associated with worsening breathlessness and increased requirement for supplementary oxygen. The GDG acknowledged there is unlikely to be any major improvement in breathlessness following exacerbation in people with IPF or those who are admitted to hospital for a respiratory cause (except in specific cases where the cause is treatable or reversible e.g. pulmonary embolism, pneumothorax). It is therefore likely that oxygen requirements will need to be reassessed following hospitalisation in order to achieve the benefits of appropriate oxygen management. It was recognised that there may be harms associated with inappropriate oxygen management if patients do not get followed-up in a timely manner and if there has been a change in symptoms due to disease progression or acute exacerbation. Currently, there may be a minority of patients who are discharged home and have oxygen requirements reassessed when they are back in the community, which carries a risk that for these patients oxygen management may not be optimal. Reassessing oxygen requirements prior to discharge brings no appreciable harm and will allow for optimal oxygen management following a potential change in the clinical status of the patient, thereby bringing clinical benefit and reducing the risk of clinical harm.
Economic considerations	There was no economic evidence identified to inform this recommendation. The GDG acknowledged the unit cost of oxygen assessment and that a new assessment would be most cost effective at a time point when the clinical status and need for oxygen had changed. The GDG agreed that they could not support a recommendation specifying exactly when a referral for oxygen assessment should

	occur, due to the lack of available evidence to inform the optimal timing or cost effectiveness of oxygen management in relation to disease progression. The event of hospitalisation for a respiratory cause, however, carries a high probability of need for reassessment in order for the benefits of oxygen to be realised, and therefore oxygen reassessment at this time point (that is,. prior to discharge) is likely to be a cost effective strategy. The GDG acknowledged that further monitoring of oxygen requirements could also potentially be cost effective strategy given the unpredictability of disease progression, however conceded there was no evidence to support in favour or against. The increased cost of monitoring at follow-up may be offset, albeit to a lesser extent, by identifying inappropriate use of oxygen. For example, people with advanced disease may be less active outside the home and may still be being prescribed expensive liquid oxygen when on assessment of the individual's needs and circumstances a cylinder or no ambulatory oxygen may be considered to be more appropriate.
Quality of evidence	This recommendation was based on GDG consensus, as no evidence was retrieved to inform this question.
Other considerations	The GDG considered the personal experiences of the patient members of the guideline group. Discussions included consideration of the following: Reassurance of monitoring of disease progression by specialist health professionals with expertise in ILD (this may be someone who runs a service seeing at least 500 ILD patients per year or has completed specialist training in ILD for at least 6 months). Experiences of availability and components of pulmonary rehabilitation . Meeting other people with IPF and advice of support groups. Coherent and concise information booklets and involvement of relatives when diagnosis is given (as patient does not often take in information at time). Warning not to access internet information immediately as it can be misleading. The GDG also acknowledged guidance on supplemental oxygen therapy published by the BTS working group [10], which states that patients with persistent resting hypoxaemia PaO_2 at or below 7.3 kPa (55 mm Hg) or below 8 kPa with clinical evidence of PH and who are breathless should be considered for palliative oxygen at home delivered by oxygen concentrator and that these individuals may also benefit from ambulatory oxygen if they remain active outside the home. However, the GDG also agreed that this is an area which warrants further research, due to the lack of evidence to show that quality of life or disease progression is improved by oxygen therapy. Research recommendation The GDG agreed that the lack of evidence for the use of oxygen therapy for people with IPF justified developing a research recommendation to address whether short-burst, ambulatory and nocturnal oxygen therapy improves outcomes for people with IPF. For further information on the research recommendations, see Appendix P.

Recommendations	**22.If the person has a cough consider:** • **treatment for causes other than idiopathic pulmonary fibrosis (such**

	as gastro-oesophageal reflux disease, post-nasal drip) • **treating with opioids if the cough is debilitating** • **discussing treatment with thalidomide[f] with a consultant respiratory physician with expertise in interstitial lung disease if the cough is intractable.**
Relative values of different outcomes	The GDG considered the critical outcome for this recommendation to be improvements in health related quality of life. Improvements in psychosocial health and cough. were considered to be important outcomes to inform this recommendation.
Trade-off between clinical benefits and harms	The GDG discussed the harms and benefits associated with symptom relief for cough. The GDG also considered the harms and benefits of thalidomide and prednisolone. They agreed that routine use of prednisolone was not appropriate for the symptomatic relief for cough, due to the lack of evidence showing any clear benefits for improving cough. The GDG agreed that there was sufficient evidence to alert physicians and patients in the UK of the possible benefits of using thalidomide for intractable cough. The precautions for thalidomide use were acknowledged as were the uncertainties regarding the long term harms of thalidomide for cough, which are unknown. The GDG agreed that contact should be made with a specialist chest physician with expertise in ILD regarding its use in IPF, as there is no known alternative treatment for cough on the few occasions when it could be debilitating. The GDG also discussed that the patient would have been offered opiates and probably anti reflux therapy, and may be on either or both but that thalidomide would essentially be used on its own (with or without those treatments) rather than being a combination with melphalan or prednisolone which is part of the same treatment regimen licensed for use in multiple myeloma as indicated in the BNF. Therefore, they recognised that thalidomide is unlicensed for use as a single drug to treat cough in people with IPF. The GDG discussed that thalidomide would be used rarely and would likely be prescribed by consultant respiratory physicians with an expertise in ILD named-patient conditions in order to support its managed uptake.
Economic considerations	No published economic evidence was identified to inform this question. The GDG considered the cost of commonly used pharmacological agents for the symptom relief of cough. The GDG agreed they could not support a recommendation in favour of routine use of thalidomide due to its relatively high acquisition cost and potential adverse effect profile. It was noted that it is also excluded from PBR tariff so use would require separate negotiation for payment at local level with strategic commissioning. The GDG agreed that if the potential clinical benefits of thalidomide could be realised using specialist expertise to identify patients whom would benefit most, it may be a cost effective means in improving quality of life for some people with IPF and intractable cough. Prednisolone was also considered. Given the adverse effect profile of prednisolone, alongside the additional need (and cost) of monitoring, there was not sufficient

[f] At the time of publication (June 2013), thalidomide did not have a UK marketing authorisation for this indication. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

	evidence of benefit to support a recommendation of this agent in the management of cough. The GDG noted that opiates had a relatively low acquisition cost compared to other available treatments for cough.
Quality of evidence	This recommendation was partly based on GDG consensus due to the lack of evidence regarding palliation of cough. Four studies were identified for the palliation of cough (low to very low quality and there was uncertainty in the effect). One study investigated the use of prednisolone; the primary aim of this study was to assess the responsiveness of IPF patients to cough inducing agents. A small sub set of patients were treated with prednisolone to investigate how the cough response is affected, but due to the lack of a direct comparison, a meta-analysis could not be carried out. Three studies investigated the use of thalidomide for the palliation of cough, two of these studies were abstracts relating to unpublished trial data and one was a two-treatment two-period cross over trial. The GDG noted the Horton 2012 trial, which has been important in alerting physicians and patients in the UK of the potential benefit of thalidomide. The main limitation was the small sample size and that the treatment was too short to actually know what the harms of thalidomide are likely to be if continued. Again, due to a lack of data reported, a meta-analysis could not be conducted. The GDG agreed that these studies indicated that thalidomide may be beneficial in treating intractable cough. The GDG agreed that it was appropriate to include abstracts retrieved as evidence to inform this review question, due to the lack of evidence found. The unit costs presented were from publically available list prices and the dosages validated by the GDG.
Other considerations	The GDG considered patient preferences for pharmacological intervention to be important considerations, whilst highlighting the potential side effects associated with medication. The placebo effect of giving a syrup or solution was discussed as a potential benefit in the management of mild cough, as was the need for effective management of debilitating cough characteristic of IPF was. Various cough syrups (including codeine, pholcodine, dextromethorphan etc.) are available but there is no evidence to recommend one over another. Of these, pholcodine and dextromethorphan may have fewer side effects. The GDG considered other relevant NICE guidance such as the Lung cancer NICE clinical guideline 121 (2011) and the Chronic obstructive pulmonary disease NICE clinical guideline 101 (2010) when making recommendations for best supportive care. Research recommendation The GDG agreed that the preliminary evidence included in this review indicated that pharmacological therapies may be of benefit in controlling intractable cough associated with IPF. Therefore, the GDG agreed to develop a research recommendation to address the value of pharmacological therapies to treat intractable cough associated with IPF. For further information on the research

	recommendations, see Appendix P.

Recommendations	**23. Ensure people with idiopathic pulmonary fibrosis, and their families and carers, have access to the full range of services offered by palliative care teams. Ensure there is collaboration between the healthcare professionals involved in the person's care, community services and the palliative care team.**
Relative values of different outcomes	The GDG acknowledged the most important outcomes for this recommendation to be improvements in quality of life measures, breathlessness, cough and psychosocial health.
Trade-off between clinical benefits and harms	The GDG agreed that currently not all people with IPF have access to the services offered by multidisciplinary palliative care teams. It was discussed that much of the symptom relief is provided as part of best supportive care and that in the majority of cases the ILD team will suffice. There was the recognition that patients can demonstrate serious adverse events profiles with the pharmacological interventions and may require withdrawal of ineffective therapies and thus may need the expertise of the ILD team to tailor alternative regimens for patients. Therefore, the GDG recognised the importance of ILD teams remaining involved in a patient's care even once they have been referred to the palliative care teams. The benefits of delivering continued care was thought to give people with IPF and their carers the feeling of control of their well-being, whilst improving their psychosocial health. There were no appreciable harms associated with palliative care, but poor quality of life was linked with lack of access of continued care.
Economic considerations	There was no published economic evaluation to inform this recommendation. The GDG noted that referral to palliative care is not universal practice and there could be a cost impact for the NHS with increased referrals. The cost effectiveness of the services provided by the palliative care teams was not examined as part of this guideline, so the cost effectiveness of the recommendation remains uncertain. Noting there were no appreciable harms, the GDG were in consensus that increased continuity of care and collaborative working between speciality teams was likely to improve outcomes of best supportive care interventions and the recommendation was likely to allow for cost effective clinical practice.
Quality of evidence	This is a consensus recommendation drawn up by the GDG on consideration of the of the wider management options available for patients.
Other considerations	The GDG considered patient preferences for pharmacological interventions for symptom relief, access and availability of community and palliative care services to be key in the management of people with IPF. The GDG recognised that these services differ according to region and discussed that the palliative care teams should ideally include input from the following services or health professionals if available: Hospice day care (ideally with adequate oxygen provision) Community nurses

	McMillan nurses Social support The GDG also discussed the importance of GP awareness of patient's with IPF to manage co-morbidities and recognising the need for timely referral to palliative and social services. The GDG considered other relevant NICE guidance such as the Lung cancer NICE clinical guideline 121 (2011) and the Chronic obstructive pulmonary disease NICE clinical guideline 101 (2010) when making recommendations for best supportive care.

10 Psychosocial support

10.1 Review Introduction

It is generally believed that there is a combination of factors that are responsible for psychosocial wellbeing. These can be divided into psychological (including thoughts, emotions, feelings and behaviours), social interaction, the environment, culture, traditions, roles within the family and society.

Individuals who have psychosocial wellbeing feel they have a role within the family and society that strengthens their perception of self and enhances their self-esteem. Psychosocial support is usually provided by on-going nurturing, unconditional relationships with family and friends. Psychosocial support is important at helping individuals to cope and manage with a threat or crisis. Given the lack of treatment options and the rapidity of functional limitation, people with IPF need psychosocial support if they are to mobilise their internal resources to adjust, cope and manage. People with IPF often find that they have lost the life they had and face an uncertain future. Functional limitation means they cannot perform the roles they once had, threatening their relationships and self-esteem. Specialist nurses are important adjuncts to providing psychosocial support by 'being there' for the patient, being easily accessible, knowledgeable and understanding, with information and advice to support both patient and carer. There are currently very few specialist ILD nurses nationally supporting patients with IPF. Good psychosocial support is likely to help patients adjust, manage and work things out for themselves, preventing escalation of problems that might require specialist psychological or social intervention. Some patients currently get psychosocial support from peers in a support group, but access to these is patchy throughout the country and difficult due to transport issues and dependence on oxygen. Assessment of psychosocial wellbeing is an important aspect of best supportive care.

10.2 Clinical question and review methodology:

The following clinical question was included in this chapter:

10.2.1 What is the specific type of psychosocial support and information that should be provided for patients diagnosed with IPF?

For full details see review protocol in Appendix C.

Table 46: PICO characteristics of review question

Population	Adults with confirmed IPF and/ or ILD
Intervention/s	• Psychosocial support • Patient information
Comparison/s	None
Outcomes/ Evaluation	Critical outcomes • Improvement in health-related quality of life Other outcomes • Dyspnoea • Improvement in psychosocial health (including depression)
Study design	Any

The objectives of this review were to determine what psychosocial support and information should be provided for patients diagnosed with IPF. No restrictions were put on sample size or study design, the population was extended to include all ILD patients in order to capture all relevant data.

10.3 Clinical evidence

Three studies were included in the review[16,71,112]. We searched for all papers studying the impact of psychosocial support in patients with IPF.

The population included patients with ILD, with a view that IPF patients would be present in the sample. Studies which only looked at specific ILD populations such as sarcoidosis were excluded as there were no people with IPF present in the sample. Also patients who suffer from sarcoidosis have a better prognosis and different treatment regimen.

One RCT[71] was identified which investigated the impact of a psychosocial support intervention in patients with IPF and their care partners (people who live with or care for the patient with IPF, as defined in study), which is presented separately in this report. This study included both quantitative and qualitative analysis. Two questionnaire surveys were also identified [16,112] which gave data on patients' experiences and needs.

Quantitative data was analysed using meta-analysis and the quality was assessed using GRADE. Qualitative data was summarised and the quality was assessed using the NICE qualitative studies checklist, taking into account biases related to qualitative study designs.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

10.3.1 Summary of included studies

Table 47: Summary of studies included in the review

Study	Intervention/topic areas surveyed	Population	Outcomes	Comments
Collard 2007[16]	**Patients experiences and opinions of:** • Education and resources • Experience with diagnosis • Experience with treatment	Pulmonary fibrosis -Patients and carers of current and deceased patients.	Patient experiences	Sampling: self-identified no confirmation of diagnosis, non-probability sampling-sampling bias. Generalisability – external validity. Large proportion of non-responses- response rate 50%. Responder bias: responders may be substantially different to non-responders. Recall bias. Misinformation bias.
Lindell 2010[71]	**Program to Reduce Idiopathic Pulmonary Fibrosis Symptoms and Improve Management (PRISIM) intervention** 6 weekly group sessions	People with IPF recruited from a university based ILD programme.	Dyspnoea Anxiety Depression Perceived	Reporting of outcomes: pre scores reported more fully than post treatment scores (graphical data only).

Study	Intervention/topic areas surveyed	Population	Outcomes	Comments
	attended by patients and care partners. **Vs. usual care**: • Seen by members of the clinical care team every 3 to 6 months. • Pulmonary clinical nurse specialist was available by phone to answer questions and conducted a monthly support group for those wanting to attend. • Psychological counselling was provided if indicated but was not offered on a routine basis.		stress QoL	Small sample size. Discrepancy in method of diagnosis between the two groups - ATS/ERS criteria not used.
Shoenheit 2011 [112]	**Patients experiences and opinions of:** • Diagnostic pathway • Diagnosis • Quality of care in treating centre • Patients aims for disease management • Commonly reported unmet medical needs • IPF on patients' quality of life and emotional well-being	People with IPF-physician confirmed diagnosis.	Patient experiences	IPF diagnosis not confirmed using current criteria. Recall bias. Small sample size. Generalisability.

10.3.2 Study quality and summary of findings

Table 48: Clinical evidence profile: PRISIM vs. usual care (patients)

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PRISM-Patients	Control	Relative risk (95% CI)	Absolute risk	
Dyspnoea (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	11	-	MD 0.37 lower (19.76 lower to 19.02 higher)	Very low
Anxiety (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Serious[6]	None	10	11	-	MD 6.57 higher (0.63 to 12.51 higher)	Low
Depression (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	11	-	MD 0.27 higher (3.49 lower to 4.03 higher)	Very low
Perceived stress (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	11	-	MD 1.12 higher (2 lower to 4.24 higher)	Very low
QoL: SF36 Physical (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Serious[6]	None	10	11	-	MD 4.98 lower (8.94 to 1.02 lower)	Low
QoL: SF36 Mental (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	11	-	MD 0.37 higher (1.95 lower to 2.69 higher)	Very low

[1] *Blinding is not reported*
[2] *Reporting of outcomes: pre scores reported more fully than post treatment scores (graphical data only)*
[3] *Small sample size*
[4] *Discrepancy in method of diagnosis between the two groups- ATS/ERS criteria not used*
[5] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*
[6] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID

Table 49: Clinical evidence profile: PRISIM vs. usual care (care partners)

Quality assessment							No of patients		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**PRISIM-Care partners**	**Control**	**Relative risk (95% CI)**	**Absolute risk**	
Anxiety (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Serious[6]	None	10	10	-	MD 2.11 lower (5.46 lower to 1.24 higher)	Low
Depression (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	10	-	MD 0.51 lower (3.39 lower to 2.37 higher)	Very low
Perceived stress (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Serious[6]	None	10	10	-	MD 3.38 lower (5.73 to 1.03 lower)	Low
QoL: SF36 Physical (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	10	-	MD 1.19 lower (6.2 lower to 3.82 higher)	Very low
QoL: SF36 Mental (Better indicated by lower values)											
1	Randomised trials	Serious[1,2,3,4]	No serious inconsistency	No serious indirectness	Very serious[5]	None	10	10	-	MD 0.47 higher (0.99	Very low

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	PRISIM-Care partners	Control	Relative risk (95% CI)	Absolute risk	
										lower to 1.93 higher)	

[1] *Blinding is not reported*
[2] *Reporting of outcomes: pre scores reported more fully than post treatment scores (graphical data only)*
[3] *Small sample size*
[4] *Discrepancy in method of diagnosis between the two groups- ATS/ERS criteria not used*
[5] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*
[6] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID

Table 50: Narrative summary of qualitative data extracted from survey studies

Topic surveyed	Study	Quotes from studies	Summary
Diagnostic pathway, diagnosis and experience with diagnosis.	Shoenheit 2011[112]	58% (26 of 45) of patients reported that 'protracted route' to confirm diagnosis was characterized by an initial dismissal of the presenting symptoms, with repeated physician visits for further evaluation and testing. This was commonly interrupted by an 'acute event,' which was often initially attributed to other causes this frequently resulted in an emergency room visit and subsequent hospitalization, where a detailed evaluation by a chest physician would eventually result in a diagnosis of IPF. This process reportedly took as long as 2–12 years, despite repeated visits to healthcare practitioners during this period. Patients who were subjected to this protracted route to diagnosis were often critical of the care they received, citing both a lack of empathy and emotional support and an apparent lack of competence among healthcare practitioners. There was a tendency among these patients to perceive the initial diagnosis not as a working hypothesis but rather as an erroneous or missed diagnosis. In a minority of cases (16%), the diagnosis was made within a month of the patient's initial presentation. Early detection was attributed to a well-informed patient researching their symptoms online, a well-informed physician detecting the distinctive 'Velcro1 rates' on chest auscultation, which prompted further evaluation for possible interstitial lung disease, or routine surveillance of known drug toxicities (e.g. amiodarone).	Diagnostic pathway, lengthy process (1- 12 years). Lack of empathy and emotional support & apparent lack of competence among healthcare practitioners. Erroneous or missed diagnosis: often saw > 1 physician & sought second opinion. Early diagnosis (≤ 1 month): well informed patient (researched symptoms). Consultation with chest physician deemed essential for accurate diagnosis: diagnosis, insensitivity and duration too short.

Topic surveyed	Study	Quotes from studies	Summary
	Shoenheit 2011[112]	"Patients expressed dissatisfaction with the manner in which the diagnosis was divulged, citing insensitivity on the part of the healthcare practitioner and insufficient time during the consultation to address the full range of patients' questions and concerns".	
	Collard 2007[16]	54.6% reported at least a 1 year delay between earliest indications of a potential breathing problem and the diagnosis of IPF. 38.2% reported seeing two or more physicians before a diagnosis of IPF was established. 53.2% sought a second opinion. 84.4% consulted a chest physician at some stage during their diagnostic evaluation.	
Quality of care in tertiary centre and quality of care in community practice.	Shoenheit 2011[112]	"Patients treated in a tertiary care centre consistently reported greater satisfaction with the quality of care, the availability of treatment options (including enrolment in a clinical trial), the knowledge and expertise of healthcare practitioners, and the frequency of follow-up visits and routine monitoring. Additionally, patients treated in a tertiary care centre commonly reported that the opportunity to interact with other IPF patients provided important benefits, including psychological support and practical disease management tips".	Quality of care in treating centre: greater satisfaction with care reported from patients treated in tertiary centre compared with community practice.
	Shoenheit 2011[112]	"Patients treated in the community practice setting consistently reported infrequent follow-up visits (typically once per year), short duration of visits (generally less than 10 minutes), and a lack of available treatment options. In general, these patients were less well informed about their disease and the available treatment options, including pulmonary rehabilitation, lung transplantation, and enrolment in a clinical trial".	
Commonly reported unmet medical needs and educational resources.	Shoenheit 2011[112]	Improved access to 'Centres of Excellence'. Clear and understandable disease education resources. Comprehensive family support/counselling programs. Fewer bureaucratic barriers to scheduling specialist appointments and obtaining supplemental oxygen. Patient advocacy and public education. Improved diagnostic techniques. More effective treatment options.	Education and support: clear and understandable disease education resources including information on treatment options (pharmacological and non-pharmacological), comprehensive family support/counselling programs, improved patient advocacy and public education.
	Collard 2007[16]	63% somewhat/ strongly agreed with the statements there was a clear lack of	

Topic surveyed	Study	Quotes from studies	Summary
		information and resources about IPF. 51.2% reported being generally/very well informed regarding the treatment options available at the present time. 38.7% reported being generally/very well informed regarding the benefits of pulmonary rehabilitation. 42.5% reported being generally/very well informed regarding the benefits of managing supplemental oxygen. 32.5% reported being generally/very well informed regarding the risks and benefits of lung transplantation.	
Patients aim for disease management.	Shoenheit 2011[112]	"Focused on disease stability and efforts to slow progression if feasible. For only a small minority of those surveyed, the emphasis was either on lung transplantation as the 'hope' for future survival beyond IPF, or some belief that their particular condition was atypical and associated with a less dire prognosis"	Patients aim for disease management: disease stability and slow progression, acceptance.
QoL and emotional well-being.		*Personal independence:* Loss of independence that coincided with the deterioration in health and inability to perform routine daily tasks. The requirement for supplemental oxygen was commonly identified as a milestone in the patients' loss of independence, as it is at this point that the disease becomes highly visible to others and excursions outside the home begin to require significant logistical planning. In many cases, this loss of independence has a notable impact on the patient's emotional well-being, as they begin to perceive themselves as a burden to both their family and society. *Relationships with others:* Considerable difficulty in continuing relationships with friends and acquaintances, due to their worsening pulmonary status and immobility, as well as a general lack of awareness and understanding of the disease. *Financial status:* 20% of respondents reported financial difficulties as a result of their inability to work and the consequent reduction in income. This further served as a stressor, as well as the concern that they were now an increasing burden to their families and loved ones.	QoL: loss of personal independence, loss of personal relationships, financial difficulties.
Experience with treatment.		74.7% of respondents reported current pharmacologic therapy for IPF Common reasons for not receiving pharmacologic therapy were:	Experience with treatment: effective treatment options, lack of referrals for pulmonary,

Topic surveyed	Study	Quotes from studies	Summary
		fear of side effects 26% ineffectiveness of therapy 23% no treatment prescribed 24% early/stable disease 22% respondents reporting use of herbs or nutritional supplements 24.4% oxygen use was reported by 61% pulmonary rehabilitation and physical therapy referrals were reported by a minority of patients 31.8% and 23.9% behavioural health counselling referrals were uncommonly reported 58.7% had transplantation discussed with them.	rehabilitation physical therapy and behavioural health counselling.

10.4 Economic evidence

Published literature

No relevant economic evaluations comparing strategies of psychosocial support or patient information for people with IPF were identified. No studies were selectively excluded.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs of the staff that may be involved in providing psychosocial support are provided below to aid consideration of cost effectiveness. The Expert Patients Programme mentioned in a patient member's testimony has a unit cost of £289 per patient (PSSRU 2011[21]).

Table 51: Unit costs for per hour of patient contact for clinical staff that may provide psychosocial support. Source PSSRU 2011[21]

Cadre of staff	Unit cost	Notes
General practitioner	£186	The difference in cost of personnel of the same grade but working in different settings and role is due to ratio of direct to indirect time of patient contact.
Medical consultant	£162	
Clinical psychologist	£135	
Band 6 hospital nurse	£122	
Band 7 community nurse specialist	£91	
Primary care counsellor	£66	
Band 6 GP practice nurse	£51	

10.5 Evidence statements

Clinical interventions:

Low to very low quality evidence with high levels of uncertainty looked at patients partaking in a programme to reduce IPF symptoms and improve management (PRISIM) intervention. The programme group reported less positively on the majority of outcomes measured including; anxiety (15.13±6.92 vs. 8.56±6.95), depression (9.71±4.34 vs. 9.44±4.35), perceived stress (19.32±3.64 vs. 18.20±3.65) and physical QOL domains (31.06±4.61 vs. 36.04±4.63) compared to patient who received usual care. There was no difference found between the groups for dyspnoea (49.51±22.64 vs. 49.88±22.64) and mental QOL domains (55.98±2.1 vs. 55.61±2.71). However post study interviews showed that patients who had partaken in the programme felt less isolated, were able to put their disease into perspective and valued participating in research which would help others (one study, n=42).

Surveys:

A narrative summary of two surveys (total n=1493) investigating opinions and experiences of patients with IPF is provided below, as it was not possible to pool results. This was low to very low quality evidence.

- Diagnostic pathway: patients who had a lengthy diagnostic process highlighted a lack of empathy emotional support, deemed healthcare professionals who dealt with them to be incompetent, and often sought a second opinion. Patients who had an early diagnosis were

usually well informed themselves or treated by a physician who was aware of the condition. It was generally felt that a consultation with a chest physician is essential for an accurate diagnosis and diagnostic consultations were too short and physicians were insensitive.

- Quality of care in treating centre: patients reported greater satisfaction with care when treated in tertiary centres compared with community practice. Patients felt the need to have improved access to Centres of Excellence, fewer bureaucratic barriers to scheduling specialist appointments and obtaining supplemental oxygen and improved diagnostic techniques.
- Education and support: patients felt the need to have clear and understandable disease education resources including information on treatment options (pharmacological and non-pharmacological treatments), comprehensive family support/counselling programmes and improved patient advocacy and public education**.**
- Quality of life: patients report a loss of personal independence, loss of personal relationships and financial difficulties
- Experience with treatment: patients felt the need to have more effective treatment options and there was a lack of referrals for pulmonary rehabilitation, physical therapy and behavioural counselling.
- Patients aims for disease management: the majority of patients reported their primary aim for disease management was disease stability and to slow progression. However a minority of patients still had problems with accepting their prognosis and hoped for a miracle cure or cure through lung transplantation.

Economic

- No economic evidence regarding strategies of psychosocial support or patient information was identified.

10.6 Recommendations and link to evidence

Recommendations	**24. NICE has produced guidance on the components of good patient experience in adult NHS services. Follow the recommendations in Patient experience in adult NHS services (NICE clinical guideline 138).** **25. An interstitial lung disease specialist nurse should be available at all stages of the care pathway to provide information and support to people with idiopathic pulmonary fibrosis and their families and carers with the person's consent.** **26. Offer advice, support and treatment to aid smoking cessation to all people with idiopathic pulmonary fibrosis who also smoke, in line with Smoking cessation services (NICE public health guidance 10).**
Relative values of different outcomes	The GDG considered improvements in mental and physical quality of life to be the critical outcomes. These outcomes were described in qualitative and quantitative studies, where patient's experiences, preferences and perceptions were reported.
Trade-off between clinical benefits and harms	The importance of continued support and continuity of care, alongside appropriate information and management of expectations for patients and carers was emphasised in discussion relating to the diagnostic, prognostic, and the review and monitoring recommendations.. Regular review allows a feeling of contact with health services and a feeling for the patient that they have not "been forgotten".
Economic considerations	There was no economic evidence identified to inform this review question. In forming this recommendation the GDG considered the setting and cadre of staff that

	should provide psychosocial support that was tailored to people with IPF. To do this they considered the unit costs presented in this chapter, alongside experiences of the patient members. Additionally, the GDG considered evidence and revisited the points raised when discussing the optimal timing of when psychosocial support should be given. The unit costs presented were from an NHS perspective and accepted to be a valid estimate by the GDG. It was noted that in current practice, attendance for tests undertaken for prognostic purposes has been potentially filling a void of regular contact, however, this may not be the most effective or cost effective means of providing for this patient need. Having a named member of the specialist team, i.e. a specialist ILD nurse, whom the patient could contact on the telephone for this support and information was felt to be a more appropriate use of resource than direct self-referral for a specialist or primary care appointment (which carries a higher unit cost per hour of patient contact than for hospital band 6 nurses) and could potentially allow a means of identifying particular patients where increased frequency of follow up was appropriate due to an unexpected decline. Given the lack of high quality clinical evidence comparing different strategies and the issues reported by observational studies, the GDG made a qualitative judgement that the benefit of involving an ILD nurse throughout the care pathway would be a cost effective use of healthcare resource, especially given the additional benefits of their involvement in other aspects of care (i.e. at diagnosis, giving information at prognosis, best supportive care referral).
Quality of evidence	This recommendation was mainly based on GDG consensus. Evidence was derived from one intervention study and two surveys, of very-low to low quality. This was due to small sample sizes, blinding not being reported, lack of information regarding post treatment scores (graphical data only), and discrepancies between methods of diagnosis used (ATS/ERS criteria not used). The intervention study showed that post-study the experimental group reported less positively on all outcomes measured including; anxiety, depression and perceived stress compared to the control group. However post study interviews showed that patients who had the intervention felt less isolated, were able to put their disease into perspective, and valued participating in research which would help others. Topics covered in the two included surveys were: diagnostic pathway; quality of care in treating centre; education and support; quality of life; and experience with treatment.
Other considerations	The GDG agreed that it was of crucial importance for patients to have access to continued support and reassurance of continuity of care alongside the provision of appropriate information in terms of the management of IPF. Given that many people with IPF will move on to receive best supportive care, the GDG agreed that IPF patients should have a named member of the specialist team to contact. The GDG also considered the personal experiences of the patient members of the guideline group regarding psychosocial support. Discussions included consideration of the following: Patient's emotions on receiving a diagnosis of IPF. Experiences of availability and components of pulmonary rehabilitation. Including a psychosocial element as well as education regarding; diet; exercise; social support; and benefits. Contact details of Specialist nurses and support groups provided at diagnosis. Coherent and concise information booklets and involvement of relatives when diagnosis is given (as patient does not often take in information at time). Warning not to access internet information immediately as it can be misleading.

	The GDG also discussed the importance of people with IPF and their carers receiving information from ILD specialists throughout their care and the reassurance felt in having an appropriate healthcare professional to contact for support. In the community, follow-up may be provided by district nurses and in such cases the GDG identified the importance of communication between these health professionals to ensure appropriate monitoring and care is provided. The GDG considered the patient experience in adult NHS services (NICE clinical guideline 138) and smoking cessation services (NICE public health guidance 10) when making recommendations for psychosocial support. Guidance in these areas was agreed to further emphasize good communication between health professionals and people with IPF, as well alert health professionals to the importance of providing smoking cessation advice where required.

11 Pharmacological interventions

11.1 Review Introduction

Idiopathic Pulmonary Fibrosis (IPF) has a deleterious impact on health status, quality of life and carries a poor prognosis. There is thus a need for effective therapies to improve the outcome for people with this condition. Unfortunately, the development of such therapies is impaired because the pathogenesis of IPF remains uncertain. Despite this limitation, a number of therapies have traditionally been widely used to treat IPF in clinical practice. These include agents that suppress pulmonary inflammation, in the belief that lung inflammation is the force driving lung fibrosis and agents that inhibit production and / or deposition of connective tissue in the lung interstitium.

The conduction of clinical trials has also faced difficulties both in terms of patient selection and choice of clinically meaningful end-points. The majority of patients with IPF are over the age of 65 and a significant number have co-morbidities. They are therefore unlikely to be fit enough to undergo a surgical lung biopsy to consolidate the diagnosis. Hence, trials which have inclusion criteria based on diagnosis by surgical lung biopsy are likely to be biased towards selecting a younger and fitter sub-group of patients. Conversely, trials which accept less strict diagnostic criteria might potentially include patients with other diagnoses.

There is currently debate about how to choose clinically meaningful end-points in trials of pharmacological treatments in IPF both in terms of demonstrating efficacy and detecting adverse effects. Whilst significant change in all-cause mortality might superficially appear to be the 'gold standard' in this regard, it is likely to be impractical in terms of the large number of patients who would need to be enrolled and length of time required for follow-up [101,130]. For this reason, for large trials, serial group change in FVC over a minimum of 12 months is considered by many as an acceptable and practical marker of disease progression. However it is not known if change in FVC is a true surrogate for mortality in IPF. In individual patients, serial trend in FVC may also be the most effective way to confirm disease decline, stability, or, incremental improvements.

As with all therapeutic interventions, clinicians treating IPF with pharmaceutical agents must balance any benefits with short and longer-term side-effects.

11.2 Clinical questions and review methodology

The following clinical questions were included in this chapter.

11.2.1 Which drug should be initiated first, for how long, and in what combination in the treatment of IPF?

- **(sub-question) What is the clinical and cost effectiveness of pharmacological interventions to manage patients with suspected or confirmed IPF:**

- Ambrisentan
- Azathioprine
- Bosentan
- Co-trimoxazole
- Mycophenolate mofetil
- *N*-acetylcysteine
- Prednisolone

- Proton-pump inhibitors
- Sildenafil
- warfarin
- Combinations: prednisolone + azathioprine and prednisolone + azathioprine + N-acetylcysteine

For full details see review protocols in Appendix C

Dosages and licensing indications for the drugs covered in this review are presented in Table1. None of the drugs are specifically licensed for IPF, so no specific doses for IPF exist. Therefore, the licensing indications identified below are broad and based on speculation or small case studies. For warfarin the dosing even within its licensed indications is variable. For prednisolone and the other immunosuppressive agents, the dose will likely be "the lowest dose that the patient tolerates".

Table 52: Dosages and licensing indications (BNF 2012)

Group	Dosing	Licensed in IPF?	Licensed Indications
PPIs	Lansoprazole 15-30mg OD	No	Gastro-oesophageal reflux disease
	Omeprazole 20-40mg OD	No	Gastro-oesophageal reflux disease
N-acetylcysteine	600mg TDS	No	None (at this dosage form)
Warfarin	Variable according to INR	No	Treatment and prevention of VTE
Prednisolone	10-60mg OD according to response and adverse effects	Under broad license	Suppression of inflammatory and allergic disorders
Co-trimoxazole	960mg BD	No	Acute respiratory/urinary tract infections
Azathioprine	50-300mg OD	No	Prophylaxis of transplant rejection
			Steroid sparing or in place of steroids in autoimmune disease
Mycophenolate Mofetil	250mg-1g BD	No	Prophylaxis of transplant rejection
Sildenafil	20mg TDS	No	Pulmonary hypertension
Bosentan	125mg BD	No	Pulmonary hypertension
Ambrisentan	5mg-10mg OD	No	Pulmonary hypertension

Table 53: PICO characteristics of clinical question on which drug should be initiated first, for how long, and what combination in the treatment of IPF?

Population	Adults with confirmed IPF
Intervention/s	• Ambrisentan • Azathioprine • Bosentan • Co-trimoxazole • Mycophenolate mofetil • N-acetylcysteine • Prednisolone • Proton-pump inhibitors • Sildenafil • warfarin • Combinations: prednisolone + azathioprine and prednisolone + azathioprine + N-

	acetylcysteine,
Comparison/s	Other pharmacological treatments/ placebo
Outcomes	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates Other outcomes • Adverse events (please see adverse events table listed in Appendix N) • Dyspnoea • Change in percent predicted DLCO • Hospitalisations due to IPF complications, including IPF exacerbations • Improvement in health-related quality of life • Change in percent predicted forced vital capacity • Performance on sub-maximal walk test (distance walked and lowest SaO_2)
Study design	Randomised controlled trials and systematic reviews of RCTs

11.2.2 Which measures can be taken to minimise the occurrence/severity of adverse events when undergoing pharmacological treatment for IPF?

Table 54: PICO characteristics of clinical question on measures can be taken to minimise the occurrence/severity of adverse events when undergoing pharmacological treatment

Population	Adults with confirmed IPF
Intervention/s	Assessing Thiopurine S-methyltransferase (TPMT)
Comparison/s	Not assessing TPMT
Outcomes	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates Other outcomes • Adverse events (please see adverse events table listed in Appendix N) • Dyspnoea • Hospitalisations due to IPF complications, including IPF exacerbations • Improvement in health-related quality of life • Performance on sub-maximal walk test (distance walked and lowest SaO_2)
Study design	Randomised controlled trials and systematic reviews

The objectives of these reviews was to determine which drug should be initiated first, for how long, and what combination in the treatment of IPF as well as the measures that can be taken to minimise the occurrence/severity of adverse events when undergoing pharmacological treatment. No restrictions were used for sample size or publication date. Studies with indirect populations such as COPD were not considered, as they have different disease trajectories and are therefore not comparable to people with IPF.

11.3 Clinical evidence

We searched for randomised control trials and systematic reviews comparing the effectiveness of the pharmacological treatments listed above with placebo or other pharmacological treatments in patients with confirmed IPF.

No studies answered the question 'Which drug should be initiated first, for how long, and in what combination in the treatment of IPF?', but fourteen included studies were used to address the clinical effectiveness of these drugs. In all studies it was unclear what line of therapy patients were undergoing. The fourteen randomised control trials are summarised below.

Two Cochrane reviews were identified[107,118]. The Cochrane Review on corticosteroids for idiopathic pulmonary fibrosis did not yield any studies. The Cochrane review on non-steroid agents for IPF was updated in line with the drugs included in the guideline scope.

Four studies[50,63,89,101] presented outcomes which were not specified in the protocol, but the GDG agreed these outcomes were important for decision making, so they have been reported. These were: hazard ratio to mortality, time to death up to study end, hazard ratio to categorical decrease in lung function and time to IPF worsening/ disease progression/ death.

The GDG prioritised the most important adverse events by drug type at the beginning of development. Only these have therefore been reported (see Appendix N).

One study[23] used both intention to treat as well as available case analysis; the GDG considered that it was important to include both types of analyses therefore these have been reported in this review.

One unpublished study which provided evidence for co-trimoxazole was included in this evidence review and used by the GDG in their decision-making, but has been included in the evidence report as academic data in confidence and therefore the relevant data has been blacked out.

No papers were identified on the clinical effectiveness of TPMT testing.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

11.3.1 Summary of included studies

Table 55: Summary of studies included in the clinical evidence review

STUDY	INTERVENTION / COMPARISON	POPULATION	OUTCOMES	COMMENTS
Demedts 2005 [23]	Corticosteroids+ azathioprine+ *N*-acetylcysteine vs azathioprine+corticosteroids.	Patients with IPF.	Lung capacity (FVC) Gas transfer (DLCO) Mortality Adverse events	High drop-out rate (only 30% of randomised patients available for follow-up at 1 year). Some patients excluded after randomisation. ITT and ACA analyses used for FVC and DLCO.
Homma 2012[45]	Nebulised N-acetylcysteine 352.4mg bd versus nil N-acetylcysteine therapy.	Early stage (I or II) IPF patients aged between 50-79 years as diagnosed by ATS/ERS.	Lung capacity (FVC) Hospitalisations due to IPF complications (including IPF exacerbations) Dyspnoea	High risk selection bias: randomisation process and allocation concealment not described. No detail provided for differences between baseline groups. Not placebo controlled comparison= 'no treatment'.

STUDY	INTERVENTION / COMPARISON	POPULATION	OUTCOMES	COMMENTS
				Blinding methods and personnel not described. Only patients aged 50-79 included. Selective reporting of data. LOCF method used for analysis. Ten patient's data not analysed due to 'protocol violations, missing data etc.' Paper suggested there was no important difference between those excluded from analysis population between arms. Reason for dropouts not given and only a subset selectively analysed.
Jackson 2010 [52]	Sildenafil vs. placebo.	Patients with IPF.	Lung capacity (FVC) Gas transfer (DLCO) Adverse events 6MWT-distance walked Dyspnoea (Borg scale)	Unclear allocation concealment. Small sample size. Study of short duration 21.4% drop out rate in placebo arm. Some outcomes were unable to be meta-analysed as standard deviations were not reported.
King 2008[62]	Bosentan vs. placebo.	Patients with IPF.	6MWT- distance walked Adverse events Time to disease progression	Allocation concealment unclear. These results include data on patients who did not complete 12 months of treatment and for whom either a last observation carried forward or an imputed value of zero was used in the analysis.
King 2011[63]	Bosentan vs. placebo.	Patients with IPF.	Mortality Adverse events Dyspnoea Time to IPF worsening or death	None.
Kubo 2005[66]	Warfarin plus prednisolone vs. Prednisolone.	Patients with IPF admitted to hospital.	Mortality Number of re-hospitalisations, 1 year survival rates 3 year survival rates	Large dropout rate. Six people dropped out of the intervention group because they were afraid of side effects and disliked the extra blood tests required, one dropped out

STUDY	INTERVENTION / COMPARISON	POPULATION	OUTCOMES	COMMENTS
				due to purpura. Population included non-smokers and hospitalised patients, therefore bias towards acutely ill. Allocation concealment not reported. No patients treated with anticoagulant alone.
Noth 2012[89]	Warfarin vs. placebo Warfarin arm stopped early due to safety concerns.	People with IPF aged between 35 to 80, as diagnosed by ATS/ERS.	Mortality Hospitalisations due to IPF complications (including IPF exacerbations) Adverse events including bleeds	All disclosures presented on an online appendix and not in paper. High risk of attrition bias as trial stopped prior to completion for safety thus all available results analysed together and high overall dropout rate.
Panther 2012[50]	Prednisolone, Azathioprine and oral NAC versus placebo versus oral NAC (this arm of the study remains ongoing with no data presented) Combination therapy arm stopped early due to safety concerns.	Patients with IPF aged 35 to 85 with mild to moderate lung function impairment.	Mortality Hospitalisations due to IPF complications (including IPF exacerbations) Adverse events	Manuscript approved by Zambon pharmaceuticals prior to submission. Risk of Bias: serious: High risk attrition bias: no overall dropout rates given prior to discontinuation of combination therapy arm at 32 week interim analysis. Discontinuation rates given for individual drugs may be for same patient no time course given or actual number of dropouts related to toxicity at 32 weeks. ITT population studied. No description of blinding methods or personnel given.
Raghu 1991 [102]	Prednisolone+ azathioprine vs. prednisolone.	Newly diagnosed people with IPF.	Lung capacity (FVC) Gas transfer (DLCO) Survival probability Adverse events Mortality	Unclear allocation concealment. Patients were allowed to cross over between groups. ATS diagnostic criteria not used (CT not mandatory).
Raghu 2012 [101]	Ambrisentan vs. placebo.	IPF	Time to IPF disease progression Mortality Categorical decrease in lung	limited data available-abstract only.

STUDY	INTERVENTION / COMPARISON	POPULATION	OUTCOMES	COMMENTS
			function	
Shulgina 2013 [115]	Co-trimoxazole vs. placebo	Fibrotic idiopathic interstitial pneumonia	Mortality Lung capacity (FVC) Gas transfer (DLCO) Health-related Quality of life (SGRQ) 6MWT (distance walked and lowest SaO_2) Dyspnoea (MRC score)	Not all patients had IPF. Patients in the co-trimoxazole group may have had shorter disease duration.
Tomioka 2005[127]	*N*-acetylcysteine vs. bromhexine hydrochloride.	Patients with IPF who had not received any form of immunosuppressiv-e therapy.	Lung capacity (FVC) Gas transfer (DLCO) 6MWT (distance walked and lowest SaO_2)	Randomisation method unclear. Allocation concealment unclear. Small sample size. *N*-acetylcysteine administered as nebulised product rather than orally.
Zisman 2010 [135]	Sildenafil vs. placebo.	Patients with IPF in an advanced stage.	Adverse events Mortality	Blinding not reported. Findings are applicable only to patients with advanced IPF. Unknown whether the treatment effect was driven by a particular subgroup of patients (e.g., those with more severe pulmonary vascular disease). Small sample size. Study of short duration. Improvements in subjective outcomes, such as quality of life, may be due to incomplete masking. SD not reported for all outcomes.

11.3.2 Summary of quality of life data

The table below summarises the QoL data as reported in the papers; where possible GRADE was applied.

Table 56: Summary of QoL data

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
King 2008[62]	SF36 domains	Reported; "When asked to rate their general health during the study period compared with 1 year prior 42.4% (n=28) of bosentan treated patients had an improvement in SF36 health transition score compared with 28.4% (n=23) of placebo recipients – a relative risk of improvement in favour of bosentan of 1.49(95% CI, 0.96-2.33; p=0.084). Changes in seven of eight domains of the SF36 survey up to month 12 were in favour of bosentan treatment, with a significant treatment effect observed in bosentan observed in the domain "role emotional"(p=0.032)".	
	Total SGRQ	Bosentan: 45.7±18.1 Placebo: 45.2±19	6 months follow up Bosentan: 45±21.3 Placebo: 47.8±21.7 12 month follow up Reported: "mean treatment difference up to month 12 continued to favour bosentan but were smaller (data not shown)".
King 2011[63]	SF36 domain: physical functioning	Bosentan: 61.1±25.4 Placebo: 58.2±24.9	1 year follow up Bosentan: 55.7±28.9 Placebo:52.8±27.6
	SF36 domain: physical role functioning	Bosentan: 63.1±30.0 Placebo:59.2±29.0	Bosentan: 58.5±32.4 Placebo:57.4±30.9
	SF36 domain: vitality	Bosentan: 55.5±21.9 Placebo:52.3±22.4	Bosentan: 51.6±24.4 Placebo:50.0±24.1
	SF36 domain: bodily pain	Bosentan: 69.9±26.5 Placebo:68.4±27.8	Bosentan: 64.3±31.1 Placebo:62.0±30.0
	SF36 domain: general health perceptions	Bosentan: 52.1±21.5 Placebo:48.7±20.0	Bosentan: 47.4±24.1 Placebo:46.9±22.9
	SF36 domain: social role functioning	Bosentan: 77.6±24.3 Placebo:72.5±27.1	Bosentan: 72.9±30.5 Placebo:69.3±29.7

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
	SF36 domain: emotional role functioning	Bosentan: 79.3±26.2 Placebo: 74.7±29.0	Bosentan: 73.4±31.6 Placebo:71.9±31.2
	SF36 domain: mental health	Bosentan: 73.6±20.1 Placebo: 71.3±21.0	Bosentan: 71.1±22.9 Placebo: 70.4±23.5
	EuroQol EQ-5D Health state score	Bosentan: 0.758±0.185 Placebo: 0.718±0.242	Bosentan: 0.660±0.386 Placebo: 0.656±0.366
	EuroQol EQ-5D Visual analogue score	Bosentan: 70.4±18.7 Placebo: 69.5±19.4	Bosentan: 65.9±24.0 Placebo: 66.4±23.2
Noth 2012[89]	Total SGRQ	Warfarin: 46.2±18.0 Placebo: 50.1±17.2	48 weeks follow up Reported: “no significant treatment effects observed”.
	SF36: aggregate physical score	Warfarin: 38.4±9.5 Placebo: 34.8±9.1	
	SF36: aggregate mental score	Warfarin: 48.2±8.6 Placebo: 48.4±9.6	
	EuroQol EQ-5D Health state score	Warfarin: 0.8±0.2 Placebo: 0.7±0.2	
	EuroQol EQ-5D Visual analogue score	Warfarin: 73.3±15.6 Placebo: 71.0±17.1	
Shulgina 2013[115]	Total SGRQ	Co-trimoxazole: 55.7±17.9 Placebo: 59.3±17.5	1 year follow up Co-trimoxazole: NR Placebo: NR
	Total SGRQ	1 year follow up Change from baseline Co-trimoxazole: 0.71±13.96 Placebo: 1.78±11.59	
	SGRQ: symptoms domain	Co-trimoxazole: -4.82±16.37 Placebo: 0.76±15.83	
	SGRQ: activity domain	Co-trimoxazole: 0.43±15.10	

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
		Placebo: 3.09±13.27	
	SGRQ: impact domain	Co-trimoxazole: 2.50±18.68 Placebo:0.99±13.88	
	EQ5D- based utility	Co-trimoxazole: -0.17±0.35 Placebo:-0.18±0.31	
Tomioka 2005[127]	SF36 domain: physical functioning	1 year follow up Change from baseline NAC: -18.2±6.6 Placebo:-17.5±6.0	
	SF36 domain: physical role functioning	NAC: -15.0±13.6 Placebo:-8.3±12.4	
	SF36 domain: vitality	NAC: -4.5±5.6 Placebo:-17.9±5.1	
	SF36 domain: bodily pain	NAC: -18.9±9.2 Placebo:-12.8±8.4	
	SF36 domain: general health perceptions	NAC: 1.6±4.8 Placebo:-4.8±4.4	
	SF36 domain: social role functioning	NAC: -3.8±7.5 Placebo:-12.5±6.9 P=0.07	
	SF36 domain: emotional role functioning	NAC: 20.0±16.5 Placebo:-22.2±15.1	
	SF36 domain: mental health	NAC: -2.0±5.1 Placebo:-14.7±4.6	
Zisman 2010[135]	SF36 domain: physical functioning	12 weeks follow up Change from baseline: mean change (95% CI) Sildenafil: -0.93(-2.24to0.38) Placebo: -1.46(-2.76to-0.17)	

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
		Absolute difference: 0.53 (-1.31 to 2.37) P value:0.57	
	SF36 domain: physical role functioning	Sildenafil: -0.87(-2.85 to 1.10) Placebo: -2.03(-3.98 to -0.08) Absolute difference: 1.16(-1.62 to 3.93) P value:0.41	
	SF36 domain: vitality	Sildenafil: 0.02(-1.70 to 1.75) Placebo:-2.01 (-3.70 to -0.31) Absolute difference: 2.03(-0.39-4.44) P value:0.10	
	SF36 domain: bodily pain	Sildenafil: -0.21(-2.13 to 1.71) Placebo: 1.97(0.08 to 3.85) Absolute difference: -2.17(-4.86 to 0.52) P value:0.11	
	SF36 domain: general health perceptions	Sildenafil: -1.04(-2.52 to 0.44) Placebo: -3.89(-5.37 to -2.42) Absolute difference:2.86 (0.76 to 4.95) P value:0.008	
	SF36 domain: social role functioning	Sildenafil: -0.72(-3.01 to 1.57) Placebo: -2.71(-4.97 to -0.46) Absolute difference: 1.99(-1.22 to 5.21) P value:0.22	
	SF36 domain: emotional role functioning	Sildenafil: -2.72(-5.56 to 0.12) Placebo: -4.82(-7.63 to -2.01) Absolute difference: 2.10(-1.90 to 6.10) P value:0.30	
	SF36 domain: mental health	Sildenafil: -0.16(-1.81 to 1.49) Placebo: -1.31(-2.93 to 0.30) Absolute difference: 1.15 (-1.15 to 3.46) P value:0.32	
	SF36: aggregate physical score	Sildenafil: -0.51(-1.86 to 0.83) Placebo: -0.35(-1.68 to 0.99) Absolute difference: -0.17(-2.06 to 1.73) P value:0.86	
	SF36: aggregate mental score	Sildenafil: 1.30(-0.59 to 3.18) Placebo: 3.02(1.15 to 4.89) Absolute difference:-1.72 (-4.38 to 0.93) P value:0.20	
	EuroQol EQ-5D	Sildenafil: -0.01(-0.06 to 0.03)	

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
	Health state score	Placebo: -0.03(-0.08 to 0.01) Absolute difference: 0.02(-0.04 to 0.08) P value:0.54	
	EuroQol EQ-5D Visual analogue score	Sildenafil: 0.48(-3.10 to 4.06) Placebo: -1.81(-5.34 to 1.73) Absolute difference: 2.28(-2.75 to 7.32) P value:0.37	
	Total SGRQ	Sildenafil: -1.64(-3.91 to 0.64) Placebo: 2.45(0.17 to 4.72) Absolute difference: -4.08(-7.30 to -0.86) P value:0.01	
	SGRQ: symptoms domain	Sildenafil: -3.58(-7.02 to -0.13) Placebo: 2.15(-1.30 to 5.61) Absolute difference:-5.73 (-10.61 to -0.85) P value:0.02	
	SGRQ: activity domain	Sildenafil: -1.15(-3.68 to 1.38) Placebo:2.49 (0.00 to 4.99) Absolute difference: -3.64(-7.20 to -0.09) P value:0.04	
	SGRQ: impact domain	Sildenafil: -0.88(-3.78 to 2.02) Placebo: 2.82(-0.03 to 5.67) Absolute difference: -3.70(-7.76 to 0.37) P value:0.07	
	Total SGRQ	Sildenafil: 54.55±16.46 Placebo: 51.72±15.86	Sildenafil: NR Placebo: NR
	SF36: aggregate physical score	Sildenafil:33.17±9.19 Placebo:34.84±8.69	Sildenafil: NR Placebo: NR
	SF36: aggregate mental score	Sildenafil:49.53±9.76 Placebo:50.58±9.52	Sildenafil: NR Placebo: NR
	EuroQol EQ-5D Health state score	Sildenafil:0.71±0.24 Placebo:0.74±0.19	Sildenafil: NR Placebo: NR
	EuroQol EQ-5D Visual analogue score	Sildenafil:66.49±17.45 Placebo:67.66±16.98	Sildenafil: NR Placebo: NR
Demedts2005[23]	Total SGRQ	NAC:50±18	6 & 12 month follow up

Reference	Outcome	Baseline (mean ± SD)	Post treatment follow-up (mean ± SD)
		Placebo: 52±16	NAC:NR Placebo: NR
Panther2012[50]	Total SGRQ	Azathioprine / prednisone /NAC: 38.7±17.4 Placebo: 39.4±17.4	60 week follow up Azathioprine / prednisone /NAC: 4.29 (-1.14, 9.73) Placebo: 7.50 (2.57, 12.4) Treatment difference: -3.20 (-10.5, 4.13) P value: 0.39
	SGRQ: symptoms domain	Azathioprine / prednisone /NAC: 49.4 ±21.1 Placebo: 45.6 ±21.8	Azathioprine / prednisone /NAC: -4.42 (-11.9, 3.1) Placebo: 8.31 (1.47, 15.2) Treatment difference: -12.7 (-22.9, -2.61) P value: 0.014
	SGRQ: activity domain	Azathioprine / prednisone /NAC: 51.1 ±19.0 Placebo: 52.7 ±21.0	Azathioprine / prednisone /NAC: 7.33 (1.05, 13.6) Placebo: 10.3 (4.66, 16.0) Treatment difference: -2.99 (-11.4, 5.46) P value: 0.49
	SGRQ: impact domain	Azathioprine / prednisone /NAC: 27.8 ±19.2 Placebo: 28.8 ±17.3	Azathioprine / prednisone /NAC: 5.23 (-0.80, 11.3) Placebo: 5.80 (0.34, 11.27) Treatment difference: -0.57 (-8.71, 7.57) P value: 0.89
	SF36: aggregate physical score	Azathioprine / prednisone /NAC: 40.3 ±9.8 Placebo: 40.6 ±9.3	Azathioprine / prednisone /NAC: -4.18 (-7.40, -0.97) Placebo: -2.96 (-5.90, -0.02) Treatment difference: -1.23 (-5.58, 3.13) P value: 0.58
	SF36: aggregate mental score	Azathioprine / prednisone /NAC: 53.9 ±9.6 Placebo: 55.7 ±7.4	Azathioprine / prednisone /NAC: 0.96 (-2.51, 4.44) Placebo: -4.35 (-7.50, -1.20) Treatment difference: 5.31 (0.62, 10.00) P value: 0.027
	EuroQol EQ-5D Health state score	Azathioprine / prednisone /NAC: 0.8±0.2 Placebo: 0.8±0.2	Azathioprine / prednisone /NAC: -0.07 (-0.14, -0.00) Placebo: -0.02 (-0.09, 0.04) Treatment difference: -0.05 (-0.14, 0.05) P value: 0.31
	EuroQol EQ-5D Visual analogue score	Azathioprine / prednisone /NAC: 76.8 ±15.5 Placebo: 78.1 ±15.4	Azathioprine / prednisone /NAC: -6.81 (-13.0, -0.67) Placebo: -6.66 (-12.4, -0.94) Treatment difference: -0.15 (-8.54, 8.24) P value: 0.93

11.3.3 Study quality and summary of findings

See Forest Plots in Appendix E, Clinical and Economic evidence tables in Appendix F and G respectively.

Where QoL data is reported

11.3.3.1 Warfarin

Two papers were identified [66, 89].

Table 57: Evidence profile for warfarin and prednisolone vs. prednisolone

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Warfarin + prednisolone	Prednisolone	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
All-cause Mortality*											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	No serious imprecision	None	5/23 (21.7%)	20/33 (60.6%)	RR 0.36 (0.16 to 0.82)	388 fewer per 1000 (from 109 fewer to 509 fewer)	Low
Number of hospitalisations due to IPF (acute) exacerbations											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	11/15 (73.3%)	21/29 (72.4%)	RR 1.01 (0.69 to 1.48)	7 more per 1000 (from 224 fewer to 348 more)	Very low
Survival at 1 year											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Serious[3]	None	29/33 (87.9%)	13/23 (56.5%)	RR 1.55 (1.06 to 2.27)	311 more per 1000 (from 34 more to 718 more)	Very low
Survival at 3 years											

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Warfarin + prednisolone	Prednisolone	Relative Risk	Absolute, Mean	
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Serious[3]	None	21/33 (63.6%)	8/23 (34.8%)	RR 1.83 (0.99 to 3.39)	289 more per 1000 (from 3 fewer to 831 more)	Very low

[1] *allocation concealment not reported; not double-blind; large number of dropouts; population of people hospitalised for IPF- possible bias*

[2] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*

[3] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID*

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Warfarin	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
All-cause Mortality*											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	No serious imprecision	None	14/72 (19.4%)	3/73 (4.1%)	4.73 (1.42 to 15.77)	153 more per 1000 (from 17 more to 607 more)	Moderate
Number of hospitalisations due to IPF (acute) exacerbations (follow-up mean 28 weeks)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[2]	None	6/72 (8.3%)	2/73 (2.7%)	3.04 (0.63 to 14.57)	56 more per 1000 (from 10 fewer to 372 more)	Very low
Adverse event: major bleed (follow-up mean 28 weeks)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[2]	None	2/72 (2.8%)	1/73 (1.4%)	2.03 (0.19 to 21.87)	14 more per 1000 (from 11 fewer to 286 more)	Very low
Adverse event: minor bleed (follow-up mean 28 weeks)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[2]	None	6/72 (8.3%)	2/73 (2.7%	3.04 (0.63 to 14.57)	56 more per 1000 (from 10 fewer to 372 more)	Very low
Total SGRQ											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	None	72	73	Baseline: Warfarin: 46.2±18.0		Low

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Warfarin	Placebo	Relative Risk	Absolute, Mean	
										Placebo: 50.1±17.2 At 48 weeks follow up reported: "no significant treatment effects observed".	
SF36: aggregate physical score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	None	72	73		Baseline: Warfarin: 38.4±9.5 Placebo: 34.8±9.1 At 48 weeks follow up reported: "no significant treatment effects observed".	Low
SF36: aggregate mental score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	None	72	73		Baseline: Warfarin: 48.2±8.6 Placebo: 48.4±9.6 At 48 weeks follow up reported: "no significant treatment effects observed".	Low
EuroQol EQ-5D, Health state score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	None	72	73		Baseline: Warfarin: 0.8±0.2 Placebo: 0.7±0.2 At 48 weeks follow up reported: "no significant treatment	Low

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Warfarin	Placebo	Relative Risk	Absolute, Mean	
									effects observed".		
EuroQol EQ-5D, Visual analogue score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	None	72	73	Baseline: Warfarin: 73.3±15.6 Placebo: 71.0±17.1 At 48 weeks follow up reported: "no significant treatment effects observed".		Low

[1] trial stopped prior to completion for safety thus all available results analysed together and high overall dropout rate
[2] Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID

11.3.3.2 Sildenafil

Two papers were identified [52, 135].

Table 58: Evidence profile for sildenafil vs. placebo

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Lung capacity (FVC)											
2	Randomised trials	Serious[1,3]	No serious inconsistency	No serious indirectness	Very serious imprecision[4]	None	103	106	N/A	MD 0.34 higher (1.06 lower to 1.75 higher)	Very low
Gas transfer (DLCO)											

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
2	Randomised trials	Serious[1,3]	No serious inconsistency	No serious indirectness	Serious imprecision[2]	None	103	106	N/A	MD 1.33 higher (0.09 lower to 2.75 higher)	Low
Dyspnoea (Borg)											
2	Randomised trials	Serious[1,3]	No serious inconsistency	No serious indirectness	Serious [2]	None	103	106	N/A	MD 0.17 lower (0.62 lower to 0.28 higher)	Low
Dyspnoea (Shortness of breath questionnaire)											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	No serious imprecision	None	89	91	N/A	MD 6.59 lower (11.45 to 1.73 lower)	Moderate
Performance on sub-maximal walk test: 6MWT (distance walked)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Serious [2]	None	14	15	N/A	MD 25 lower (70.59 lower to 20.59 higher)	Low
Mortality											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	No serious imprecision	None	2/81 (2.5%)	4/85 (4.7%)	RR 0.52 (0.1 to 2.79)	23 fewer per 1000 (from 42 fewer to 84 more)	Moderate
Adverse event: chest pain/ coronary artery disease											
2	Randomised trials	Serious[1,3]	No serious inconsistency	No serious indirectness	Very serious[4]	None	1/103 (0.97%)	1/106 (0.94%)	RR 1.04 (0.15 to 7.13)	0 more per 1000 (from 8 fewer to 58 more)	Very low
Adverse event: facial flushing											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[4]		1/14 (7.1%)	1/15 (6%)	RR 1.07 (0.07 to	5 more per 1000 (from 62	Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									15.54)	fewer to 969 more)	
Adverse event: visual disturbance											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[4]		1/14 (7.1%)	0/15 (0%)	RR 3.2 (0.14 to 72.62)	*	Very low
SF36 domain: physical functioning											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI) Sildenafil: -0.93(-2.24to0.38) Placebo: -1.46(-2.76to-0.17) Absolute difference: 0.53 (-1.31 to 2.37) P value:0.57		Low
SF36 domain: physical role functioning											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI) Sildenafil: -0.87(-2.85 to 1.10) Placebo: -2.03(-3.98 to -0.08) Absolute difference: 1.16(-1.62 to 3.93) P value:0.41		Low
SF36 domain: vitality											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI)		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									Sildenafil: 0.02(-1.70 to 1.75) Placebo:-2.01 (-3.70 to -0.31) Absolute difference: 2.03(-0.39-4.44) P value:0.10		
SF36 domain: bodily pain											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI) Sildenafil: -0.21(-2.13 to 1.71) Placebo: 1.97(0.08 to 3.85) Absolute difference: -2.17(-4.86 to 0.52) P value:0.11		Low
SF36 domain: general health perceptions											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -1.04(-2.52 to 0.44) Placebo: -3.89(-5.37 to -2.42) Absolute difference:2.86 (0.76 to 4.95) P value:0.008		Low
SF36 domain: social role functioning											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI):		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									Sildenafil: -0.72(-3.01 to 1.57) Placebo: -2.71(-4.97 to -0.46) Absolute difference: 1.99(-1.22 to 5.21) P value:0.22		
SF36 domain: emotional role functioning											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -2.72(-5.56 to 0.12) Placebo: -4.82(-7.63 to -2.01) Absolute difference: 2.10(-1.90 to 6.10) P value:0.30		Low
SF36 domain: mental health											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -0.16(-1.81 to 1.49) Placebo: -1.31(-2.93 to 0.30) Absolute difference: 1.15 (-1.15 to 3.46) P value:0.32		Low
SF36: aggregate physical score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									(95% CI): Sildenafil: -0.51(-1.86 to 0.83) Placebo: -0.35(-1.68 to 0.99) Absolute difference: -0.17(-2.06 to 1.73) P value:0.86		
SF36: aggregate mental score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: 1.30(-0.59 to 3.18) Placebo: 3.02(1.15 to 4.89) Absolute difference:-1.72 (-4.38 to 0.93) P value:0.20		Low
EuroQol EQ-5D, Health state score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -0.01(-0.06 to 0.03) Placebo: -0.03(-0.08 to 0.01) Absolute difference: 0.02(-0.04 to 0.08) P value:0.54		Low
EuroQol EQ-5D, Visual analogue score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									(95% CI): Sildenafil: 0.48(-3.10 to 4.06) Placebo: -1.81(-5.34 to 1.73) Absolute difference: 2.28(-2.75 to 7.32) P value:0.37		
Total SGRQ											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -1.64(-3.91 to 0.64) Placebo: 2.45(0.17 to 4.72) Absolute difference: -4.08(-7.30 to -0.86) P value:0.01		Low
SGRQ: symptoms domain											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -3.58(-7.02 to -0.13) Placebo: 2.15(-1.30 to 5.61) Absolute difference:-5.73 (-10.61 to -0.85) P value:0.02		Low
SGRQ: activity domain											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI):		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
									Sildenafil: -1.15(-3.68 to 1.38) Placebo:2.49 (0.00 to 4.99) Absolute difference: -3.64(-7.20 to -0.09) P value:0.04		
SGRQ: impact domain											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	At 12 weeks follow up, mean change from baseline (95% CI): Sildenafil: -0.88(-3.78 to 2.02) Placebo: 2.82(-0.03 to 5.67) Absolute difference: -3.70(-7.76 to 0.37) P value:0.07		Low
Total SGRQ											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	Baseline: mean (±SD): Sildenafil: 54.55±16.46 Placebo: 51.72±15.86		Low
SF36: aggregate physical score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	Baseline: mean (±SD): Sildenafil:33.17±9.19 Placebo:34.84±8.69		Low
SF36: aggregate mental score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	Baseline: mean (±SD): Sildenafil:49.53±9.76 Placebo:50.58±9.52		Low
EuroQoL EQ-5D, Health state score											
1	Randomised	Serious[3]	Not applicable	No serious	Could not be	None	89	91	Baseline: mean (±SD):		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Sildenafil	Placebo	Relative Risk	Absolute, Mean	
	trials			indirectness	calculated				Sildenafil:0.71±0.24 Placebo:0.74±0.19		
EuroQoL EQ-5D, Visual analogue score											
1	Randomised trials	Serious[3]	Not applicable	No serious indirectness	Could not be calculated	None	89	91	Baseline: mean (±SD): Sildenafil:66.49±17.45 Placebo:67.66±16.98		Low

[1] *unclear allocation concealment and investigator blinding*
[2] *The confidence interval crosses one minimally important difference making the effect size uncertain*
[3] *Blinding not reported*
[4] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID *no events in control group*

11.3.3.3 Bosentan

Two papers were identified [61] [63]

Table 59: Evidence profile for Bosentan vs. placebo

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Performance on sub-maximal walk test: 6MWT (distance)											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Serious imprecision[3]	None	71	83	N/A	MD 18 lower (57.23 lower to 21.23 higher)	Low
Mortality											

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk	Absolute, Mean	
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	No serious imprecision	None	11/407 (2.7%)	6/209 (2.9%)	RR 0.94 (0.35 to 2.51)	2 fewer per 1000 (from 19 fewer to 43 more)	Moderate
Adverse events: drug hypersensitivity											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Very serious[2]	None	1/406 (0.2%)	0/209 (0%)	RR 1.55 (0.06 to 37.83)	N/A	Very low
Adverse events: abnormal LFTs											
2	Randomised trials	Serious[1]	No serious inconsistency	No serious indirectness	No serious imprecision		39/480 (8.1%)	0/293 (0%)	RR 27.34 (3.57 to 209.53)	N/A	Moderate
Dyspnoea											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[2]		383	199	N/A	MD 0 higher (0.61 lower to 0.61 higher)	Low
SF36 domains											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	71	83	Reported; "general health compared with 1 year prior 42.4% (n=28) of bosentan treated patients had an improvement in SF36 health transition score compared with 28.4% (n=23) of placebo recipients – a relative risk of improvement in favour of bosentan of 1.49(95% CI, 0.96-2.33; p=0.084). Changes in seven of eight domains of the SF36 survey up to month 12 were		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk	Absolute, Mean	
									in favour of bosentan treatment, with a significant treatment effect observed in bosentan observed in the domain "role emotional"(p=0.032)".		
Total SGRQ											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	71	83	Absolute, Mean difference (95% CI) for baseline to 6 months: -2.80(-9.61 to 4.01) 12 month follow up: Reported: "mean treatment difference up to month 12 continued to favour bosentan but were smaller (data not shown)".		Low
SF36 domain: physical functioning											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 61.1±25.4 Placebo: 58.2±24.9 1 year follow up: Bosentan: 55.7±28.9 Placebo:52.8±27.6		Low
SF36 domain: physical role functioning											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 63.1±30.0 Placebo:59.2±29.0 1 year follow up: Bosentan: 58.5±32.4		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk	Absolute, Mean	
									Placebo:57.4±30.9		
SF36 domain: vitality											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 55.5±21.9 Placebo:52.3±22.4 1 year follow up: Bosentan: 51.6±24.4 Placebo:50.0±24.1		Low
SF36 domain: bodily pain											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 69.9±26.5 Placebo:68.4±27.8 1 year follow up: Bosentan: 64.3±31.1 Placebo:62.0±30.0		Low
SF36 domain: general health perceptions											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 52.1±21.5 Placebo:48.7±20.0 1 year follow up: Bosentan: 47.4±24.1 Placebo:46.9±22.9		Low
SF36 domain: social role functioning											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 77.6±24.3 Placebo:72.5±27.1 1 year follow up: Bosentan: 72.9±30.5		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk	Absolute, Mean	
									Placebo:69.3±29.7		
SF36 domain: emotional role functioning											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 79.3±26.2 Placebo: 74.7±29.0 1 year follow up: Bosentan: 73.4±31.6 Placebo:71.9±31.2		Low
SF36 domain: mental health											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 73.6±20.1 Placebo: 71.3±21.0 1 year follow up: Bosentan: 71.1±22.9 Placebo: 70.4±23.5		Low
EuroQoL EQ-5D, Health state score											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 0.758±0.185 Placebo: 0.718±0.242 1 year follow up: Bosentan: 0.660±0.386 Placebo: 0.656±0.366		Low
EuroQoL EQ-5D, Visual analogue score											
1	Randomised trials	Serious[1]	No serious inconsistency	Not applicable	Could not be calculated	None	407	209	Baseline: Bosentan: 70.4±18.7 Placebo: 69.5±19.4 1 year follow up: Bosentan: 65.9±24.0		Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Bosentan	Placebo	Relative Risk	Absolute, Mean	
									Placebo: 66.4±23.2		

[1] *Allocation concealment unclear*

[2] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*

[3] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID

11.3.3.4 Mycophenolate mofetil

No clinical evidence was found.

11.3.3.5 N-acetylcysteine

Two papers were found [45] [127]. These are reported and analysed separately due to the different dosages of NAC in each case. The route of administration of NAC was by inhalation in both cases.

Table 60: Evidence profile for N-acetylcysteine vs. placebo

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	*N*-acetylcysteine	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Lung capacity (FVC)											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	N/A	MD 2.4 higher (9.81 lower to 14.61 higher)	Very low
Gas transfer (DLCO)											
1	Randomised	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	N/A	MD 1.1 lower (18.99 lower to	Very

Quality assessment							Summary of findings				Quality
							No of patients		Effect		
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	*N*-acetylcysteine	Placebo	Relative Risk	Absolute, Mean difference	
	trials									16.79 higher)	low
Performance on sub-maximal exercise testing: 6MWT (distance walked)											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very Serious[3]	None	10	12	N/A	MD 66.4 higher (37.98 lower to 170.78 higher)	Very low
Lowest SaO_2 during 6MWT (change)											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	No serious imprecision	None	10	12	N/A	MD 5.5 higher (3.85 to 7.15 higher)	Very low
SF36 domain: physical functioning											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: -0.7(-6.02 to 4.62)		Very low
SF36 domain: physical role functioning											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: -6.70(-17.67 to 4.27)		Very low
SF36 domain: vitality											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: 13.4(8.89 to 17.91)		Very low
SF36 domain: bodily pain											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: -6.1(-13.52 to 1.32)		Very low
SF36 domain: general health perceptions											

Quality assessment							Summary of findings				Quality
							No of patients		Effect		
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	*N*-acetylcysteine	Placebo	Relative Risk	Absolute, Mean difference	
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: 6.4(2.52 to 10.28)		Very low
SF36 domain: social role functioning											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: 8.7(2.63 to 14.77)		Very low
SF36 domain: emotional role functioning											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: 42.2(28.87 to 55.53)		Very low
SF36 domain: mental health											
1	Randomised trials	Serious[1]	Not applicable	Serious[2]	Very serious[3]	None	10	12	Absolute, Mean difference (95% CI) baseline to 1 year follow up: 12.7(8.22 to 17.18)		Very low

[1] Randomisation method unclear; allocation concealment unclear; small sample size; open label study

[2] Japanese populations with a different course of disease.

[3] Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID

Table 61: Evidence profile for N-acetylcysteine vs. no treatment

Quality assessment							Summary of findings				Quality
							No of patients		Effect		
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Nebulised acetylcysteine 352.4mg bd	No treatment	Relative Risk (95% CI)	Absolute, Mean difference	

										(95% CI)	
Total number of patients with IPF exacerbation (follow-up mean 48 weeks)											
1	Randomised trials	very serious[1]	Not applicable	Serious[3]	Very serious[2]	none	1/44 (2.3%)	4/46 (8.7%)	RR 0.26 (0.03 to 2.25)	64 fewer per 1000 (from 84 fewer to 109 more)	Very low
Number of patients who subjectively felt their dyspnoea had improved compared to deteriorated at 48 weeks (follow-up mean 48 weeks)											
1	Randomised trials	Very serious[1]	Not applicable	Serious[3]	No serious	None	33/38 (86.8%)	32/38 (84.2%)	RR 1.03 (0.86 to 1.24)	25 more per 1000 (from 118 fewer to 202 more)	Low
Mean change in lung capacity (FVC) from baseline (%) at 48 weeks (follow-up mean 48 weeks)											
1	Randomised trials	Very serious[1]	Not applicable	Serious[3]	Serious[4]	None	38	38	N/A	MD 0.06 higher (0.05 lower to 0.17 higher)	Low

[1] Methodological limitations comprised of one or more of the following: unclear allocation concealment, the lack of blinding, inadequate allowance for drop-outs in the analysis, selective outcome reporting or unadjusted baseline inequality.
[2] Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID. Default MIDs were set at RRs of 0.75 and 1.25 for dichotomous, and at 2-6% change in FEV baseline
[3] Japanese population with a different course of disease
[4] The confidence interval crosses one minimal important difference making the effect size uncertain

11.3.3.6 Proton pump inhibitors

No RCTs were retrieved for proton pump inhibitors.

11.3.3.7 Co-trimoxazole

Table 62: Evidence profile for Co-trimoxazole vs. placebo

Quality assessment						No of patients		Effect		Quality

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Co-trimoxazole	Placebo	Relative risk (95% CI)	Absolute risk	
Mortality (ITT)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	No serious imprecision	None	18/95 (18.9%)	19/86 (22.1%)	RR 0.86 (0.48 to 1.52)	31 fewer per 1000 (from 115 fewer to 115 more)	Moderate
Mortality (per protocol)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	No serious imprecision	None	3/53 (5.7%)	14/65 (21.5%)	RR 0.26 (0.08 to 0.87)	159 fewer per 1000 (from 28 fewer to 198 fewer)	Moderate
Lung capacity: FVC (ml)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	Very serious[2]	None	63	60	-	MD 13.45 higher (96.04 lower to 122.94 higher)	Very low
Lung capacity: FVC % predicted											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	Very serious[2]	None	63	60	-	MD 0.14 higher (3.16 lower to 3.44 higher)	Very low
Gas transfer: DLCO (mmol/min/KPa)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	Very serious[2]	None	45	50	-	MD 0.08 lower (0.38 lower to 0.22 higher)	Very low
Gas transfer: DLCO % predicted											
1	Randomised trials	No serious	No serious inconsistency	Serious[1]	Very serious[2]	None	45	50	-	MD 0.21 higher (3.6	Very low

Quality assessment							No of patients		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Co-trimoxazole**	**Placebo**	**Relative risk (95% CI)**	**Absolute risk**	
		risk of bias								lower to 4.02 higher)	
SGRQ total (units)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	No serious imprecision	None	49	52	-	MD 1.07 lower (6.09 lower to 3.95 higher)	Moderate
6MWT (distance)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	No serious imprecision	None	20	31	-	MD 0.78 higher (44.15 lower to 45.71 higher)	Moderate
6MWT (desaturation of 4% or more)											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	Serious[3]	None	16/20 (80%)	31/35 (88.6%)	RR 0.9 (0.7 to 1.16)	89 fewer per 1000 (from 266 fewer to 142 more)	Low
MRC dyspnoea score											
1	Randomised trials	No serious risk of bias	No serious inconsistency	Serious[1]	Serious[3]	None	54	56	-	MD 0.14 lower (0.43 lower to 0.15 higher)	Low

11.3.3.8 Ambrisentan

One paper was identified [101]. This was available in abstract form only therefore limited data were able to be extracted and included.

All outcomes for this paper were outside of the protocol but were included as they were felt to be important for decision making by the GDG. Please see table 64 for extra outcomes not specified in the protocol but identified in studies for all treatments.

11.3.3.9 Combination

Three RCTS were retrieved [23] [102] [50].

Table 63: Evidence profile for azathioprine + prednisolone vs. prednisolone

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Azathioprine + Prednisolone	Prednisolone	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Lung capacity: FVC											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	14	13	N/A	MD 4.8 higher (16.53 lower to 26.13 higher)	Very low
Gas transfer: DLCO											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Serious[3]	None	14	13	N/A	MD 6.4 higher (11.8 lower to 24.6 higher)	Very low
Mortality											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	No serious imprecision	None	4/14 (28.6%)	4/13 (30.8%)	RR 0.93 (0.29 to 2.97)	22 fewer per 1000 (from 218 fewer to	Low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Azathioprine +	Prednisolone	Relative Risk	Absolute, Mean	
										606 more)	
Adverse events: elevated liver enzymes											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	1/14 (70.1%)	0/13 (0%)	RR 2.8 (0.12 to 63.2)	N/A	Very low
Adverse events- infections											
1	Randomised trials	Very serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	4/14 (28.6%)	1/13 (7.7%)	RR 3.71 (0.47 to 29.06)	208 more per 1000 (from 41 fewer to 1000 more)	Very low

[1] *Unclear allocation concealment. Patients allowed to cross-over between groups; ATS diagnostic criteria not used (CT not mandatory)*
[2] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*
[3] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID

Table 64: Evidence profile for prednisolone + azathioprine + N-acetylcysteine vs. azathioprine + prednisolone

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Prednisolone+ azathioprine+ N-acetylcysteine	Azathioprine + Prednisolone	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Lung capacity: FVC- Available case analysis (ACA)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[3]	None	55	51	N/A	MD 0.05 higher (0.24 lower to 0.34 higher)	Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Prednisolone+ azathioprine+	Azathioprine +	Relative Risk	Absolute, Mean	
Lung capacity: FVC- Intention to Treat analysis (ITT)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious imprecision[3]	None	71	68	N/A	MD 0.05 higher (0.2 lower to 0.3 higher)	Very low
Gas transfer: DLCO-ACA											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Serious[2]	None	48	47	N/A	MD 0.74 higher (0.06 to 1.42 higher)	Very low
Gas transfer: DLCO- ITT											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Serious[2]	None	48	47	N/A	MD 0.54 higher (0.03 lower to 1.11 higher)	Very low
Mortality											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	No serious imprecision	None	7/80 (8.8%)	8/75 (10.6%)	RR 0.82 (0.31 to 2.15)	19 fewer per 1000 (from 74 fewer to 123 more)	Moderate
Adverse event: abnormal LFTs											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	14/80 (17.5%)	11/75 (14.7%)	RR 1.19 (0.58 to 2.46)	28 more per 1000 (from 62 fewer to 214 more)	Very low

[1] *High drop-out rate; Patients excluded after randomisation*

[2] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID
[3] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*

Table 65: Evidence profile for Prednisolone + Azathioprine + N-acetylcysteine vs. Placebo

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
All-cause mortality											
1	Randomised trials	Serious[1]	Not applicable	No serious	No serious impression	None	8/77 (10.4%)	1/78 (1.3%)	8.1 (1.04 to 63.26)	91 more per 1000 (from 1 more to 798 more)	Low
IPF exacerbation											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[2]	None	5/77 (6.5%)	0/78 (0%)	11.14 (0.63 to 198.09)	*	Very low
Adverse events (infections)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[2]	none	5/77 (6.5%)	1/78 (1.3%)	RR 5.06 (0.61 to 42.36)	52 more per 1000 (from 5 fewer to 530 more)	Very low
Adverse events (GI)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[2]	none	1/77 (1.3%)	3/78 (3.8%)	RR 0.34 (0.04 to 3.18)	25 fewer per 1000 (from 37 fewer to 84 more)	Very low
Adverse events (metabolic)											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Very serious[2]	none	1/77 (1.3%)	0/78 (0%)	RR 3.04 (0.13 to 73.45)	*	Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
Total SGRQ											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 38.7±17.4 Placebo: 39.4±17.4 60 week follow up Azathioprine / prednisone /NAC: 4.29 (-1.14, 9.73) Placebo: 7.50 (2.57, 12.4) Treatment difference: -3.20 (-10.5, 4.13) P value: 0.39		Very low
SGRQ: symptoms domain											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 49.4 ±21.1 Placebo: 45.6 ±21.8 60 week follow-up Azathioprine / prednisone /NAC: -4.42 □-11.9, 3.1□ Placebo: 8.31 □1.47, 15.2□ Treatment difference: -12.7 □-22.9, -2.61□ P value: 0.014		Very low
SGRQ: activity domain											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 51.1		Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
									±19.0 Placebo: 52.7 ±21.0 At 60 weeks: Azathioprine / prednisone /NAC: 7.33 (1.05, 13.6) Placebo: 10.3 (4.66, 16.0) Treatment difference: -2.99 (-11.4, 5.46) P value: 0.49	
SGRQ: impact domain											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 27.8 ±19.2 Placebo: 28.8 ±17.3 At 60 weeks Azathioprine / prednisone /NAC: 5.23 (-0.80, 11.3) Placebo: 5.80 (0.34, 11.27) Treatment difference: -0.57 (-8.71, 7.57) P value: 0.89		Very low
SF36: aggregate physical score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 40.3 ±9.8 Placebo: 40.6 ±9.3 At 60 weeks: Azathioprine		Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
									/ prednisone /NAC: -4.18 (-7.40, -0.97) Placebo: -2.96 (-5.90, -0.02) Treatment difference: -1.23 (-5.58, 3.13) P value: 0.58		
SF36: aggregate mental score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 53.9 ±9.6 Placebo: 55.7 ±7.4 At 60 weeks: Azathioprine / prednisone /NAC: 0.96 (-2.51, 4.44) Placebo: -4.35 (-7.50, -1.20) Treatment difference: 5.31 (0.62, 10.00) P value: 0.027		Very low
EuroQol EQ-5D: Health state score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 0.8±0.2 Placebo: 0.8±0.2 At 60 weeks: Azathioprine / prednisone /NAC: -0.07 (-0.14, -0.00)		Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
									Placebo: -0.02 (-0.09, 0.04) Treatment difference: -0.05 (-0.14, 0.05) P value: 0.31		
EuroQol EQ-5D: Visual analogue score											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 76.8 ±15.5 Placebo: 78.1 ±15.4 At 60 weeks: Azathioprine / prednisone /NAC: -6.81 (-13.0, -0.67) Placebo: -6.66 (-12.4, -0.94) Treatment difference: -0.15 (-8.54, 8.24) P value: 0.93		Very low
SGRQ: symptoms domain											
1	Randomised trials	Serious[1]	Not applicable	No serious indirectness	Could not be calculated	none	77	78	Baseline: Azathioprine / prednisone /NAC: 49.4 ±21.1 Placebo: 45.6 ±21.8 60 week follow-up Azathioprine / prednisone /NAC: -4.42 □-11.9, 3.1□ Placebo: 8.31 □1.47, 15.2□ Treatment difference: -		Very low

Quality assessment							Summary of findings				
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Pred/AZA/NAC	Placebo	Relative Risk (95% CI)	Absolute, Mean difference (95% CI)	
									12.7 -22.9, -2.61 P value: 0.014		

[1] Risk of Bias: Serious: High risk attrition bias: No description of blinding methods or personnel given, unclear allocation concealment; Patients were allowed to cross over between groups ATS diagnostic criteria not used (CT not mandatory)
[2] Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID
*(*No events in control group therefore absolute difference cannot be calculated)*

Table 66: Evidence profile for extra outcomes not specified in the protocol but identified in studies for all treatments

Quality assessment						
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Effect size HR (95%CI)
Mortality HR (Noth 2012 warfarin vs. placebo)						
1	Randomised trials	Very serious[1]	Not applicable	Serious[4]	Could not be calculated*	1.58 (0.32, 2.83)
Time to IPF worsening/ death (King2011 bosentan vs. placebo)						
1	Randomised trials	Very serious[2]	Not applicable	No serious indirectness	Could not be calculated*	0.85 (0.653, 1.107)
Time to death up to study end (King 2011 bosentan vs. placebo)						
1	Randomised trials	Very serious[2]	Not applicable	No serious indirectness	Could not be calculated*	1.039 (0.6, 1.798)
Mortality HR (Panther 2012 triple therapy vs. placebo)						
1	Randomised	Serious[3]	Not applicable	No serious	Could not be calculated*	9.26 (NR)

Quality assessment						
No of studies	**Design**	**Limitations**	**Inconsistency**	**Indirectness**	**Imprecision**	**Effect size HR (95%CI)**
	trials			indirectness		
Mortality HR (Raghu 2012 ambrisentan vs. placebo)						
1	Randomised trials	Serious[5]	Not applicable	No serious indirectness	Could not be calculated*	2.05 (0.75, 5.76)
Categorical decrease in lung function (a 10% decrease in FVC with a 5% decrease in DLCO or a 15% decrease in DLCO with a 5% decrease in FVC) (Raghu 2012 ambrisentan vs. placebo)						
1	Randomised trails	Serious[5]	Not applicable	No serious indirectness	Could not be calculated*	1.53 (0.84, 2.78)
Time to IPF disease progression (Raghu 2012 ambrisentan vs. placebo)						
1	Randomised trials	Serious[5]	Not applicable	No serious indirectness	Could not be calculated*	1.74 (1.14, 2.66)

The GDG considered these outcomes to be important for decision making despite not being identified at protocol stage and have been included here for extra information.
**imprecision could not be calculated*
[1]*Trial stopped prior to completion for safety thus all available results analysed together and high overall dropout rate*
[2] *Unclear allocation concealment*
[3] *Japanese populations with a different course of disease*
[4] *Risk of Bias: Serious: high risk attrition bias, no description of blinding methods or personnel given, unclear allocation concealment, patients were allowed to cross over between groups and ATS diagnostic criteria not used (CT not mandatory)*
[5] *limited data available- abstract only*

11.4 Economic evidence summary

11.4.1 Literature review

One relevant economic evaluation was identified that compared a triple therapy of steroids, *N*-acetylcysteine and azathioprine to conservative treatment in the IPF population. The same study assessed whether thiopurine *S*-methyltransferase (TPMT) testing was cost effective prior to triple therapy.. See also the full study evidence tables in appendix G. No studies were selectively excluded.

Table 67: Economic evidence profile: Thiopurine *S*-methyltransferase testing versus no thiopurine *S*-methyltransferase testing compared to conservative treatment

Study	Applicability	Limitations	Other comments	Total cost per patient [d]	Total Effect (QALY per patient)	Cost effectiveness	Uncertainty
Hagaman [38] (USA)	Partially applicable [a]	Potentially serious limitations [b]	Decision analytic Markov model. Examines three strategies: Intvn 1: Conservative treatment [c] Intvn 2: Azathioprine, N-acetylcysteine and prednisone without testing Intvn 3: Azathioprine, N-acetylcysteine and prednisone with testing.	Intvn 1: £6,250 ($9691) Intvn 2: £10,191 ($15802) [e] Intvn 3: £10,201 ($15818) [f]	Intvn 1: 2.50 Intvn 2: 2.61 Intvn 3: 2.62	Intvn 1: reference Intvn 2: Extendedly Dominated Intvn 3 vs. Intvn 1: ICER = £31,701 ($49,156) Intvn 3 vs. Intvn 2: ICER = £19,130 ($26,663) (g).	Inspection from graph suggests that in order for TPMT testing to be cost effective compared to no testing, the prevalence of abnormal TPMT activity needs to be 2.5%. At prevalence above 13.5% TPMT testing dominates. If the probability of leukopenia on low dose of azathioprine increases above 12% over the base case value (21.4% with intermediate TPMT activity) then testing is no longer cost effective at $50,000 threshold [results not reported].

(a) Addresses appropriate population and intervention, with assessment of appropriate health effects, expressed in terms of Quality Adjusted Life Years. However, conducted from USA Medicare perspective, and some costs are reported as substantially higher than in the current UK context. Marginal costs between health states likely to be smaller in UK setting, in particular that between conservative and triple therapy. Discounting of costs and health outcomes not reported although a lifetime horizon was taken.

(b) Time horizon of 1 year, with extrapolation to lifetime horizon. Implicit assumption that if you have an adverse event due to inappropriate dosage it will occur in first year of treatment, and potentially some of the benefits of having appropriate dose beyond first year are not captured. Relevant health outcomes are included. Where possible RCT data is used, supplemented by observational data and expert opinion. Unclear if cost estimates come from the best source of data. Deterministic sensitivity performed and incremental analysis presented. No probabilistic sensitivity to explore uncertainty in results. No apparent conflict of interest.

(c) Costs converted from USA dollars to UK pounds using 2007 purchasing power parities.

(d) Reported as having a Diagnostic Resource Group resource code of 99243 (medical history and exam) 4 times annually. Authors note the efficacy for treatment effect was derived from the placebo arm of Ifigenia trial in which patients received azathioprine and prednisone with an N-acetylcysteine placebo. This is considered a reasonable approximation of effect for conservative treatment.

(e) Cost of azathioprine, N-acetylcysteine, and prednisone at standard dose, medical history and exam 3 times annually, monthly CBC for 1 year and bimonthly after, LFT and renal function biannually, PFT and CT scan annually, DEXA scanning, bisphosphate therapy, calcium, and vitamin D, co-trimoxazole 3 times weekly. Dose is not reported.

(f) Cost of azathioprine, N-acetylcysteine, and prednisone at reduced dose, medical history and exam 3 times annually, monthly CBC for 1 year and bimonthly after, LFT and renal function biannually, PFT and CT scan annually, DEXA scanning, bisphosphate therapy, calcium, and vitamin D, co-trimoxazole 3 times weekly. Dose is not reported. Assumption that reduced dose of therapy has the same efficacy as normal dose.

(g) Cost effectiveness of Thiopurine S-methyltransferase testing versus no thiopurine S-methyltransferase testing, without consideration of conservative treatment as a comparator

11.4.2 Unit costs

In the absence of recent UK cost-effectiveness analysis for many of the interventions identified as having potential to modify disease progression, relevant unit costs are provided in Appendix O to aid consideration of cost effectiveness. The below table summarises the total cost expected per patient per year's course of treatment and associated with interventions listed from least expensive to most expensive in terms of the total of the unit cost and additional costs associated with therapeutic drug monitoring.

Table 68: Unit cost of drug and associated monitoring cost

Item	Cost per year	Notes
Proton-pump inhibitors – Lansoprazole	• Cost of drug = £22 **Total = £22**	• Maintenance 30 mg once daily No monitoring required
N-acetylcysteine (oral)	• Cost of drug = £158 **Total = £158**	• 600mg 3 times daily No monitoring required As *N*-acetylcysteine is unlicensed in the UK costs are variable dependent upon brand of imported product. £158 should be considered at the lower range ofcost per year.
Warfarin	• Cost of drug = £10 • Additional costs = £202 **Total = £212**	• Dose according to INR • Assumed dose of 3mg daily INR be determined daily or on alternate days in early days of treatment, then at longer intervals, 4-6 weeks, then up to every 12 weeks. Assumed to equate to 19 visits to outpatient anticoagulation clinic.
Prednisolone	• Cost of drug = £24 • Additional costs = £220 **Total = £244**	• 15mg daily for first 6 weeks • 5 mg daily thereafter Corticosteroid monitoring in primary care and vitamin supplements given. Dexa scan included.
Co-trimoxazole	• Cost of drug = £171 • Additional costs = £138 **Total = £309**	• 960mg given twice daily 12 full blood counts taken in primary care
Azathioprine	• Cost of drug = £114 • Additional costs (inc. TPMT) = £280 **Total = £394**	• 2mg/kg – max 150mg per day • Assume 125 mg per day 13 Liver function tests, 7 full blood counts and TPMT TPMT activity measured
Mycophenolate mofetil	• Cost of drug = £207 • Additional costs = £218 **Total = £425**	• 1g twice daily Complete blood counts weekly during the first month, twice monthly for the second and third months of treatment, then monthly through the first year (19 nurse procedures in primary care)
Sildenafil - Revatio®	• Cost of drug = £4531 **Total= £4,532**	• By mouth, 20 mg 3 times daily;
Bosentan - Tracleer®	• Cost of drug = £19,633 • Additional costs = £171	• Initially 62.5 mg twice daily increased after 4 weeks to 125 mg twice daily; max. 250 mg twice daily 13 Liver Function Tests (nurse in primary care)

Item	Cost per year	Notes
	Total = £19,804	
Ambrisentan - Volibris®	• Cost of drug = £20,033 • Additional costs = £228 **Total = £20261**	• 5mg given daily 13 Liver Function Tests and 5 full blood counts (nurse in primary care)

Source: Please refer to Appendix J for details of cost breakdown and reference. Drug costs as per the March 2013 Drugs Tariff or MIMS March 2013 online database(for drugs excluded from the tariff)[3,87]. Cost for N-acetylcysteine costed as per quote obtained directly from pharmaceutical supplier.

11.5 Evidence statements

11.5.1 Clinical review evidence statements

11.5.1.1 Warfarin and Prednisolone vs. Prednisolone

All-cause mortality

Low quality evidence showed that warfarin and prednisolone is clinically more effective at reducing deaths compared to prednisolone alone (one study, N=56)

Hospitalisations due to IPF complications (including exacerbations)

Very low quality evidence showed that there may be no clinical difference between warfarin and prednisolone and prednisolone alone at reducing the number of hospitalisations due to IPF exacerbations (one study, N=56)

Survival

Very low quality evidence showed that warfarin and prednisolone is potentially more clinically effective than prednisolone alone at improving 1 year survival (one study, N=56).

Very low quality evidence shows that warfarin and prednisolone is potentially more clinically effective than prednisolone alone at improving 3 year survival (one study, N=56)

11.5.1.2 Warfarin vs. placebo

Mortality

Low quality evidence showed that warfarin is less clinically effective at reducing deaths when compared to placebo (one study, N=145)

Hospitalisations due to IPF complications (including exacerbations)

Very low quality evidence showed that warfarin is less clinically effective in reducing IPF exacerbations than placebo (one study, N=145)

Adverse event (major bleed)

Very low quality evidence showed that warfarin is less clinically effective at reducing adverse events (major bleeds) than placebo (one study, N=145)

Adverse event (minor bleed)

Very low quality evidence showed that warfarin is less clinically effective at reducing adverse events (minor bleeds) than placebo (one study, N=145)

11.5.1.3 Sildenafil vs. placebo

Lung capacity (FVC)

Very low quality evidence showed that sildenafil may be clinically effective compared with placebo at improving FVC but the direction of the estimate could favour either intervention (two studies, N=209)

Gas transfer (DLCO)

Low quality evidence showed that sildenafil may be clinically effective compared with placebo at improving DLCO but the direction of the estimate could favour either intervention (two studies, N=209)

Dyspnoea (Borg scale)

Low quality evidence showed there may be no difference between sildenafil and placebo at improving dyspnoea; the direction of the estimate of effect favoured sildenafil (two studies, N=209)

Dyspnoea (Shortness of breath questionnaire)

Low quality evidence showed that there may be no clinical difference between sildenafil and placebo at reducing dyspnoea (one study, N= 180)

Performance on 6MWT

Moderate quality evidence showed that sildenafil is less clinically effective than placebo in improving distance walked in the 6MWT (one study, N=29).

Mortality

Moderate quality evidence showed that sildenafil may be clinically more effective than placebo at reducing mortality but the direction of the estimate of effect could favour either intervention (one study, N=29).

Adverse events (chest pain/coronary artery disease)

Very low quality evidence showed that there may be no clinical difference between sildenafil and placebo in causing adverse events due to coronary artery disease but the direction of the estimate of effect could favour either intervention (two studies, N=209)

Adverse events (facial flushing)

Very low quality evidence showed that there may be no clinical difference between sildenafil and placebo in causing adverse events (facial flushing) but the direction of the estimate of effect could favour either intervention (one study, N=29).

Adverse events (visual disturbance)

Very low quality evidence showed that sildenafil may be more likely to cause visual disturbance compared with placebo but the direction of the estimate of effect could favour either intervention (one study, N=29).

11.5.1.4 Bosentan vs. placebo

Performance on 6MWT

Low quality evidence showed that there may be no clinical difference between bosentan and placebo in improving distance walked in the 6MWT (one study, N=154).

Mortality

Moderate quality evidence showed that bosentan is more effective than placebo in reducing mortality (one study, N=154).

Dyspnoea

High quality evidence showed that there may be no difference between bosentan and placebo in reducing dyspnoea but the direction of the estimate of effect would favour either intervention (one study, N=154).

Adverse events (drug hypersensitivity)

Very low quality evidence showed that placebo may be more clinically effective than bosentan at minimising adverse events (drug hypersensitivity) but the direction of the estimate of effect could favour either intervention (one study, N=154).

Adverse events (abnormal liver function tests)

Moderate quality evidence showed that placebo is more effective than bosentan at preventing abnormal liver function tests (one study, N=154).

11.5.1.5 *N*-acetylcysteine vs. placebo

Lung capacity (FVC)

Very low quality evidence showed that N-acetylcysteine may be clinically effective compared with placebo in improving FVC but the direction of the effect could favour either intervention (one study, N=22).

Gas transfer (DLCO)

Very low quality evidence showed that DLCO is reduced when using N-acetylcysteine compared with placebo but the direction of the estimate of effect could favour either intervention (one study, N=22).

Performance on 6MWT

Very low quality evidence showed that N-acetylcysteine may be clinically effective compared with placebo at improving distance walked in the 6MWT but the direction of the estimate of effect could favour either intervention (one study, N=22).

Low quality evidence showed that N-acetylcysteine may be clinically effective compared with placebo at improving the lowest SaO_2 in the 6MWT but the direction of the estimate of effect could favour either intervention (one study, N=22)

11.5.1.6 N-acetylcysteine vs. no treatment

Lung capacity (FVC)

Low quality evidence showed that N-acetylcysteine may be clinically effective compared with no treatment at improving FVC (one study, N=76).

Hospitalisations due to IPF complications (including IPF exacerbations)

Very low quality evidence showed that there may be no clinical difference between N-acetylcysteine and placebo in hospitalisations due to IPF complications, including IPF exacerbations (one study, N=76).

Dyspnoea

Very low quality evidence showed that there is no clinical difference between N-acetylcysteine and no treatment in improving dyspnoea (one study, N=76).

11.5.1.7 Co-trimoxazole vs. placebo

Lung capacity (FVC (ml))

Very low quality evidence showed that there was no difference between co-trimoxazole and placebo at improving FVC (one study, N=181).

Lung capacity (FVC (% predicted))

Very low quality evidence showed that there was no difference between co-trimoxazole and placebo at improving FVC (one study, N=181).

Gas transfer (DLCO (mmol/min/kPa))

Very low quality evidence showed that there was no difference between co-trimoxazole and placebo at improving DLCO (one study, N=181).

Gas transfer (DLCO (% predicted))

Very low quality evidence showed that there was no difference between co-trimoxazole and placebo at improving DLCO (one study, N=181)

Mortality (ITT analysis)

Moderate quality evidence showed that cotrimoxazole is clinically effective compared with placebo at reducing mortality (one study, N=181)

Mortality (per protocol analysis)

Moderate quality evidence showed that cotrimoxazole is clinically effective compared with placebo at reducing mortality (one study, sample size not clearly reported)

Health related quality of life-SGRQ

Moderate quality evidence showed that there was no difference between co-trimoxazole and placebo at improving health-related quality of life (one study, N=181)

Performance on sub-maximal exercise testing, 6MWT (distance walked)

Moderate quality evidence showed that there was no difference between co-trimoxazole and placebo at improving performance on sub-maximal exercise testing (one study, N=181).

Dyspnoea (MRC score)

Low quality evidence showed that there was no difference between co-trimoxazole and placebo at improving dyspnoea (one study, N=181).

11.5.1.8 Ambrisentan vs. placebo

No outcomes listed in the protocol were found for this study.

11.5.1.9 Azathioprine + Prednisolone vs. Prednisolone

Lung capacity (FVC)

Very low quality evidence showed that azathioprine + prednisolone may be more clinically effective than prednisolone in improving FVC but the direction of the estimate of effect could favour either intervention (one study, N=27)

Gas transfer (DLCO)

Low quality evidence showed that there may be no clinical difference between a combination of prednisolone + azathioprine and prednisolone alone in improving DLCO (one study, N=27).

Mortality

Low quality evidence showed that a combination of prednisolone + azathioprine is less effective than prednisolone alone at reducing mortality (one study, N=27).

Adverse events: elevated liver enzymes

Very low quality evidence showed that prednisolone may be more clinically effective than a combination of prednisolone + azathioprine in reducing adverse events (elevated liver enzymes) but the direction of the estimate of effect could favour either intervention (one study, N=27).

Adverse events: infections

Very low quality evidence showed that prednisolone may be more clinically effective than a combination of prednisolone + azathioprine in reducing adverse events (infections) but the direction of the estimate of effect could favour either intervention (one study, N=27).

11.5.1.10 Prednisolone + Azathioprine + *N*-acetylcysteine vs. Azathioprine + Prednisolone

FVC: available case analysis

Very low quality evidence showed that prednisolone + Azathioprine + N-acetylcysteine may be more clinically effective than azathioprine + prednisolone but the direction of the estimate of effect could favour either intervention (one study, N=106).

FVC: intention to treat analysis

Very low quality evidence showed that prednisolone + Azathioprine + N-acetylcysteine may be more clinically effective than azathioprine + prednisolone but the direction of the estimate of effect could favour either intervention (one study, N=106).

DLCO: available case analysis

Low quality evidence showed that prednisolone + azathioprine + N-acetylcysteine is potentially more clinically effective than azathioprine + prednisolone in improving DLCO (one study, N=106).

DLCO: intention to treat analysis

Low quality evidence showed that prednisolone + azathioprine + N-acetylcysteine is potentially more clinically effective than azathioprine + prednisolone in improving DLCO (one study, N=106).

Mortality

Moderate quality evidence showed that prednisolone + azathioprine + N-acetylcysteine is more clinically effective than azathioprine + prednisolone at reducing mortality (one study, N=155).

Adverse events: abnormal liver function tests

Very low quality evidence showed that there was too much uncertainty to determine whether there is a difference between prednisolone + azathioprine + *N*-acetylcysteine and azathioprine + prednisolone in the incidence of abnormal liver function tests from baseline when assessed at 12 months follow-up (one study, N=155).

11.5.1.11 Prednisolone and AZA + NAC vs. placebo

Mortality

Low quality evidence showed that placebo is more effective than a combination of prednisolone, Azathioprine +N-Acetylcysteine in reducing all-cause mortality (one study, N=155).

Hospitalisations due to IPF complications (including IPF exacerbations)

Very low quality evidence showed that a combination of prednisolone, Azathioprine +N-Acetylcysteine may be less clinically effective than placebo at reducing IPF exacerbations but the direction of the estimate of effect could favour either intervention (one study, N=155).

Side effects (infectious)

Very low quality evidence showed that a combination of prednisolone, Azathioprine +N-Acetylcysteine may be less clinically effective than placebo at reducing side effects (infections) but the direction of the estimate of effect could favour either intervention (one study, N=155).

Side effects (gastrointestinal)

Very low quality evidence showed that placebo may be less clinically effective than a combination of prednisolone, Azathioprine +N-Acetylcysteine at reducing side effects (GI) but the direction of the estimate of effect could favour either intervention (one study, N=155).

Side effects (metabolic)

Very low quality evidence showed that a combination of prednisolone, Azathioprine +N-Acetylcysteine may be less clinically effective than placebo at reducing side effects (metabolic) but the direction of the estimate of effect could favour either intervention (one study, N=155).

11.5.1.12 Quality of life:

Low to very low quality evidence showed that there was no clinically effective difference between the drug investigated and placebo/ no treatment in any QOL measures**.**

11.5.2 Health economic evidence statements

- TPMT testing before prescription of azathioprine, steroids and N acetylcysteine, in IPF patients is a cost effective strategy compared to no testing and likely to be cost saving in the UK setting.
- Azathioprine, steroids and *N*-acetylcysteine (with TPMT testing prior to initiation) is unlikely to be a cost effective strategy in the treatment of IPF when compared to conservative treatment.
- These statements are based on evidence that is partially applicable and with potentially serious limitations.
- It is unclear whether co trimoxazole is cost effective in modifying IPF disease progression. This is based on evidence of direct applicability and with potentially serious limitations.

11.6 Recommendations and link to evidence

	There is no conclusive evidence to support the use of any drugs to increase the survival of people with idiopathic pulmonary fibrosis.
Recommendations	**27.For guidance on pirfenidone, see the NICE technology appraisal on**

	pirfenidone for the treatment of idiopathic pulmonary fibrosis. For guidance on nintedanib, see the NICE technology appraisal on nintedanib for the treatment of idiopathic pulmonary fibrosis. **28. Do not use any of the drugs below, either alone or in combination, to modify disease progression in idiopathic pulmonary fibrosis:** • **ambrisentan** • **azathioprine** • **bosentan** • **co-trimoxazole** • **mycophenolate mofetil** • **prednisolone** • **sildenafil** • **warfarin.** **29. Advise the person that oral *N*-acetylcysteine[g] is used for managing idiopathic pulmonary fibrosis, but its benefits are uncertain.** **30. If people with idiopathic pulmonary fibrosis are already using prednisolone or azathioprine, discuss the potential risks and benefits of discontinuing, continuing or altering therapy.** **31. Manage any comorbidities according to best practice. For gastro-oesophageal reflux disease, see Managing dyspepsia in adults in primary care (NICE clinical guideline 17).**
Relative values of different outcomes	All-cause mortality and changes in lung function measures (FVC and DLCO) were considered the critical outcome measures in assessing the efficacy of pharmacological treatments. 6MWD and adverse events were also considered by the GDG to be important outcomes to inform decision making for these recommendations.
Trade-off between clinical benefits and harms	The GDG considered the evidence for the clinical benefits against the adverse events of the individual pharmacological treatments. No evidence was retrieved for mycophenolate mofetil, which can cause bone marrow suppression and hepatotoxic reactions. No evidence was also found for proton pump inhibitors in treating IPF, which can cause gastrointestinal disturbance and rarely hepatotoxicity. One abstract comparing ambrisentan to placebo showed differences between number of deaths and trial participants with a categorical decrease in lung function between groups. There were more respiratory hospitalisations in the ambrisentan group compared to the placebo.

[g] At the time of publication (June 2013), *N*-acetylcysteine did not have a UK marketing authorisation. The prescriber should follow relevant professional guidance, taking full responsibility for the decision. Informed consent should be obtained and documented. See the General Medical Council's Good practice in prescribing medicines – guidance for doctors for further information.

Bosentan is associated with hypotension and oedema. The GDG considered the evidence for bosentan, which showed no difference in the 6MWD and dyspnoea compared with placebo, but resulted in more abnormal LFTs and drug hypersensitivity with bosentan. There was no appreciable benefit in quality of life for bosentan.

The GDG considered the cost effectiveness evidence for ambrisentan and bosentan, and concluded not to recommend these drugs as they are unlikely to be cost effective and because there remained uncertainty around their potential to result in negative outcomes.

Co-trimoxazole can cause a variety of hypersensitivity reactions including Stevens-Johnson syndrome. The evidence for co-trimoxazole showed no clinically important difference between the intervention and control groups for change in FVC, TLCO, DLCO or 6MWD. Clinically important differences in symptom domain or SGRQ and percentage of patients requiring increase in oxygen therapy in favour of co-trimoxazole treatment.

Sildenafil is also associated with hypotension and oedema. The evidence for sildenafil showed an improvement in mortality and FVC compared to placebo, but there were more visual disturbances. DLCO and 6MWT were also worsened, however there was much uncertainty. There was no appreciable benefit in quality of life. The Borg score showed no difference and the shortness of breath questionnaire was worse. The GDG did not recommend sildenafil based on the uncertainty in these effects.

Warfarin treatment is associated with a significant risk of serious bleeding and in rare cases with hepatotoxicity and skin necrosis. The evidence for warfarin versus placebo showed higher mortality, hospitalisations and bleeding in the group treated with warfarin. The evidence for warfarin versus prednisolone showed warfarin had greater improvements on mortality and survival compared to prednisolone, but the study was of very low quality and had an indirect population, as patients were hospitalised. On balance, the GDG did not recommend warfarin due to harms associated with its use.

The effects of long-term steroid therapy include immunosuppression, increased risk of osteoporosis, weight gain, diabetes, peptic ulceration and Cushing's syndrome and those of azathioprine include hepatotoxicity, bone marrow suppression and immunosuppression. Most clinical experience lies with a regimen comprising triple therapy of prednisolone, azathioprine and N-acetylcysteine. This combination was shown to reduce the rate of decline in FVC and TLCO compared to a combination of prednisolone and azathioprine in one study. However, evidence retrieved from the PANTHER trial showed that the combination of prednisolone with azathioprine and N-acetylcysteine was associated with increased risk of death and significant adverse effects compared to placebo. Because data from the same study suggested that N-acetylcysteine alone was not likely to be harmful, the GDG concluded that it was prednisolone or azathioprine which was the toxic component of triple therapy , but acknowledged that it is currently not known whether a single treatment or the combination of these treatments were responsible for the adverse effects seen. The N-acetylcysteine arm of the Panther trial is on-going. Oral N-acetylcysteine appears relatively safe in therapeutic doses although it may alter the viscosity of gastrointestinal mucous and cause upset. The GDG acknowledged that N-acetylcysteine is not licensed for IPF disease modifying purposes.

No evidence was retrieved for either azathioprine or corticosteroids used as monotherapy in IPF. Both drugs have known adverse effects and after considerable deliberation the GDG considered that both azathioprine and prednisolone were not

	to be recommended in combination or alone on the basis of their adverse events profile and current concern with their safety when used in triple therapy form. The GDG acknowledged that corticosteroids may have unproven beneficial effects on patient symptoms e.g. cough and high doses may have unproven benefits in people experiencing acute exacerbations of IPF. However the GDG decided that corticosteroids should not be used to modify disease progression in IPF. No evidence was retrieved which informed thiopurine S-methyltransferase testing (TPMT). TPMT measurements prior to commencing treatment with thiopurine drugs (in order to anticipate the possible accumulation and toxicity of un-metabolised drug), such as azathioprine is encouraged. However, due to the harms associated with azathioprine the GDG agreed that making a recommendation or research recommendation for TPMT testing was inappropriate as the use of azathioprine is not recommended.
Economic considerations	There was one published health economic study that was identified to inform this question and a further draft in confidence was given after identification of an abstract. The published study explored the cost effectiveness of a regimen azathioprine, N-acetylcysteine and steroids with and without testing of thiopurine S-methyltransferase (TPMT) testing to conservative treatment. The efficacy of triple therapy was taken from the intention to treat dichotomised data of the Demedts et al (2005) trial and assumed that the efficacy of conservative treatment was similar to that found for a regimen of azathioprine, steroids and an N-acetylcysteine placebo. This was considered an appropriate approximation to estimate disease progression for patients having conservative treatment. The GDG considered the costs from the USA Medicare setting were thought to be higher than costs in the UK current practice. To note in particular, the TPMT test itself cost $300 (£197), whereas in the UK the assay can be provided for approximately £29. If the cost of £197 was substituted with £29, TPMT testing would dominate the no TPMT testing strategy (being more effective and less costly). The cost difference between conservative treatment and the triple drug regimen in the UK setting is likely to be smaller than that found in the study. In the incremental analysis, the triple therapy regimen without TPMT testing is extendedly dominated and therefore excluded from further consideration. This is because a combination of conservative treatment and triple therapy with testing would be more cost effective than triple therapy without testing. When we consider the cost effectiveness of triple therapy plus testing compared to conservative therapy, the incremental cost effectiveness ratio (ICER) is £31,701. This suggests triple therapy, even with testing, is not a cost effective strategy. There could be great uncertainty in the results from this analysis, however this was not assessed formally or quantified by probabilistic sensitivity analysis. The GDG assessed the study in light of the findings from the clinical review and raised concerns over the adverse effect profile of the drugs alone or in combination. For patients in whom azathioprine is not contraindicated by TPMT testing, it is still uncertain whether there would be a significant adverse effect profile of the drugs combined in triple therapy. Overall the GDG concluded that on account of concerns regarding the quality and applicability of the included economic study, and potential safety concerns, triple therapy should not be recommended. The GDG also considered the cost of N-acetylcysteine as a single intervention. Although one supplier quoted an acquisition cost which would amount to £158 per

annum, the GDG noted that in practice the cost of this intervention is variable because as an unlicensed "special" drug it is not included in the drug tariff or regulated. The GDG also commented that the drug was available from health food shops at a much lower cost at approximately £28 per month. However, this was likely to be prohibitively expensive to the patient (also noting that the NICE reference case only considers costs from a NHS perspective and excludes out of pocket expenditure by the patient). One clinical member recalled that a recent community pharmacy prescription was £298 for a month's supply (£3600 per annum). The GDG agreed that it was likely the cost of supplying this drug varied greatly and when also taking into account hospital prescription and dispensing costs, it was likely £158 per annum was at the lower end of the potential range of where the average cost of this drug lies at the time of development. Nevertheless, given the potential health benefit and reduction in healthcare contacts evidenced by the clinical review, the GDG deliberated that clinicians should at least advise patients of the uncertain but potential benefit of N-acetylcysteine. The cost effectiveness of this intervention remains unclear.

The GDG decided they could not recommend co-trimoxazole for the purpose of modifying disease progression. This was because the TIPAC study showed no effect on decline in FVC, the primary outcome. The GDG acknowledged that co-trimoxazole might be a cost effective therapy for improving mortality by reducing respiratory infections in IPF and might be cost effective for improving quality of life in IPF, but the GDG considered that the evidence for efficacy in both regards was not conclusive

In regards to the other pharmacological agents and in the absence of health economic evidence, the acquisition cost of ambrisentan, bosentan, mycophenolate mofetil, PPIs and co-trimoxazole were presented to the GDG alongside relevant therapeutic monitoring costs. The GDG also considered the probability and cost of adverse events in their deliberations. The unit costs presented were from publically available list prices and the dosages were considered appropriate to establish an estimate of cost for each drug.

The review suggested bosentan did not have any appreciable benefit in terms of survival or quality of life, and no economic evidence was available for ambrisentan. The yearly acquisition cost of ambrisentan and bosentan was sufficiently high that both these drugs were considered extremely unlikely to be cost effective.

Without any formal health economic evidence, the cost effectiveness of sildenafil was unclear given that the review showed a point estimate of improved mortality in favour of sildenafil (albeit statistically non-significant) and that the annual cost was relatively high. Modelling in this clinical topic area was not prioritised. Therefore, to aid the informal assessment of the cost effectiveness of sildenafil, the GDG considered the absolute difference in mortality found at 6 months compared to the cost of a 6 month course for sildenafil. Treating 1000 patients with sildenafil would save 22 lives at 6 months at an additional cost of £2,265,500 compared to a 'do nothing approach'. In order to offset this initial 6 month treatment cost and make the incremental cost effectiveness ratio fall at or below the £20,000 threshold, the surviving 22 patients would need to live an additional 5 to 7 years to those in the non-treatment group (assuming survival was lived with a utility of 1 (full health) to 0.7 respectively). Although this is a simplistic calculation in that only one trade off was taken into account (6 month treatment cost versus potential QALY gain), it indicates the impact on survival (via disease modification) that is required to make the intervention cost effective at a £20,000 threshold. Taking this into account (noting its simplifications), and the finding from the clinical review that it is

	uncertain that sildenafil reduces mortality, the GDG came to a consensus that sildenafil was unlikely to be cost effective at the current time. Mycophenolate mofetil also had a relatively high acquisition cost compared to alternative interventions considered in this review and no clinical evidence was identified. As there was no evidence to suggest appreciable benefit, along with an appreciable cost and side effect profile, the GDG decided to recommend against its use. However, the GDG acknowledged that the cost effectiveness of this intervention has not been formally assessed. In the absence of health economic evidence, the acquisition cost of PPIs was presented to the GDG. In comparison to the other pharmacological interventions the acquisition cost of PPIs is relatively low. However, the relative cost effectiveness of PPIs when administered as a single therapy remains unclear. There was no available evidence to inform on the cost effectiveness of different sequencing of pharmacological interventions.
Quality of evidence	The GDG considered the clinical evidence and the health economic findings for all the drugs listed in the scope. Evidence quality was downgraded across some of the studies for indirect population as in some instances the populations were exclusively Japanese, high drop- out rate, unclear allocation concealment and unclear blinding. Disease progression in IPF was considered by the GDG to imply decline in lung function, most appropriately signified by decline in FVC. The GDG considered that a drug should be given a 'do not use to modify disease progression in IPF' recommendation if one or more of the following criteria was met and the decision was agreed by consensus within the GDG: There was evidence of no effect or evidence of harm There was evidence that the drug was not cost-effective There was insufficient evidence to demonstrate efficacy. The evidence for warfarin was of very low quality, had an indirect population, as all patients were hospitalised and the GDG agreed that due to potential adverse events this drug should not be recommended for disease modification in IPF. The GDG considered the cost effectiveness evidence for ambrisentan and bosentan, and concluded not to recommend these drugs as they are unlikely to be cost effective and because uncertainty remained around their potential to result in negative outcomes. The GDG discussed the evidence retrieved for sildenfil alone when compared to placebo, but agreed that this evidence was not sufficient to support their use in modifying IPF disease. No evidence was retrieved for mycophenolate mofetil and proton pump inhibitors in treating IPF. Due to the lack of evidence the GDG questioned the safety of these drugs and agreed that further research was needed to test the efficacy of these drugs. The Panther trial showed higher mortality, exacerbations and adverse events in the triple therapy group. Most clinical experience lies with a regimen comprising triple therapy of prednisolone, azathioprine and N-acetylcysteine. The quality of evidence was very low due to attrition bias and unclear blinding method, as well as very imprecise outcomes. The GDG considered that both azathioprine and prednisolone were not to be recommended on the basis of their adverse events profile and current concern with their safety when used as in triple therapy form. The GDG noted that N-acetylcysteine was included in the triple therapy trial, but considered that the evidence for its single use showed some improvement in outcomes and was

	overall relatively safe when compared to other pharmacological options. The studies measured the effects of inhaled and oral N-acetylcysteine on indirect populations (Japanese populations) and most of the outcomes were of very low quality. The GDG noted that even though this drug is not licensed The GDG noted that even though this drug is not licensed, the fact that it is relatively safe, and that given that it is frequently prescribed for people with IPF and bought over the counter at low doses, built the case for the GDG to advise patients that it is used in managing IPF, but when non-other exists, but that its benefits remain uncertain.
Other considerations	Overall there was not sufficient information reported in the studies to make any conclusion regarding timing of initiation of any drug treatment. The GDG acknowledged that whilst there was a lack of evidence found for the effectiveness of drugs for disease modifying purposes in IPF, people may be prescribed drugs for purposes other than for treatment of IPF. For example, the GDG considered the use of PPIs for gastro-oesophageal reflux disease in people with IPF, but agreed that there was insufficient evidence to suggest a therapeutic effect of PPIs on disease progression in people with IPF. Similarly, warfarin should be used in people with IPF who develop venous-thromboembolism and sildenafil and endothelin antagonists can be used to treat pulmonary hypertension in people with IPF. Research recommendations The GDG agreed that the lack of evidence for the use of anti-reflux therapy, corticosteroids and co-trimoxazole for people with IPF justified developing a research recommendation to address whether these treatments are effective in people with IPF. For further information on the research recommendations see Appendix P.

12 Lung transplantation

12.1 Review introduction

In cases of advanced fibrotic lung disease associated with a poor prognosis or refractory limiting symptoms, selected patients may be suitable for lung transplantation. It is important that patients are referred for transplant assessment in a timely fashion. Either single or bilateral pulmonary transplantation may be considered although the latter is associated with superior short and long term survival. Both provide a gain in life expectancy and relief of symptoms of breathlessness with improved quality of life. Patients must have no contraindications to transplantation and be in the 'transplant window' where the risks of surgery are acceptable given the patient's status but the patient is robust enough to make transplantation feasible. In the longer term the risks of immunosuppression are important. Donor organ shortages mean that even when listed actively for transplantation, some patients may not benefit.

12.2 Clinical question and review methodology

The following clinical question was included in this chapter:

12.2.1 What is the optimal timing to consider a patient with IPF for lung transplantation referral?

For full details see review protocol in Appendix C.

Table 69: PICO characteristics of review question

Population	Adults with confirmed IPF
Intervention/s	Time of assessment for lung/pulmonary transplantation
Comparison/s	• Different timings in the IPF care pathway according to the different levels of disease severity • No assessment
Outcomes	Critical outcomes • All cause and IPF related mortality • 1 and 3 year survival rates Other outcomes • Cross-over time • Hospitalisations due to IPF complications (including IPF exacerbations) • Improvement of health-related quality of life • Occurrence of lung transplantation
Study design	RCTs, systematic reviews of RCTs and cohorts

The objective of this review question was to determine when in the IPF care pathway a patient with confirmed IPF should be considered for lung transplantation referral. The GDG put particular importance on the timing or the stage of disease progression when the patient should be referred for lung transplantation, in order to fully benefit from the intervention and be at the optimal point for consideration by the transplantation centre. No restrictions were used for sample size or publication date and studies in abstract form were also considered in order to capture all relevant data. The population was restricted to IPF patients only as other ILD and respiratory diseases have different disease trajectories and outcomes to lung transplantation compared to IPF patients.

12.3 Clinical evidence

No directly relevant clinical studies comparing different timing s for the assessment of lung transplantation (LTX) in a population of IPF were identified. Due to the lack of directly relevant evidence we reviewed several studies which were identified in the search which gave information on different areas related to the clinical question for the GDG to consider when making their decision.

Two papers on the Lung Allocation Score (LAS),[15,22] one paper on 6MWD and waiting list for LTX mortality,[58] and two papers on waiting list for LTX mortality[14,97] were included.

The LAS is a system used in the USA to inform donor organ allocation to registered patients. It is designed to estimate their survival benefit from a LTX. The LAS is calculated on the basis of clinical data collected for each patient, including information such as functional status, exercise capacity, lung function, haemodynamic data and the need for oxygen or ventilation support. Transplant benefit, and thus priority, is determined by predictive models that weigh medical urgency (risk of death while on waiting list) against expected outcome (post-transplant survival at 1 year). The main objectives guiding development of the LAS were to minimize waiting list mortality, maximize transplant benefit, and ensure the efficient and equitable allocation of donor lungs. The two papers[15,22] included look at patient outcomes before the LAS was implemented and after. Both were observational studies.

Charman et al[14] looked at the waiting list mortality of patients with various chronic lung diseases, including patients with pulmonary fibrosis. They stratified their analysis for patients listed for single, double and heart and LTX as part of their observational study. Paik et al[97] also undertook a study looking at the waiting list mortality of patients with various chronic lung diseases listed for LTX and stratified the results for IPF patients. Kadikar et al[58] conducted an observational study in which the usefulness of the 6MWT as a guide for LTX assessment was investigated. The population studied included patients with various chronic lung diseases such as cystic fibrosis and alpha -1-antitrypsin deficiency; however results were stratified for people with IPF therefore only data on the 26 IPF patients has been reported on in this review.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F study and selection flow chart in Appendix Q and exclusion list in Appendix R.

12.3.1 Summary of included studies

Table 70: Summary of studies included in this review

Study	Intervention / control	N	Outcomes	Comments
Lung allocation score				
Chen 2009[15]	Pre LAS vs. post LAS (3 year period after the implementation of LAS).	Pre LAS N= 4119 (IPF= 1418, 34%) Post LAS N=3833 (IPF= 1563, 41%)	Proportion receiving LTX at 6 and 12 months Waiting list mortality at 6 and 12 months Post LTX mortality at 6 and 12 months	Change in referral patterns. Secular trends (advances in surgical techniques, perioperative management and immunosuppressant therapy). Factors determined at organ matching may have a large impact on who receives LTX. No indication of disease severity at baseline.
De Oliveira	Pre LAS vs. post LAS	Pre LAS N=	Hospital mortality	Similar baseline

Study	Intervention / control	N	Outcomes	Comments
2012[22]		51(IPF= 33, 64.7%) Post LAS N=56 (IPF=46, 82.1%)	Survival at 1, 3 and 5 years.	characteristics – however higher frequency of history of diabetes, and smoking in LAS group (p=0.02). Small sample size. Single centre study – lack of generalisability. Changes in medical management between post and pre LAS time.
6MWD & waiting list mortality				
Kadikar 1997[58]	Cohort collected from January 1991 to June 2005	N= 144 (26 IPF)	Transplanted Remained on waiting list Died on waiting list / during assessment.	6MWD not documented for 7/26 IPF patients and no analysis conducted for IPF alone. Single centre study – lack of generalisability.
IPF and waiting list mortality				
Charman 2002[14]	Cohort collected from April 1984 – September 1999. Outcomes of patients accepted for LTX.	N=653 (100 PF)	Died on Waiting List Removed or still waiting Number transplanted Days Waiting Post-transplant survival days Risk of death after transplant relative to that of continued waiting at 1, 6 and 12 months	Doesn't specify IPF: cohort is all pulmonary fibrosis. Doesn't account for any confounders presented as crude data.
Paik 2012[97]	Cohort collected from May 1996 to May 2011.	146 (61 IPF)	Died on Waiting List Number transplanted	Doesn't account for any confounders presented as crude data.

12.3.2 Study quality and summary of findings

Table 71: Clinical evidence profile: Lung allocation score, pre and post implementation.

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LAS	Control	Relative risk (95% CI)	Absolute risk	
Survival at 1 year, De Oliveira 2012[22]											
1	Observational study	Serious[1,2,5,7,8,9]	Non-applicable	No serious indirectness	Serious[6]	None	27/32 (84.4%)	28/36 (77.8%)	RR 1.08 (0.86 to 1.36)	62 more per 1000 (from 109 fewer to 280 more)	Very low
Survival at 3 years, De Oliveira 2012[22]											
1	Observational study	Serious[1,2,5,7,8,9]	Non-applicable	No serious indirectness	Very serious[10]	None	3/4 (75%)	17/26 (65.4%)	RR 1.15 (0.61 to 2.16)	98 more per 1000 (from 255 fewer to 758 more)	Very low
Hospital mortality, De Oliveira 2012[22]											
1	Observational study	Serious[1,2,5,7,8,9]	Non-applicable	No serious indirectness	Very serious[10]	None	2/46 (4.3%)	3/33 (9.1%)	RR 0.48 (0.08 to 2.7)	47 fewer per 1000 (from 84 fewer to 155 more)	Very low
Transplanted at 6 months, Chen 2009[15]											
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Serious[6]	None	1063/1563 (68%)	369/1418 (26%)	RR 2.61 (2.38 to 2.87)	419 more per 1000 (from 359 more to 487 more)	Very low
Transplanted at 12 months, Chen 2009[15]											
1	Observational	Serious[1,2,3,4,5]	Non-	No serious	Serious[6]	None	1204/1563	539/1418	RR 2.03	392 more per	Very

Quality assessment							No of patients		Effect		Quality
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**LAS**	**Control**	**Relative risk (95% CI)**	**Absolute risk**	
	study		applicable	indirectness			(77%)	(38%)	(1.89 to 2.18)	1000 (from 338 more to 449 more)	low
Post LTX mortality at 6 months, Chen 2009[15]											
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	No serious imprecision	None	219/1563 (14%)	199/1418 (14%)	RR 1 (0.84 to 1.19)	0 fewer per 1000 (from 22 fewer to 27 more)	Very low
Post LTX mortality at 12 months, Chen 2009[15]											
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	No serious imprecision	None	313/1563 (20%)	298/1418 (21%)	RR 0.95 (0.83 to 1.1)	11 fewer per 1000 (from 36 fewer to 21 more)	Very low
Waiting list mortality at 6 months, Chen 2009[15]											
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Serious[6]	None	141/1563 (9%)	213/1418 (15%)	RR 0.6 (0.49 to 0.73)	60 fewer per 1000 (from 41 fewer to 77 fewer)	Very low
Waiting list mortality at 12 months, Chen 2009[15]											
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Serious[6]	None	172/1563 (11%)	298/1418 (21%)	RR 0.52 (0.44 to 0.62)	101 fewer per 1000 (from 80 fewer to 118 fewer)	Very low
Re-admission <30 days, De Oliveira 2012[22]											
1	Observational study	Serious[1,2,5,7,8,9]	Non-applicable	No serious indirectness	Very serious[10]	None	11/46 (23.9%)	7/33 (21.2%)	RR 1.13 (0.49 to	28 more per 1000 (from	Very low

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LAS	Control	Relative risk (95% CI)	Absolute risk	
									2.6)	108 fewer to 339 more)	

[1] *Changes in referral patterns and listing practices may have contributed to the effect size*
[2] *Secular trends such as advances in surgical techniques, preoperative management and immunosuppressive therapy may have contributed to the effect size*
[3] *No indication of disease severity at baseline - possibility that one group may have had sicker population*
[4] *The post LAS group had a slightly older population compared to the pre LAS group - 55-9 vs. 58-9*
[5] *No blinding of investigators was reported*
[6] *Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower* 95% CI crossed the upper MID
[7] *Higher frequency of history of diabetes, and smoking in the post LAS group (p=0.02)*
[8] *Small sample sizes*
[9] *Single centre study, results may not be generalisable to other populations*
[10] *Outcomes were downgraded by two increments if the upper CI simultaneously crossed the upper MID and the lower CI crossed the lower MID*

Table 72: Clinical evidence profile: Lung allocation score, pre and post implementation – skewed data

The table below summarises the findings from the De Oliveira 2012[22], which report median and IQR, these data are reported as skewed data. As raw figures/ mean ± SD weren't reported this data was not meta-analysed and the findings are presented below as reported in the paper.

Quality assessment							No of patients	Effect Median (IQR)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Time on waiting list, De Oliveira 2012[22]									
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Could not be calculated	None	79 IPF Pre LAS N= 33, Post LAS N=46	Pre LAS: 209(113-379) Post LAS: 65(14-209) P: <0.01	Very low

Quality assessment							No of patients	Effect Median (IQR)	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Length of ICU stay, De Oliveira 2012[22]									
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Could not be calculated	None	79 IPF Pre LAS N= 33, Post LAS N=46	Pre LAS: 6(4-16) Post LAS: 3(2-7) P: <0.01	Very low
Length of hospital stay, De Oliveira 2012[22]									
1	Observational study	Serious[1,2,3,4,5]	Non-applicable	No serious indirectness	Could not be calculated	None	79 IPF Pre LAS N= 33, Post LAS N=46	Pre LAS: 23(16-42) Post LAS: 11(9-17) P: <0.01	Very low

[1] Changes in referral patterns and listing practices may have contributed to the effect size
[2] Secular trends such as advances in surgical techniques, preoperative management and immunosuppressive therapy may have contributed to the effect size
[3] No indication of disease severity at baseline - possibility that one group may have had sicker population
[4] The post LAS group had a slightly older population compared to the pre LAS group - 55-9 vs. 58-9
[5] No blinding of investigators was reported
Note: imprecision could not be calculated for observational data which provided only median or IQRs for outcomes.

Table 73: Clinical evidence profile: 6MWD & waiting list mortality

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Number of patients transplanted (n), Kadikar 1997[58]									

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
1	Observational study	Serious[1,2,3]	Non-applicable	No serious indirectness	Could not be calculated	None	26 IPF	6/26 (23%)	Very Low
Number of patients remaining on waiting list (n), Kadikar 1997[58]									
1	Observational study	Serious[1,2,3]	Non-applicable	No serious indirectness	Could not be calculated	None	26 IPF	9/26 (35%)	Very low
Number of patients who died on waiting list/ during assessment (n), Kadikar 1997[58]									
1	Observational study	Serious[1,2,3]	Non-applicable	No serious indirectness	Could not be calculated	None	26 IPF	11/26 (42%)	Very low
6MWD (m) Kadikar 1997[58]									
1	Observational study	Very serious[1,2,3,4]	Non-applicable	Serious[5]	Could not be calculated	None	26 IPF	Patients on waiting list/transplanted: 364.3±122.8 N=13 Patient who died: 214.9±143.6 N=6 P=0.057	Very low

[1] Small sample size and single centre - results may not be generalizable to other populations
[2] No blinding of investigators was reported
[3] Doesn't account for any confounders presented as crude data
[4] 6MWD not documented for 7/26 IPF patients
[5] No analysis conducted for IPF alone
Note: imprecision could not be calculated for observational data which provided only median or IQRs for outcomes.

Table 74: Clinical evidence profile: IPF and waiting list mortality

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
Number of patients who died on waiting List (n), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 33 (33%) Single LTX: 18 (29%) Double/Heart LTX: 15 (41%)	Very low
Number of patients who died on waiting List (n), Paik 2012[97]									
1	Observational study	Serious[3,4]	Non-applicable	No serious indirectness	Could not be calculated	None	61 IPF	35 (57.4%)	Very low
Number of patients removed or still waiting (n), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 7(7%) Single LTX: 3(5%) Double/Heart LTX: 4(11%)	Very low
Number of patients removed or still waiting (n), Paik 2012[97]									
1	Observational study	Serious[3,4]	Non-applicable	No serious indirectness	Could not be calculated	None	61 IPF	3(4.9%)	Very low
Number of patients transplanted (n), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 60(60%) Single LTX: 42(67%) Double/Heart LTX: 18(49%)	Very low
Number of patients transplanted (n), Paik 2012[97]									
1	Observational study	Serious[3,4]	Non-applicable	No serious indirectness	Could not be calculated	None	61 IPF	23 (37.7%)	Very low
Days Waiting (median (IQR)), Charman 2002[14]									

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations			
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 117 (43, 231) Single LTX: 104 (5,194) Double/Heart LTX: 147 (94,305)	Very low
Post-transplant survival days (median (95% CI)), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 931 (98,1764) Single LTX: 449 (0,1287) Double/Heart LTX: 1121 (0, 3024)	Very low
Risk of death after transplant relative to that of continued waiting at 1 month (RR), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 2.23 Single LTX: 1.96 Double/Heart LTX: 2.88	Very low
Risk of death after transplant relative to that of continued waiting at 6 months (RR), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 0.65 Single LTX: 0.71 Double/Heart LTX: 0.57	Very low
Risk of death after transplant relative to that of continued waiting at 12 months (RR), Charman 2002[14]									
1	Observational study	Serious[2,3,4]	Non-applicable	Serious[1]	Could not be calculated	None	All : 100 Single LTX : 63 Double/Heart LTX : 37	All: 0.46 Single LTX: 0.54 Double/Heart LTX: 0.36	Very low

[1] Population studied was patients with pulmonary fibrosis the proportion of IPF patients is unknown, if there were any at all.
[2] Single centre study - results may not be generalizable to other populations

[3] No blinding of investigators was reported
[4] Doesn't account for any confounders presented as crude data
Note: imprecision could not be calculated for observational data which provided only median or IQRs for outcomes.

12.4 Economic evidence

Published literature

No relevant economic evaluations comparing different timing of LTX in a population of IPF were identified.

One study that met the inclusion criteria was selectively excluded due to the sample of the study having a low proportion of IPF patients [105] this is summarised in Appendix R, with reasons for exclusion given.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs were provided to aid consideration of cost effectiveness. The unit cost of an elective inpatient LTX was £41,684 (IQR: £25,203 to £49,045), with an average length of stay of 18.6 days. Each additional excess bed day costed £578 (IQR: £349 to £769). A unit cost weighted by excess bed days was calculated to be £42,018[25].

12.5 Evidence statements

Clinical

No directly relevant clinical studies comparing different timings of lung transplantation in a population of IPF were identified.

Indirect evidence used by GDG for information – observational data:

Lung allocation score:

Survival

Very low quality evidence showed that there was no clinically effective difference in survival at 1 and 3 years between patients who underwent LTX post LAS implementation compared to those who underwent LTX pre LAS implementation [1 study N=79].

Mortality

Very low quality evidence showed that the LAS is clinically effective at reducing hospital mortality compared to pre LAS implementation [1 study N=79].

Very low quality evidence showed that the LAS is not clinically effective at reducing post LTX mortality at 6 and 12 months compared to pre LAS implementation [1 study N=2981].

Very low quality evidence showed that the LAS is not clinically effective at reducing waiting list mortality at 6 and 12 months compared to pre LAS implementation [1 study N=2981].

Transplantation

Very low quality evidence showed that the LAS may be clinically effective at increasing chances of having a LTX at 6 and 12 months from listing compared to pre LAS implementation [1 study N=2981].

Re-admission

Very low quality evidence showed that the LAS is not clinically effective at reducing re admission to hospital within 30days or less from being discharged after LTX compared to pre LAS implementation, but the direction of the estimate of the effect could favour either [1 study N=79].

Time on waiting list

Imprecision and clinical effectiveness could not be assessed for the following outcomes.

Very low quality evidence assessed showed that LAS reduced time on LTX waiting list compared to pre LAS implementation [1 study N=79].

Very low quality evidence showed that LAS reduced length of ICU stay compared to pre LAS implementation [1 study N=79].

Very low quality evidence for which imprecision and clinical effectiveness could not be assessed showed that LAS reduced length of hospital stay compared to pre LAS implementation [1 study N=79].

6MWD and IPF waiting list characteristics

Imprecision and clinical effectiveness could not be assessed for the following outcome.

Very low quality evidence showed a non-significant difference in 6MWD in people with IPF between those who died on waiting list and those alive/transplanted [1 study N= 26].

IPF waiting list characteristics

Imprecision and clinical effectiveness could not be assessed for the following outcomes.

Very low quality evidence showed that 23% of IPF patients waiting for LTX were transplanted during the 5 year study period [1 study N= 26].

Very low quality evidence showed that 37.7% of IPF patients waiting for LTX were transplanted during the 9 year study period [1 study N= 61].

Very low quality evidence showed that 35% of IPF patients waiting for LTX still remained on the waiting list after the 5 year study period [1 study N= 26].

Very low quality evidence showed that 4.9% of IPF patients waiting for LTX still remained on the waiting list after the 9 year study period [1 study N= 61].

Very low quality evidence showed that 42% of IPF patients placed on LTX waiting list had died waiting or during the assessment period [1 study N= 26].

Very low quality evidence showed that 57.4% of IPF patients placed on LTX waiting list had died waiting or during the 9 year study period [1 study N= 61].

Very low quality evidence showed that 33% of IPF patients placed on LTX waiting list had died waiting of which 29% were waiting for a single LTX and 41% were waiting for double/heart transplant [1 study N= 100].

Very low quality evidence showed that 7% of IPF patients placed on LTX waiting list were still waiting, of which 5% were waiting for a single LTX and 11% were waiting for double/heart transplant [1 study N= 100].

Very low quality evidence showed that 60% of IPF patients placed on LTX waiting list had been transplanted of which 67% were waiting for a single LTX and 49% were waiting for double/heart transplant [1 study N= 100].

Very low quality evidence showed that IPF patients placed on LTX waiting list had waited a median (IQR) 117 (43,231) days. This was lower for patients waiting for a single LTX and longer for patients waiting for double/heart LTX [1 study N= 100].

Very low quality evidence showed that IPF patients placed on LTX waiting list had a post LTX survival of median (95% CI) 931 (98, 1764) days. This was lower for patients waiting for a single LTX and higher for patients waiting for double/heart LTX [1 study N= 100].

Very low quality evidence showed that IPF patients placed on LTX waiting list had a risk of death relative to that of continued waiting at 1 month of 2.23. This was lower for patients waiting for a single LTX and higher for patients waiting for double/heart LTX [1 study N= 100].

Very low quality evidence showed that IPF patients placed on LTX waiting list had a risk of death relative to that of continued waiting at 6 month of 0.65. This was higher for patients waiting for a single LTX and lower for patients waiting for double/heart LTX [1 study N= 100].

Very low quality evidence showed that IPF patients placed on LTX waiting list had a risk of death relative to that of continued waiting at 12 month of 0.46. This was higher for patients waiting for a single LTX and lower for patients waiting for double/heart LTX [1 study N= 100].

Economic

- No relevant economic evaluations were identified.

12.6 Recommendations and link to evidence

Recommendations	**32. Discuss lung transplantation as a treatment option for people with idiopathic pulmonary fibrosis who do not have absolute contraindications. Discussions should:** • **take place between 3 and 6 months after diagnosis or sooner if clinically indicated** • **be supported by an interstitial lung disease specialist nurse** • **include the risks and benefits of lung transplantation** • **involve the person's family and carers with the person's consent.** **(See recommendations 18 – 23 about best supportive care.)** **33. Refer people with idiopathic pulmonary fibrosis for lung transplantation assessment if they wish to explore lung transplantation and if there are no absolute contraindications. Ask the transplant centre for an initial response within 4 weeks.**
Relative values of different outcomes	The GDG considered mortality and survival to be the critical outcomes to inform this recommendation. The GDG considered the prognosis of people with IPF pre and post lung transplantation, whilst also considering when a patient had been assessed and referred for lung transplantation, as well as the length of assessment, length of waiting times and availability of lung donors.
Trade-off between clinical benefits and harms	This recommendation was based on GDG consensus as no directly relevant studies on the optimal timing to refer a patient with IPF for lung transplantation were retrieved. The GDG discussed the harms and benefits associated with referring people with IPF for lung transplantation at different time-points in the care pathway, as well as the complications post-surgery associated with the single and bilateral procedures. Early assessment and referral for lung transplantation could increase the

	probability of survival; improve symptoms, and quality of life (physical and mental components) post transplantation. It was recognised that a patient's prognosis, the unknown rate of disease progression, risk of acute exacerbation and length of waiting times due to donor organ availability, were all factors to acknowledge when considering whether a patient would benefit from lung transplantation. That the status of a patient with IPF initially deemed suitable for lung transplantation, may change and some patients accepted for transplantation may deteriorate to the point of no longer being actively listed. As well as considering a patient's prognosis and clinical suitability for lung transplantation, the GDG acknowledged that a patient's social, financial and mental well-being (support from family and carers, and psychosocial support) would have a considerable impact on their eligibility for an invasive procedure. The GDG agreed that a patient should also be assessed on their social and mental capacity for lung transplantation. Complications associated with transplantation may include cellular or humeral rejection, infection, and primary organ dysfunction and airway complications.
Economic considerations	There was no economic evaluation to inform this recommendation. The GDG discussed the economic implications of different referral strategies in the context of a limited supply of suitable donor organ availability. Principle drivers of cost effectiveness of lung transplant were identified and discussed in deliberations. These included: • the high cost of transplant, with an estimated two thirds of care costs arising post transplantation (with the majority of care costs arising in the first year and decreasing thereafter). • the frequency/cost of post-transplant rehospitalisation and post-transplant medication compared to the frequency/cost of hospitalisation and medication whilst waiting for transplant. • the marginal gains in life expectancy (and quality of life) compared with conservative care. It was acknowledged that lung transplant carries a very high unit cost and therefore should be offered to patients who would achieve maximal health incremental benefit in comparison to what would have been achieved without lung transplant. Because of the potential rapid decline of IPF patients, and a short life expectancy of these patients, IPF patients were considered likely to be high priority candidates on this basis, given the lack of any alternative disease modifying treatment. However, the review question does not seek to answer whether LTX is cost effective, but rather at which time point is it most cost effective to refer patients with IPF for LTX. In regard to optimal timing of lung transplant, the GDG discussed the difficulties in predicting a sudden decline. It was felt however, that given the potential that marginal life expectancy gain in IPF patients may be greater than for other respiratory conditions given lack of available treatment and short life expectancy, and therefore using this comparison the lung transplant is likely to be seen as a cost effective intervention, it should be made available as an intervention in this population group as much as possible. Therefore the best means of ensuring this was to alert a lung transplant centre as soon as possible about any potential candidates without absolute contraindications, i.e. at the point of diagnosis. Referral of all potential candidates without absolute contraindications on the point of diagnosis would incur additional cost to the NHS, as currently this is not common

	practice; however the cost impact implications are not clear, in part because it's uncertain how many newly diagnosed patients would not have absolute contraindications. There is also variation in practice in what clinical information is required to make a referral, and therefore resource implications were difficult to estimate. Some clinical members reported the need to undertake many investigations ranging from up to date CT scans and routine urine and blood samples (which would incur minimal incremental cost to the recommended care pathway) to DEXA scans, HIV tests and angiograms (which would incur additional cost to the recommended care pathway). However, if such interventions are a driver of cost effectiveness of the lung transplant by selecting the patients who could benefit most, the cost of providing such information could be justified. Additionally, given the small numbers of patients who do not have absolute contraindications, overall the contribution of the cost of the tests (in comparison to the cost of transplant and downstream care) to the average cost per referred patient would be relatively small, especially if the number of transplants in this group increases. Overall, the GDG considered it appropriate to alert transplant centres through a referral letter which required minimal resource use, and if the transplant centre deemed it necessary to have further information regarding the eligibility of the patient for transplant, this information could be collected on a case by case basis. The GDG also commented that as MDTs are a recommended element of the diagnostic pathway, and referral should be considered soon after, depending on local circumstances there may be efficiencies and reduction in duplication of services with closer relationships and liaison being built between MDTs and transplant centres. The use of satellite clinics for transplant assessment was also discussed as an efficient means of allowing access.
Quality of evidence	This recommendation was based on informal GDG consensus as no studies on the optimal timing to refer a patient with IPF for lung transplantation were retrieved. Due to the lack of evidence, the GDG considered indirect data from two studies on Lung Allocation Score (LAS), one study on 6MWD and waiting list mortality, and one study on waiting list mortality. The GDG acknowledged that LAS is not used in the U.K, but agreed that the outcomes presented in the two studies which investigated the implementation of LAS and studies that provided data on waiting list mortality were important considerations for UK practice. These studies were deemed low to very low quality with a serious risk of bias affecting all outcomes as there was no accounting for confounding factors such as changes in referral patterns and listing practices, changes in secular trends such as advances in surgical techniques and baseline disease severity. One study showed that implementation of LAS increased the likelihood of lung transplantation and reduced waiting list mortality at 6 and 12 months from the time of initial listing for lung transplantation / time from lung transplantation. Another study showed that implementation of LAS improved hospital mortality as well as time on waiting list, length of ICU stay and length of hospital stay, but this data was taken from a skewed distribution and may not be generalizable to other populations. Of the three studies that reported on waiting list mortality one showed a non-significant clinical difference in 6MWD in people with IPF between those who died on waiting list and those alive/transplanted. The second showed patients waiting for single lung transplantation had a shorter waiting time, lower waiting list mortality, and more transplantations occurring during the study period than patients listed for double/heart lung transplantation. Patients who also had a double/heart lung transplant had lower post-transplant mortality. The study also reported that the relative risk of death after transplant relative to that of continued waiting decreased for all patients and both sub groups with time (1, 6 and 12 months). The

	population group in both studies was patients with pulmonary fibrosis and neither stratified analysis for IPF alone. The third study showed that of the cohort of IPF patients listed for LTX over half died whilst waiting The GDG considered the indirect evidence retrieved and agreed that determining whether a person is suitable for lung transplantation should occur as soon as possible once a confident diagnosis has been made. It was acknowledge that confirming a diagnosis in people suspected with IPF and providing adequate information and support (regarding diagnosis, prognosis and management options) may take a couple of months and that first discussions to inform people with IPF if they are suitable for lung transplantation should begin around this time. Obtaining a firm diagnosis assessing whether a person has absolute contraindications was agreed by the GDG to take around 3-6 months.
Other considerations	During GDG discussions both the pulmonary scientific council of the international society for heart and lung transplantation guidelines for the selection of lung transplantation candidates[94] and the ATS international guidelines for the selection of lung transplantation candidates[1] were considered. The GDG acknowledged that the ATS and BTS guidelines did not cover when a patient should be considered for referral to lung transplantation, but that the ATS specified referral criteria in a minority of people with IPF. The GDG considered that clinical judgement should be used to determine whether the patient is willing to be considered for lung transplantation and the importance of social, financial and mental support. People with IPF deemed suitable for lung transplantation are likely to fare better with support from family and friends to care for them pre and post operatively. This support may also help with the financial and emotional burden experienced during this time when a patient may not be fit enough to work and waiting for the procedure can impact negatively on psychosocial health.

13 Ventilation

13.1 Review introduction

People with IPF may experience acute episodes of deterioration with worsening hypoxia, increasing breathlessness and a high 'work' of breathing. Acute, or acute-on-chronic respiratory failure in IPF may be caused by a number of factors including respiratory infection, left ventricular failure, pulmonary embolism, pneumothorax and acute exacerbation of IPF (in which other causes of acute deterioration have been excluded). However, acute respiratory failure (ARF) from any cause may warrant respiratory support in the form of mechanical (invasive) or non-invasive ventilation.

The decision to ventilate should be based on the likelihood that ventilation will enhance recovery balanced against the risks. Assisted ventilation is more likely to be beneficial if there is a definite reversible cause for acute deterioration. However in IPF, even potentially reversible causes occur on a background of a progressive form of lung fibrosis. Invasive ventilation in particular has associated risks, including the possibility of further harming the lung due to ventilator-associated injury and infection. Ventilating people with IPF is difficult because the lungs are stiff and noncompliant. Ventilation requires intensive monitoring in high dependency or intensive care units and consumes resources. It is known that the vast majority of people with IPF will die while receiving mechanical ventilation or shortly after discharge from the intensive care unit (ICU). Instituting assisted ventilation in people with IPF is of questionable value, and may even be futile. Appropriate discussion regarding ventilation is an important component of the management of people with IPF. People with IPF being considered for lung transplantation are only rarely transplanted if mechanically ventilated due to the high associated mortality. 'Bridging' people with IPF and acute respiratory failure to pulmonary transplantation using extra-corporeal membrane oxygenation (ECMO) or Novalung is increasingly used to avoid ventilation in selected patients already on the waiting list.

13.2 Clinical question and review methodology

The following clinical question was included in this chapter:

13.2.1 In acute or acute-on chronic respiratory failure in patients with IPF, what is the value of non-invasive and invasive ventilation?

For full details see review protocol in Appendix C.

Table 75: PICO characteristics of review question

Population	Adults with confirmed IPF
Intervention/s	Invasive ventilation
Comparison/s	• Non-invasive ventilation • No ventilation
Outcomes	Critical outcomes • Mortality (in hospital and post discharge) Other outcomes • Improvement of health-related quality of life • Hospital length of stay
Study design	RCTs and cohort studies

The objectives of this review were to determine the value of non-invasive and invasive ventilation in IPF patients with acute or acute-on chronic respiratory failure. No restrictions were used for sample

size and publication date. Studies in abstract form were also included in order to capture all relevant data. Studies with indirect populations such as COPD were not considered as the GDG considered that they have different disease trajectories and needs and are thus not comparable with people with IPF.

13.3 Clinical evidence

We searched for randomised control trials and cohort studies comparing the effectiveness of invasive mechanical ventilation (IMV) versus non-invasive mechanical ventilation (NIMV) for patients with IPF.

Seven studies were included in this review [6,9,35,77,110,120,132]. No randomised controlled trials were identified. All the studies included were retrospective cohorts.

Four papers compared the outcomes of patients receiving IMV versus NIMV[6,9,77,132]. In one of the studies, by Yokoyama et al[132], the primary aim was to investigate the outcomes of patients receiving NIMV. However, due to the severity of disease and complications some of the participants were also given IMV, and a post-hoc analysis was carried out for IMV versus NIMV. A major limitation of this was that baseline data was not provided per group, but only for the cohort as a whole.

An inherent limitation of all the studies reviewed was treatment cross over, as some patients was given NIMV initially but as their condition worsened IMV was initiated. In addition, patients in the IMV group may have had significantly worse disease severity at baseline, which may have confounded the results.

Due to the limited amount of evidence available we have also included studies which only looked at one intervention (IMV or NIMV) in a cohort of patients and reported the relevant outcomes. Two studies were identified which looked at a cohort of patients receiving IMV alone.[35,120] Another study included by Saydain et al[110] , is a retrospective study describing the clinical course of IPF patients admitted to ICU. The investigators described the difference in mortality of patients receiving ventilation (NIMV and/or IMV) and non-ventilated patients.

Evidence from these are summarised in the clinical GRADE evidence profile below. See also the forest plots in Appendix E, study evidence tables in Appendix F, study and selection flow chart in Appendix Q and exclusion list in Appendix R.

13.3.1 Summary of included studies

Table 76: Summary of studies included in the review

Study	Intervention/ comparison	Population	Outcomes	Comments
Alhameed 2004[6]	IMV vs. NIMV	All patients with IPF requiring MV for unknown causes of ARF N= 25	In hospital mortality Mortality at 6 months	• Cross over between treatment groups • Generalisability • Confounding factors weren't accounted for
Blivet 2001[9]	IMV vs. NIMV	IPF patients admitted for ARF N= 15	In hospital mortality	• Generalisability • Cross over between treatment groups • Confounding factors weren't accounted for
Mollica 2010[77]	IMV vs. NIMV	IPF patients admitted for ARF N= 34	In hospital mortality Mortality at 6	• The disease severity was quite different between the 2 groups, with patients

Study	Intervention/ comparison	Population	Outcomes	Comments
			months	undergoing IMV showing a significantly higher APACHE II score as compared with subjects undergoing NIV • Confounding factors weren't accounted for
Yokoyama 2010 [132]	IMV vs. NIMV	Patients with acute exacerbation of IPF N=11	In hospital mortality	• Post hoc analysis of NIMV vs. IMV • Baseline data not given per group
Fumeaux 2001 [35]	Observational data for patients receiving IMV	IPF patients admitted for ARF who required IMV N=14	In hospital mortality	• No comparison/ control group • Observational data • Generalisability
Stern 2001[120]	Observational data for patients receiving IMV	Patients with pulmonary fibrosis requiring MV for ARF N=23	In hospital mortality	• No comparison/ control group • Observational data • Generalisability
Saydain 2002[110]	Observational data for ventilated patients vs. no ventilation	Patients with IPF admitted to ICU N= 38	In hospital mortality	• Cause of admission may not have been ARF.

13.3.2 Study quality and summary of findings

Table 77: Clinical evidence profile: invasive mechanical ventilation vs. non-invasive mechanical ventilation

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mechanical ventilation	Non-invasive ventilation	Relative risk (95% CI)	Absolute risk	
Mortality (in hospital), Alhameed 2004[6], Blivet 2001[9], Mollica 2010[77], Yokoyama 2010[132]											
4	Observational studies	Very serious[1,2,3,4,5]	No serious inconsistency	No serious indirectness	Serious[6]	None	50/52 (96.2%)	20/33 (60.6%)	RR 1.57 (1.18 to 2.09)	345 more per 1000 (from 109 more to 661 more)	Very low
Mortality (6 months), Alhameed 2004[6], Mollica 2010[77]											
2	Observational studies	Very serious[1,2,3,4,5]	No serious inconsistency	No serious indirectness	No serious imprecision	None	36/36 (100%)	22/23 (95.7%)	RR 1.03 (0.90 to 1.19)	29 more per 1000 (from 96 fewer to 182 more)	Very low

[1] Observational data biases
[2] Generalisability of findings is limited as the data come from single centres with small population sizes, effect size could be impacted by variations in patient characteristics and practice
[3] Cross over in treatment between the groups
[4] Confounding factors aren't accounted for
[5] Disease severity is worse in the mechanical ventilation groups at baseline compared to the non-invasive ventilation group. (Mollica 2010[77])
[6] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID

Table 78: Clinical evidence profile: patients receiving invasive mechanical ventilation alone

Quality assessment							No of patients Mechanical ventilation	Results	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Mortality (in hospital)	
Mortality (in hospital),Fumeaux 2001[35]									

Quality assessment							No of patients Mechanical ventilation	Results	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Mortality (in hospital)	
1	Observational study	Very serious[1,2,3]	No serious inconsistency	No serious indirectness	Could not be calculated	None	N=14	14/14 (100%)	Very low
Mortality (in hospital),Stern 2001[120]									
1	Observational study	Very serious[1,2,3]	No serious inconsistency	No serious indirectness	Could not be calculated	None	N=23	22/23 (96%)	Very low

[1] Observational data biases
[2] Generalisability of findings is limited as the data comes from single centres with small population sizes, effect size could be impacted by variations in patient characteristics and practice
[3] No comparison or control groups
Note: imprecision could not be calculated for observational studies with no comparison groups.

Table 79: Clinical evidence profile: patients receiving ventilation vs. no ventilation

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Ventilation	No ventilation	Relative risk (95% CI)	Absolute risk	
Mortality (in hospital)											
1	Observational studies	Very serious[1,2,3]	No serious inconsistency	Serious[5]	Serious[4]	None	13/19 (68.4%)	10/19 (52.6%)	RR 1.3 (0.77 to 2.2)	158 more per 1000 (from 121 fewer to 632 more)	Very low

[1] Descriptive study which plainly describes the clinical course of patients admitted to ICU-observational data biases
[2] No baseline characteristics provided
[3] Single centre study data and a small sample size, lacks generalisability the effect size could be impacted by variations in patient characteristics and variations in practice.
[4] Outcomes were downgraded by one increment if the upper or lower 95% CI crossed the lower MID or the upper or lower 95% CI crossed the upper MID
[5] Patients may not have been admitted for ARF as the study looks at all ICU admissions

13.4 Economic evidence

Published literature

No relevant economic evaluations comparing invasive and non-invasive ventilation strategies were identified. No studies that met the inclusion criteria were selectively excluded.

Unit costs

In the absence of recent UK cost-effectiveness analysis, relevant unit costs are provided in Table 80 below.

Table 80: NHS reference costs[25] for invasive and non-invasive ventilation

Reference cost HRG	National average unit cost	Lower Quartile Unit Cost	Upper Quartile Unit Cost	Average cost of excess bed day	Lower Quartile Unit Cost	Upper Quartile Unit Cost	Weighted national average	Weighted average length of stay
Respiratory Failure with Intubation with Major CC (DZ27A); as recorded for Non Elective Inpatients (a)	£2,973	£1,130	£3,504	£251	£224	£284	£3,426	8.63
Respiratory Failure with Intubation with CC (DZ27B); as recorded for Non Elective Inpatients (b)	£1,631	£487	£2,301	£466	£168	£227	£1,864	6.27
Respiratory Failure with Intubation without CC (DZ27C); as recorded for Non Elective Inpatients (c)	£5,218	£1,839	£8,056	NA	NA	NA	£5,218	11.33
Weighted for complications and co morbidities for HRG codes: DZ27A, DZ27B and DZ27C; as recorded for Non Elective long stay inpatients							**£3,275**	**8.40**
Respiratory Failure without Intubation with Major CC (DZ27D); as recorded for Non Elective Inpatients (d)	£2,395	£1,628	£2,613	£236	£176	£271	£2,706	8.57
Respiratory Failure without Intubation with CC (DZ27E); as recorded for Non Elective Inpatients (e)	£1,974	£1,293	£2,263	£235	£186	£280	£2,255	7.02
Respiratory Failure without Intubation without CC (DZ27F); as recorded for Non Elective Inpatients (f)	£1,358	£906	£1,625	£217	£183	£245	£1,743	5.39
Weighted for complications and co morbidities for HRG codes: DZ27D, DZ27E and DZ27F; as recorded for Non Elective long stay inpatients							**£2,570**	**8.10**
Non-Invasive Ventilation Support Assessment 19 years and over (DZ37A); as recorded for Non Elective Inpatients (g)	£996	£298	£880	NA	NA	NA	**£996**	**1.82**

Note that COPD patients would not be coded under HRG code DZ27 as a separate code is available.NA = Not applicable as no data submissions recorded.

(a) The number of data submissions for this code was 65, with 167 units of activity.

(b) The number of data submissions for this code was 18, with 22 units of activity.

(c) The number of data submissions for this code was 3, with 3 units of activity.

(d) The number of data submissions for this code was 156, with 6000 units of activity.

(e) The number of data submissions for this code was 151, with 2138 units of activity.

(f) The number of data submissions for this code was 78, with 173 units of activity.

(g) The number of data submissions for this code was 41, with 850 units of activity.

13.5 Evidence statements

Clinical

NIMV versus IMV:

Mortality

Very low quality evidence suggested that NIMV is potentially more clinically effective when compared to IMV at reducing in hospital mortality (4 studies, N=85).

Very low quality evidence showed that NIMV is more effective when compared to IMV at reducing mortality at 6 months but the effect size is too small to be clinically important (2 studies, N=59).

Observation data for IMV alone:

Mortality

Very low quality evidence from two studies showed that patients with IPF with acute respiratory failure had an in-hospital mortality rate of 100% and 96% for patients receiving invasive mechanical ventilation (2 studies, N=37).

Ventilation versus no ventilation:

Mortality

Very low quality evidence suggested that patients admitted to ICU who were ventilated (invasive or non-invasive ventilation) had a higher in hospital mortality rate compared to patients who had not been ventilated (one study, N=38).

Economic

No relevant economic evaluations were identified.

13.6 Recommendations and link to evidence

Recommendations	**34. A respiratory physician or specialist nurse with an interest in interstitial lung disease should discuss the poor outcomes associated with mechanical ventilation (including non-invasive mechanical ventilation) for respiratory failure with people with idiopathic pulmonary fibrosis. These discussions should ideally take place between 3 to 6 months after diagnosis or sooner if clinically indicated. (See recommendations 18 – 23 about best supportive care.)** **35. Do not routinely offer mechanical ventilation (including non-invasive mechanical ventilation) to people with idiopathic pulmonary fibrosis who develop life-threatening respiratory failure.**
Relative values of different outcomes	The GDG considered mortality to be the critical outcome for informing health professionals whether to pursue ventilation (invasive or non-invasive) in people with IPF. Length or hospital stay and improvements in health related quality of life were also considered important outcomes, especially as the GDG recognised that the decision to ventilate a patient would also impact on the patient's suitability for lung transplantation.

Trade-off between clinical benefits and harms	The GDG considered the harms and benefits of ventilating people with IPF and discussed the importance of potentially reversible causes of deterioration in IPF such as infections and pneumothorax when considering whether to ventilate. The difficulty of ventilating people with IPF was acknowledged, as their lungs are poorly compliant with widespread collapse and shunting. The high mortality rates of mechanical ventilation and non-invasive ventilation (unless there is a reversible cause) and the potential for mechanical bridging with Novalung or ECMO in specialised centres was also discussed. However, Novalung and ECMO are not available nationally, and commissioned only in transplant unit centres. The GDG agreed that the harms associated with ventilation outweighed the benefits.
Economic considerations	There was no economic evidence to review to help inform recommendations regarding invasive and non-invasive ventilation. The GDG considered the relatively high unit costs for this intervention, noting the clinical experience and the evidence retrieved by the review did not suggest a clinical benefit of using this intervention for the majority of IPF patients. The lack of clinical benefit was a key factor in deciding against recommending invasive or non-invasive ventilation in the routine management of people with IPF, and unless there was a high probability of recovery (i.e. used as a bridge to transplant or used whilst treating a reversible cause) either form of ventilation was unlikely to be cost effective.
Quality of evidence	The GDG considered evidence from three studies that investigated the outcomes for patients receiving invasive mechanical ventilation (IMV) versus invasive mechanical ventilation (NIMV) alone), and four studies that compared the effectiveness of invasive mechanical ventilation (IMV) versus non-invasive mechanical ventilation (NIMV), for people with IPF. An inherent limitation of the four studies comparing ventilation types was the treatment cross over, as some patients were given NIMV initially and also received IMV as their condition worsened. In addition, patients in the IMV group may have had significantly worse disease severity at baseline, which may have confounded the results, but these data were not always provided. The GDG agreed by consensus that it was important to provide people with IPF information on ventilation (the risks and benefits) as soon as possible once a confident diagnosis of IPF had been made, because they recognised that this is a topic which is often not discussed. Three to six months was chosen as the optimum time to have this informative discussion regarding ventilation as this is often the time it takes to confirm a diagnosis of IPF. The GDG also recognised that beyond this time frame, the risk of a person with IPF becoming acutely unwell and requiring ventilation may increase.
Other considerations	The GDG discussed the importance of patient's preference for ventilation. It was recognised that when patients experience severe respiratory failure, the patient, their family and carers often do not have time to come to terms with what is happening. Therefore, it was considered that support from healthcare professionals to educate patients on prognosis, as well as the options and side effects regarding types of ventilation and lung transplantation, should be discussed with patients and their family soon after diagnosis.

14 Review and follow-up

14.1 Review introduction

The majority of people with IPF are given their diagnosis during a secondary care consultation in a hospital setting. Their management plan and follow-up arrangements are also traditionally organised in secondary care. Given its poor prognosis many people with IPF remain under the care of a hospital consultant, most commonly a chest physician. In a small number of patients who may be suitable for lung transplantation, close monitoring is required. The commonest symptoms of IPF are progressive breathlessness and cough which is sometimes intractable. People with IPF have an increased risk of developing lung cancer and cardiovascular problems. The coordination of palliative care, smoking cessation strategies and prevention and treatment of secondary complications require close liaison between primary and secondary care. However, there is national variation in the frequency and duration of follow-up of people with IPF. There are a small but increasing number of specialist interstitial lung disease nurses in the UK who, where available, play an important role in bridging the gap between secondary and primary care.

14.2 Clinical question and review methodology:

The following clinical questions were included in this chapter.

14.2.1 How often should a patient with confirmed diagnosis of IPF be reviewed?

14.2.2 In which healthcare setting and by whom should a review appointment for patients with confirmed IPF be conducted?

For full details see review protocol in Appendix C.

Table 81: PICO characteristics of review question

Population	Adults with confirmed IPF
Intervention/s	• Review at 3 and 6 months • Review earlier than 3 months if clinically indicated • Review at yearly intervals
Comparison/s	• Different timing of review • No review
Outcomes	Critical outcomes • Change in percent predicted DLCO • Change in percent predicted FVC Other outcomes • Oxygen saturation at rest • Oxygen saturation on exertion • Distance walked on 6 min walk or incremental shuttle walk test • Eligibility for lung transplant
Study design	RCTs, Systematic reviews of RCTs, and cohort studies

14.3 Clinical evidence

No relevant clinical studies comparing different timings and delivery of review appointments were identified.

14.4 Economic evidence

Published literature

No relevant economic evaluations comparing different review and monitoring strategies were identified. No studies were selectively excluded.

14.5 Evidence statements

Clinical

No relevant clinical studies comparing different timings and delivery of review appointments were identified.

Econom

No relevant economic evaluations were identified.

14.6 Recommendations and link to evidence

Recommendations	**36. In follow-up appointments for people with idiopathic pulmonary fibrosis:** • **assess lung function** • **assess for oxygen therapy** • **assess for pulmonary rehabilitation** • **offer smoking cessation advice, in line with Smoking cessation services (NICE public health guidance 10)** • **identify exacerbations and previous respiratory hospital admissions** • **consider referral for assessment for lung transplantation in people who do not have absolute contraindications (see recommendations 32 and 33)** • **consider psychosocial needs and referral to relevant services as appropriate** • **consider referral to palliative care services** • **assess for comorbidities (which may include anxiety, bronchiectasis, depression, diabetes, dyspepsia, ischaemic heart disease, lung cancer and pulmonary hypertension).** **37. Consider follow-up of people with idiopathic pulmonary fibrosis:** • **every 3 months or sooner if they are showing rapid disease progression or rapid deterioration of symptoms or** • **every 6 months or sooner if they have steadily progressing disease or**

	• **initially every 6 months if they have stable disease and then annually if they have stable disease after 1 year.**
Relative values of different outcomes	The GDG considered changes in lung function to be the most critical outcome. Oxygen management and 6MWD were also considered to be important outcomes for assessing patient's prognosis at regular intervals, in order to determine the rate of disease progression and when clinical management should be reviewed.
Trade-off between clinical benefits and harms	Regular review allows the opportunity to identify when a change of management and a timely intervention is required. Frequent review is therefore a mechanism of improving clinical benefit (and cost effectiveness) of interventions that follow. No appreciable harms were identified in having regular review appointments. Potential quality of life improvements may not be realised with review of therapeutic options. The GDG also considered the evidence presented in the prognostic review and discussed whether it was possible to predict a change in a patient's symptoms to determine optimal frequency of review appointments. There was consensus that although measurement of identified prognostic markers could give a sense of likely rate of physiological decline, they may not be helpful in predicting when a change of symptoms would occur. However, with a lack of alternative available evidence, prognostic information alongside clinical history and judgement could be helpful in indicating which patients were more likely to decline rapidly (and therefore more likely to experience a change in symptoms sooner) than others. Assuming disease progression generally occurs at a linear rate, it was thought reasonable to offer review appointments at a closer time interval following diagnosis, and for patients where disease progression is established to be reasonably stable to offer review appointments over a longer interval (given it is unlikely symptoms and need for a change in management will have changed between review appointments). In acknowledgement that the natural course of IPF may be unpredictable, the GDG felt strongly that if a change in disease progression was suspected, i.e. post-acute exacerbation or through self-reported change in symptoms, review appointments should again be offered more frequently for these patients.
Economic considerations	There was no economic evidence identified to inform this recommendation. The GDG agreed that due to the specialist nature of providing care for an IPF patient, that review and follow up should involve clinical staff with specialist expertise in ILD (this may be someone who runs a service seeing at least 500 ILD patients per year or has completed specialist training in ILD for at least 6 months), and increasing the frequency of review (i.e. reducing the time interval between review appointments) would involve additional cost. The resource use and cost of frequent review appointments therefore needs to be justified by the clinical benefit brought by enabling timely intervention when a change in clinical management is required. The cost effectiveness of a more frequent review is in part determined by whether the cost effectiveness of the interventions that follow is driven by the timing or stage of disease at which they are offered. The GDG considered the purpose of the review, the change in clinical management that may occur and the impact this may make on the cost effectiveness of the interventions that may be offered (i.e. whether preventing a delay in initiating these interventions justify the cost of increased follow up). Therapeutic interventions that are recommended in this guideline for people with IPF mainly focus on symptom control and relief provided as best supportive care. However, no evidence was

	identified in regards to the optimal timing of these interventions. It was agreed that to maximise clinical benefit, a change in a person's symptoms should be identified as quickly as possible. One means of achieving this, at lower cost than scheduled review appointments, is for patient self-referral when they consider that their symptoms have changed. However, patient self-referral could result in additional and inappropriate use of specialist time, as patients may over-refer given the information publicly available on IPF and anxiety felt on learning the short median life expectancy of IPF. Finally, the GDG noted that regular review provided opportunity to discontinue ineffective or cost ineffective management. This is particularly important in the context of emerging new evidence and potentially cost effective interventions becoming available.
Quality of evidence	This recommendation was based on GDG consensus, as no evidence was retrieved to inform this question. The GDG considered the personal experiences of the patient members of the guideline group. Discussions included consideration of the following: Reassurance of monitoring of disease progression by specialist health professionals with expertise in ILD (this may be someone who runs a service seeing at least 500 ILD patients per year or has completed specialist training in ILD for at least 6 months). Experiences of availability and components of pulmonary rehabilitation Meeting other people with IPF and advice of support groups Coherent and concise information booklets and involvement of relatives when diagnosis is given (as patient does not often take in information at time) Warning not to access internet information immediately as it can be misleading.
Other considerations	The GDG emphasised the importance of continued support, continuity of care, alongside appropriate information and management of expectations for patients and carers, and considered that regular review and follow up could provide a mechanism for these aspects of care. Regular review allows patients to feel they are in contact with health services. Having a named member of the specialist team, i.e. a specialist ILD nurse, whom the patient could contact on the telephone for this support and information was considered to be a more appropriate use of resource than direct self-referral for a specialist appointment and could potentially allow a means of identifying particular patients where increased frequency of follow up was appropriate due to an unexpected decline. The GDG considered other relevant NICE guidance when making recommendations for patient review and follow-up, such as: The patient experience in adult NHS services (NICE clinical guideline 138) Smoking cessation services (NICE public health guidance 10) Managing dyspepsia in primary care (NICE clinical guideline17) (see recommendation 1.7.4) Lung cancer (NICE clinical guideline 121) Guidance in these areas was agreed to further emphasise good communication between health professionals and people with IPF, as well as alerting health professionals to the importance of treating co-morbidities, as well as providing smoking cessation advice where required.

15 Reference list

1 International guidelines for the selection of lung transplant candidates. The American Society for Transplant Physicians (ASTP)/American Thoracic Society(ATS)/European Respiratory Society(ERS)/International Society for Heart and Lung Transplantation(ISHLT). American Journal of Respiratory and Critical Care Medicine. 1998; 158(1):335-339. *(Guideline Ref ID LALAATSP1998)*

2 6MWD and Risk of Mortality in IPF. 2012 *(Guideline Ref ID DUBOIS2012A)*

3 MIMS. 2013. Available from: http://www.mims.co.uk/home/ [Last accessed: 13 March 2013] *(Guideline Ref ID MIMS2013)*

4 Aalokken TM, Naalsund A, Mynarek G, Berstad AE, Solberg S, Strom EH et al. Diagnostic accuracy of computed tomography and histopathology in the diagnosis of usual interstitial pneumonia. Acta Radiologica. 2012; 53(3):296-302. *(Guideline Ref ID AALOKKEN2012)*

5 Al Moamary MS. Impact of a pulmonary rehabilitation programme on respiratory parameters and health care utilization in patients with chronic lung diseases other than COPD. Eastern Mediterranean Health Journal. 2012; 18(2):120-126. *(Guideline Ref ID ALMOAMARY2012)*

6 Al-Hameed FM, Sharma S. Outcome of patients admitted to the intensive care unit for acute exacerbation of idiopathic pulmonary fibrosis. Canadian Respiratory Journal. 2004; 11(2):117-122. *(Guideline Ref ID ALHAMEED2004)*

7 Association for Palliative Medicine of Great Britain and Ireland, British Pain Society. The use of drugs beyond licence in palliative care and pain management. 2005. Available from: http://www.britishpainsociety.org/book_usingdrugs_main.pdf [Last accessed: 14 June 2012] *(Guideline Ref ID BPS2005)*

8 Best AC, Meng J, Lynch AM, Bozic CM, Miller D, Grunwald GK et al. Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. Radiology. 2008; 246(3):935-940. *(Guideline Ref ID BEST2008)*

9 Blivet S, Philit F, Sab JM, Langevin B, Paret M, Guerin C et al. Outcome of patients with idiopathic pulmonary fibrosis admitted to the ICU for respiratory failure. Chest. 2001; 120(1):209-212. *(Guideline Ref ID BLIVET2001)*

10 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 Suppl 5:v1-58. *(Guideline Ref ID BTS2008)*

11 Braghiroli A, Ioli F, Spada EL, Vecchio C, Donner CF. LTOT in pulmonary fibrosis. Monaldi Archives for Chest Disease. 1993; 48(5):437-440. *(Guideline Ref ID BRAGHIROLI1993)*

12 British Thoracic Society. Clinical component for the home oxygen service in England and Wales, 2006 Available from: http://www.brit-thoracic.org.uk/Portals/0/Clinical%20Information/Home%20Oxygen%20Service/clinical%20adultoxygenjan06.pdf *(Guideline Ref ID BTS2006)*

13 Caminati A, Bianchi A, Cassandro R, Mirenda MR, Harari S. Walking distance on 6-MWT is a prognostic factor in idiopathic pulmonary fibrosis. Respiratory Medicine. 2009; 103(1):117-123. *(Guideline Ref ID CAMINATI2009)*

14 Charman SC, Sharples LD, McNeil KD, Wallwork J. Assessment of survival benefit after lung transplantation by patient diagnosis. Journal of Heart and Lung Transplantation. 2002; 21(2):226-232. *(Guideline Ref ID CHARMAN2002)*

15 Chen H, Shiboski SC, Golden JA, Gould MK, Hays SR, Hoopes CW et al. Impact of the lung allocation score on lung transplantation for pulmonary arterial hypertension. American Journal of Respiratory and Critical Care Medicine. 2009; 180(5):468-474. *(Guideline Ref ID CHEN2009)*

16 Collard HR, Anstrom KJ, Schwarz MI, Zisman DA. Sildenafil improves walk distance in idiopathic pulmonary fibrosis. Chest. 2007; 131(3):897-899. *(Guideline Ref ID COLLARD2007)*

17 Collard HR, King TE, Jr., Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. American Journal of Respiratory & Critical Care Medicine. 2003; 168(5):538-542. *(Guideline Ref ID COLLARD2003)*

18 Coutinho GF, Pancas R, Magalhaes E, Bernardo JE, Eugenio L, Antunes MJ. Diagnostic value of surgical lung biopsy: comparison with clinical and radiological diagnosis. European Journal of Cardio-Thoracic Surgery. 2008; 33(5):781-785. *(Guideline Ref ID COUTINHO2008)*

19 Crockett AJ, Cranston JM, Antic N. Domiciliary oxygen for interstitial lung disease. Cochrane Database of Systematic Reviews. 2001; Issue 3:CD002883. *(Guideline Ref ID CROCKETT2001)*

20 Currow DC, McDonald C, Oaten S, Kenny B, Allcroft P, Frith P et al. Once-daily opioids for chronic dyspnea: a dose increment and pharmacovigilance study. Journal of Pain and Symptom Management. 2011; 42(3):388-399. *(Guideline Ref ID CURROW2011)*

21 Curtis L. Unit costs of health and social care. Canterbury: Personal Social Services Reseach Unit, University of Kent; 2011. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2011/index.php *(Guideline Ref ID PSSRU2011)*

22 De Oliveira NC, Osaki S, Maloney J, Cornwell RD, Meyer KC. Lung transplant for interstitial lung disease: outcomes for single versus bilateral lung transplantation. Interactive Cardiovascular and Thoracic Surgery. 2012; 14(3):263-267. *(Guideline Ref ID DEOLIVEIRA2012)*

23 Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. New England Journal of Medicine. 2005; 353(21):2229-2242. *(Guideline Ref ID DEMEDTS2005)*

24 Department of Health. NHS reference costs 2009-2010: Appendix NSRC04: NHS trust and PCT combined reference cost schedules, 2011 Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123459 *(Guideline Ref ID NHSNSRC2011)*

25 Department of Health. NHS reference costs 2010-11. 2012. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_131140 [Last accessed: 27 March 2012] *(Guideline Ref ID DOH2012)*

26 Du Bois R, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. American Journal of Respiratory and Critical Care Medicine. 2011; 184(4):459-466. *(Guideline Ref ID DUBOIS2011A)*

27 du Bois RM, Albera C, Bradford WZ, Costabel U, Noble PW, Sahn SA et al. A Novel Clinical Prediction Model For Near-Term Mortality In Patients With Idiopathic Pulmonary Fibrosis (IPF). American Journal of Respiratory & Critical Care Medicine. American Thoracic Society International Conference Abstracts: American Thoracic Society. 2013; 187:A2357. *(Guideline Ref ID DUBOIS2013)*

28 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. American Journal of Respiratory & Critical Care Medicine. 2011; 184(12):1382-1389. *(Guideline Ref ID DUBOIS2011B)*

29 du Bois RM. An earlier and more confident diagnosis of idiopathic pulmonary fibrosis. European Respiratory Review. 2012; 21(124):141-146. *(Guideline Ref ID DUBOIS2012)*

30 Ferreira A, Garvey C, Connors GL, Hilling L, Rigler J, Farrell S et al. Pulmonary rehabilitation in interstitial lung disease: benefits and predictors of response. Chest. 2009; 135(2):442-447. *(Guideline Ref ID FERREIRA2009)*

31 Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. American Journal of Respiratory and Critical Care Medicine. 2003; 168(5):543-548. *(Guideline Ref ID FLAHERTY2003)*

32 Flaherty KR, Thwaite EL, Kazerooni EA, Gross BH, Toews GB, Colby TV et al. Radiological versus histological diagnosis in UIP and NSIP: survival implications. Thorax. 2003; 58(2):143-148. *(Guideline Ref ID FLAHERTY2003A)*

33 Flaherty KR, Toews GB, Travis WD, Colby TV, Kazerooni EA, Gross BH et al. Clinical significance of histological classification of idiopathic interstitial pneumonia. European Respiratory Journal. 2002; 19(2):275-283. *(Guideline Ref ID FLAHERTY2002)*

34 Flaherty KR, Andrei AC, King TEJ, Raghu G, Colby TV, Wells A et al. Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? American Journal of Respiratory and Critical Care Medicine. 2007; 175(10):1054-1060. *(Guideline Ref ID FLAHERTY2007)*

35 Fumeaux T, Rothmeier C, Jolliet P. Outcome of mechanical ventilation for acute respiratory failure in patients with pulmonary fibrosis. Intensive Care Medicine. 2001; 27(12):1868-1874. *(Guideline Ref ID FUMEAUX2001)*

36 Gaunaurd I, Eustis N, Cohen M, Ramos C, Sol C, Cardenas D et al. Rehabilitation of patients with idiopathic pulmonary fibrosis: changes in quality of life, functional mobility, and oxygen metabolism. Cardiopulmonary Physical Therapy Journal. 2011; 22(4):30-31. *(Guideline Ref ID GAUNAURD2011)*

37 GRADE Working Group. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group website. 2011. Available from: http://www.gradeworkinggroup.org/ [Last accessed: 1 October 2011] *(Guideline Ref ID GRADE2011)*

38 Hagaman JT, Kinder BW, Eckman MH. Thiopurine S-methyltranferase testing in idiopathic pulmonary fibrosis: a pharmacogenetic cost-effectiveness analysis. Lung. United States 2010; 188(2):125-132. *(Guideline Ref ID HAGAMAN2010)*

39 Hallstrand TS, Boitano LJ, Johnson WC, Spada CA, Hayes JG, Raghu G. The timed walk test as a measure of severity and survival in idiopathic pulmonary fibrosis. European Respiratory Journal. 2005; 25(1):96-103. *(Guideline Ref ID HALLSTRAND2005)*

40 Hamada K, Nagai S, Tanaka S, Handa T, Shigematsu M, Nagao T et al. Significance of pulmonary arterial pressure and diffusion capacity of the lung as prognosticator in patients with idiopathic pulmonary fibrosis. Chest. 2007; 131(3):650-656. *(Guideline Ref ID HAMADA2007)*

41 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(6):549-554. *(Guideline Ref ID HOLLAND2008A)*

42 Holland A, Hill C. Physical training for interstitial lung disease. Cochrane Database of Systematic Reviews. 2008; Issue 4:CD006322. *(Guideline Ref ID HOLLAND2008)*

43 Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. Respiratory Medicine. 2009; 103(10):1430-1435. *(Guideline Ref ID HOLLAND2009)*

44 Holland AE, Hill CJ, Glaspole I, Goh N, McDonald CF. Predictors of benefit following pulmonary rehabilitation for interstitial lung disease. Respiratory Medicine. 2012; 106(3):429-435. *(Guideline Ref ID HOLLAND2012)*

45 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(3):467-477. *(Guideline Ref ID HOMMA2012)*

46 Hope-Gill BDM, Hilldrup S, Davies C, Newton RP, Harrison NK. A study of the cough reflex in idiopathic pulmonary fibrosis. American Journal of Respiratory and Critical Care Medicine. 2003; 168(8):995-1002. *(Guideline Ref ID HOPEGILL2003)*

47 Horton MR, Danoff SK, Lechtzin N. Thalidomide inhibits the intractable cough of idiopathic pulmonary fibrosis. Thorax. 2008; 63(8):749. *(Guideline Ref ID HORTON2008)*

48 Horton MR, Santopietro V, Mathew L, Horton KM, Polito AJ, Liu MC et al. Thalidomide for the treatment of cough in idiopathic pulmonary fibrosis: a randomized trial. Annals of Internal Medicine. 2012; 157(6):398-406. *(Guideline Ref ID HORTON2012)*

49 Hunninghake GW, Zimmerman MB, Schwartz DA, King TE, Jr., Lynch J, Hegele R et al. Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. American Journal of Respiratory & Critical Care Medicine. 2001; 164(2):193-196. *(Guideline Ref ID HUNNINGHAKE2001)*

50 Idiopathic Pulmonary Fibrosis Clinical Research Network, Raghu G, Anstrom KJ, King TEJ, Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. New England Journal of Medicine. 2012; 366(21):1968-1977. *(Guideline Ref ID PANTHER2012)*

51 Ishie RT, Cardoso JJdD, Silveira RJ, Stocco L. Video-assisted thoracoscopy for the diagnosis of diffuse parenchymal lung disease. Jornal Brasileiro De Pneumologia. 2009; 35(3):234-241. *(Guideline Ref ID ISHIE2009)*

52 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(2):115-123. *(Guideline Ref ID JACKSON2010)*

53 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clinical Trials. 1989; 10(4):407-415. *(Guideline Ref ID JAESCHKE1989)*

54 Jamaati HR, Tabarsi P, Emami M, Mohammadi F, Bakhshayesh KM, Masjedi MR. Clinical pattern of idiopathic pulmonary fibrosis: a retrospective study. Tanaffos. 2006; 5(2):27-32. *(Guideline Ref ID JAMAATI2006)*

55 Jastrzebski D, Gumola A, Gawlik R, Kozielski J. Dyspnea and quality of life in patients with pulmonary fibrosis after six weeks of respiratory rehabilitation. Journal of Physiology and Pharmacology. 2006; 57 Suppl 4:139-148. *(Guideline Ref ID JASTRZEBSKI2006)*

56 Jegal Y, Kim DS, Shim TS, Lim CM, Do Lee S, Koh Y et al. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. American Journal of Respiratory and Critical Care Medicine. 2005; 171(6):639-644. *(Guideline Ref ID JEGAL2005)*

57 Jeon K, Chung MP, Lee KS, Chung MJ, Han J, Koh W-J et al. Prognostic factors and causes of death in Korean patients with idiopathic pulmonary fibrosis. Respiratory Medicine. 2006; 100(3):451-457. *(Guideline Ref ID JEON2006)*

58 Kadikar A, Maurer J, Kesten S. The six-minute walk test: a guide to assessment for lung transplantation. Journal of Heart and Lung Transplantation. 1997; 16(3):313-319. *(Guideline Ref ID KADIKAR1997)*

59 Katzenstein AL, Fiorelli RF. Nonspecific interstitial pneumonia/fibrosis. Histologic features and clinical significance. American Journal of Surgical Pathology. 1994; 18(2):136-147. *(Guideline Ref ID KATZENSTEIN1994)*

60 Katzenstein AL, Myers JL. Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. American Journal of Respiratory and Critical Care Medicine. 1998; 157(4 Pt 1):1301-1315. *(Guideline Ref ID KATZENSTEIN1998)*

61 King TE, Jr. Bosentan for idiopathic pulmonary fibrosis. Current Opinion in Investigational Drugs. 2008; 9(11):1171-1179. *(Guideline Ref ID KING2008)*

62 King TE, Jr., 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. American Journal of Respiratory & Critical Care Medicine. 2008; 177(1):75-81. *(Guideline Ref ID KING2008A)*

63 King TE, Jr., 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. American Journal of Respiratory & Critical Care Medicine. 2011; 184(1):92-99. *(Guideline Ref ID KING2011)*

64 King TE, Jr., 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(1):171-177. *(Guideline Ref ID KING2005)*

65 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; International Review of Thoracic Diseases. 2011; 81(3):196-205. *(Guideline Ref ID KOZU2011)*

66 Kubo H, Nakayama K, Yanai M, Suzuki T, Yamaya M, Watanabe M et al. Anticoagulant therapy for idiopathic pulmonary fibrosis. Chest. 2005; 128(3):1475-1482. *(Guideline Ref ID KUBO2005)*

67 Kurashima K, Takayanagi N, Tsuchiya N, Kanauchi T, Ueda M, Hoshi T et al. The effect of emphysema on lung function and survival in patients with idiopathic pulmonary fibrosis. Respirology. 2010; 15(5):843-848. *(Guideline Ref ID KURASHIMA2010)*

68 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. American Journal of Respiratory and Critical Care Medicine. 2003; 168(5):531-537. *(Guideline Ref ID LATSI2003)*

69 Lettieri CJ, Veerappan GR, Helman DL, Mulligan CR, Shorr AF. Outcomes and safety of surgical lung biopsy for interstitial lung disease. Chest. 2005; 127(5):1600-1605. *(Guideline Ref ID LETTIERI2005A)*

70 Lettieri CJ, Veerappan GR, Parker JM, Franks TJ, Hayden D, Travis WD et al. Discordance between general and pulmonary pathologists in the diagnosis of interstitial lung disease. Respiratory Medicine. 2005; 99(11):1425-1430. *(Guideline Ref ID LETTIERI2005)*

71 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 and Lung. 2010; 39(4):304-313. *(Guideline Ref ID LINDELL2010)*

72 Lynch DA, Godwin JD, Safrin S, Starko KM, Hormel P, Brown KK et al. High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. American Journal of Respiratory and Critical Care Medicine. 2005; 172(4):488-493. *(Guideline Ref ID LYNCH2005)*

73 Manali ED, Stathopoulos GT, Kollintza A, Kalomenidis I, Emili JM, Sotiropoulou C et al. The Medical Research Council chronic dyspnea score predicts the survival of patients with idiopathic pulmonary fibrosis. Respiratory Medicine. 2008; 102(4):586-592. *(Guideline Ref ID MANALI2008)*

74 Mejia M, Carrillo G, Rojas-Serrano J, Estrada A, Suarez T, Alonso D et al. Idiopathic pulmonary fibrosis and emphysema: decreased survival associated with severe pulmonary arterial hypertension. Chest. 2009; 136(1):10-15. *(Guideline Ref ID MEJIA2009)*

75 Mogulkoc N, Brutsche MH, Bishop PW, Greaves SM, Horrocks AW, Egan JJ et al. Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. American Journal of Respiratory and Critical Care Medicine. 2001; 164(1):103-108. *(Guideline Ref ID MOGULKOC2001A)*

76 Molin LJ, Steinberg JB, Lanza LA. VATS increases costs in patients undergoing lung biopsy for interstitial lung disease. Annals of Thoracic Surgery. United States 1994; 58(6):1595-1598. *(Guideline Ref ID MOLIN1994)*

77 Mollica C, Paone G, Conti V, Ceccarelli D, Schmid G, Mattia P et al. Mechanical ventilation in patients with end-stage idiopathic pulmonary fibrosis. Respiration; International Review of Thoracic Diseases. 2010; 79(3):209-215. *(Guideline Ref ID MOLLICA2010)*

78 Monaghan H, Wells AU, Colby TV, du Bois RM, Hansell DM, Nicholson AG. Prognostic implications of histologic patterns in multiple surgical lung biopsies from patients with idiopathic interstitial pneumonias. Chest. 2004; 125(2):522-526. *(Guideline Ref ID MONAGHAN2004)*

79 Mura M, Porretta MA, Bargagli E, Sergiacomi G, Zompatori M, Sverzellati N et al. Predicting survival in newly diagnosed idiopathic pulmonary fibrosis: a 3-year prospective study. European Respiratory Journal. 2012; 40(1):101-109. *(Guideline Ref ID MURA2012)*

80 Naji NA, Connor MC, Donnelly SC, McDonnell TJ. Effectiveness of pulmonary rehabilitation in restrictive lung disease. Journal of Cardiopulmonary Rehabilitation. 2006; 26(4):237-243. *(Guideline Ref ID NAJI2006)*

81 National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisals. 2008. [Last accessed: 19 December 2008] *(Guideline Ref ID NICE2008)*

82 National Institute for Health and Clinical Excellence. Equality and diversity policy. London: National Institute for Health and Clinical Excellence; 2009. Available from: http://www.nice.org.uk/media/0B6/F7/EqualityDiversityPolicy.pdf *(Guideline Ref ID NICE2009B)*

83 National Institute for Health and Clinical Excellence. The guidelines manual 2009. 2009. Available from: http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/GuidelinesManual2009.jsp [Last accessed: 13 January 2009] *(Guideline Ref ID NICE2009)*

84 National Institute for Health and Clinical Excellence. The guidelines manual 2012. London. National Institute for Health and Clinical Excellence, 2012 Available from: http://publications.nice.org.uk/the-guidelines-manual-pmg6 *(Guideline Ref ID NICE2012)*

85 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(6):462-467. *(Guideline Ref ID NAVARATNAM2011)*

86 Neri M, Fedi L, Spanevello A, Mazzucchelli G, Grandi M, Ambrosetti M et al. Savings obtained using an oxygen economizer device: a cost-minimization analysis. Monaldi Archives for Chest Disease. Italy 1999; 54(4):311-314. *(Guideline Ref ID NERI1999)*

87 NHS Business Services Authority. NHS electronic drug tariff. 2013. Available from: http://www.ppa.org.uk/edt/March_2013/mindex.htm *(Guideline Ref ID NHS2013)*

88 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(3):394-399. *(Guideline Ref ID NISHIYAMA2008)*

89 Noth I, Anstrom KJ, Calvert SB, de Andrade JA, Flaherty KR, Glazer C et al. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. American Journal of Respiratory and Critical Care Medicine. 2012; 186(1):88-95. *(Guideline Ref ID NOTH2012)*

90 Obi I, Shlobin OA, Weir N, Ahmad S, Barnett S, Nathan SD. Effects of oxygen therapy in patients with advanced lung diseases. American Journal of Respiratory and Critical Care Medicine. 2010; 181(1 MeetingAbstracts). *(Guideline Ref ID OBI2010)*

91 Ohshimo S, Bonella F, Cui A, Beume M, Kohno N, Guzman J et al. Significance of bronchoalveolar lavage for the diagnosis of idiopathic pulmonary fibrosis. American Journal of Respiratory and Critical Care Medicine. 2009; 179(11):1043-1047. *(Guideline Ref ID OHSHIMO2009)*

92 Oliveira CC, Fabro AT, Ribeiro SM, Defaveri J, Capelozzi VL, Queluz THT et al. Evaluation of the use of transbronchial biopsy in patients with clinical suspicion of interstitial lung disease. Jornal Brasileiro De Pneumologia. 2011; 37(2):168-175. *(Guideline Ref ID OLIVEIRA2011)*

93 Ooi A, Iyenger S, Ferguson J, Ritchie AJ. VATS lung biopsy in suspected, diffuse interstitial lung disease provides diagnosis, and alters management strategies. Heart, Lung and Circulation. 2005; 14(2):90-92. *(Guideline Ref ID OOI2005)*

94 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. Journal of Heart and Lung Transplantation. 2006; 25(7):745-755. *(Guideline Ref ID ORENS2006)*

95 Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2011. Available from: http://www.oecd.org/std/ppp [Last accessed: 16 January 2011] *(Guideline Ref ID OECD2011)*

96 Ozalevli S, Karaali HK, Ilgin D, Ucan ES. Effect of home-based pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. Multidisciplinary Respiratory Medicine. 2010; 5(1):31-37. *(Guideline Ref ID OZALEVLI2010)*

97 Paik HC, Haam SJ, Lee DY, Yi GJ, Song SW, Kim YT et al. The fate of patients on the waiting list for lung transplantation in Korea. Transplantation Proceedings. 2012; 44(4):865-869. *(Guideline Ref ID PAIK2012)*

98 Peckham RM, Shorr AF, Helman DLJ. Potential limitations of clinical criteria for the diagnosis of idiopathic pulmonary fibrosis/cryptogenic fibrosing alveolitis. Respiration; International Review of Thoracic Diseases. 2004; 71(2):165-169. *(Guideline Ref ID PECKHAM2004)*

99 Personal Social Services Research Unit. Unit Costs of Health and Social Care 2010. University of Kent: 2010 Available from: http://www.pssru.ac.uk/uc/uc2010contents.htm *(Guideline Ref ID PSSRU2010)*

100 Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health and Quality of Life Outcomes. 2007; 5:70. *(Guideline Ref ID PICKARD2007)*

101 Raghu G, Behr J, Brown KK, Egan JJ, Kawut SM, Flaherty KR et al. Artemis-Ipf: a placebo-controlled trial of ambrisentan in idiopathic pulmonary fibrosis. American Journal of Respiratory & Critical Care Medicine. 2012; 185(Suppl):A3632. *(Guideline Ref ID RAGHU2012)*

102 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. American Review of Respiratory Disease. 1991; 144(2):291-296. *(Guideline Ref ID RAGHU1991)*

103 Raghu G, Mageto YN, Lockhart D, Schmidt RA, Wood DE, Godwin JD. The accuracy of the clinical diagnosis of new-onset idiopathic pulmonary fibrosis and other interstitial lung disease: A prospective study. Chest. 1999; 116(5):1168-1174. *(Guideline Ref ID RAGHU1999)*

104 Rammaert B, Leroy S, Cavestri B, Wallaert B, Grosbois JM. Home-based pulmonary rehabilitation in idiopathic pulmonary fibrosis. Revue Des Maladies Respiratoires. 2011; 28(7):e52-e57. *(Guideline Ref ID RAMMAERT2011)*

105 Ramsey SD, Patrick DL, Albert RK, Larson EB, Wood DE, Raghu G. The cost-effectiveness of lung transplantation: a pilot study. Chest. 1995; 108(6):1594-1601. *(Guideline Ref ID RAMSEY1995)*

106 Rena O, Casadio C, Leo F, Giobbe R, Cianci R, Baldi S et al. Videothoracoscopic lung biopsy in the diagnosis of interstitial lung disease. European Journal of Cardio-Thoracic Surgery. 1999; 16(6):624-627. *(Guideline Ref ID RENA1999)*

107 Richeldi L, Davies Huw Richard HR, Spagnolo P, Luppi F. Corticosteroids for idiopathic pulmonary fibrosis. Cochrane Database of Systematic Reviews. 2003; Issue 3:CD002880. DOI:10.1002/14651858.CD002880. *(Guideline Ref ID RICHELDI2003)*

108 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(5):407-411. *(Guideline Ref ID RICHELDI2012A)*

109 Saini G, McKeever T, Johnson S, Jenkins G. Thalidomide as treatment for IPF associated cough. Thorax. 2011; 66(Suppl 4):A103. *(Guideline Ref ID SAINI2011)*

110 Saydain G, Islam A, Afessa B, Ryu JH, Scott JP, Peters SG. Outcome of patients with idiopathic pulmonary fibrosis admitted to the intensive care unit. American Journal of Respiratory and Critical Care Medicine. 2002; 166(6):839-842. *(Guideline Ref ID SAYDAIN2002)*

111 Schmidt SL, Nambiar AM, Tayob N, Sundaram B, Han MK, Gross BH et al. Pulmonary function measures predict mortality differently in IPF versus combined pulmonary fibrosis and emphysema. European Respiratory Journal. 2011; 38(1):176-183. *(Guideline Ref ID SCHMIDT2011)*

112 Schoenheit G, Becattelli I, Cohen AH. Living with idiopathic pulmonary fibrosis: an in-depth qualitative survey of European patients. Chronic Respiratory Disease. 2011; 8(4):225-231. *(Guideline Ref ID SCHOENHEIT2011)*

113 Schunemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? Health Services Research. 2005; 40(2):593-597. *(Guideline Ref ID SCHUNEMANN2005)*

114 Schunemann HJ, Puhan M, Goldstein R, Jaeschke R, Guyatt GH. Measurement properties and interpretability of the Chronic respiratory disease questionnaire (CRQ). COPD. 2005; 2(1):81-89. *(Guideline Ref ID SCHUNEMANN2005A)*

115 Shulgina L, Cahn AP, Chilvers ER, Parfrey H, Clark AB, Wilson ECF et al. Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: a randomised controlled trial. Thorax. 2013; 68(2):155-162. *(Guideline Ref ID SHULGINA2013)*

116 Sigurdsson MI, Isaksson HJ, Gudmundsson G, Gudbjartsson T. Diagnostic surgical lung biopsies for suspected interstitial lung diseases: a retrospective study. Annals of Thoracic Surgery. 2009; 88(1):227-232. *(Guideline Ref ID SIGURDSSON2009A)*

117 Slodkowska J, Wesolowski S, Onish K, Kus J. Usual interstitial pneumonia: A comparative analysis of histopathology and high resolution computed tomography (HRCT) in relation to the clinical course of the disease. Electronic Journal of Pathology and Histology. 2000; 6(4):88-99. *(Guideline Ref ID SLODKOWSKA2000)*

118 Spagnolo P, Del GC, Luppi F, Cerri S, Balduzzi S, Walters E Haydn et al. Non-steroid agents for idiopathic pulmonary fibrosis. Cochrane Database of Systematic Reviews. 2010; Issue 9:CD003134. *(Guideline Ref ID SPAGNOLO2010)*

119 Spencer L, Grundy S, Greaves M, Bishop P, Duck A, Leonard C. Demonstration of diagnostic and prognostic benefits of an interstitial lung disease (ILD) multidisciplinary team meeting. European Respiratory Society Congress 2011 2011. *(Guideline Ref ID SPENCER2011)*

120 Stern JB, Mal H, Groussard O, Brugiere O, Marceau A, Jebrak G et al. Prognosis of patients with advanced idiopathic pulmonary fibrosis requiring mechanical ventilation for acute respiratory failure. Chest. 2001; 120(1):213-219. *(Guideline Ref ID STERN2001)*

121 Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. American Journal of Respiratory and Critical Care Medicine. 2008; 177(4):433-439. *(Guideline Ref ID SUMIKAWA2008)*

122 Sverzellati N, Wells AU, Tomassetti S, Desai SR, Copley SJ, Aziz ZA et al. Biopsy-proved idiopathic pulmonary fibrosis: spectrum of nondiagnostic thin-section CT diagnoses. Radiology. 2010; 254(3):957-964. *(Guideline Ref ID SVERZELLATI2010)*

123 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. Respiratory Medicine. 2010; 104(2):296-304. *(Guideline Ref ID SWIGRIS2010)*

124 Swigris JJ, Fairclough DL, Morrison M, Make B, Kozora E, Brown KK et al. Benefits of pulmonary rehabilitation in idiopathic pulmonary fibrosis. Respiratory Care. 2011; 56(6):783-789. *(Guideline Ref ID SWIGRIS2011)*

125 Swinburn CR, Mould H, Stone TN, Corris PA, Gibson GJ. Symptomatic benefit of supplemental oxygen in hypoxemic patients with chronic lung disease. American Review of Respiratory Disease. 1991; 143(5 Pt 1):913-915. *(Guideline Ref ID SWINBURN1991)*

126 Thomeer M, Verschakelen J, Laurent F, Nicholson AG, Verbeken EK, Capron F et al. Multidisciplinary interobserver agreement in the diagnosis of idiopathic pulmonary fibrosis. European Respiratory Journal. 2008; 31(3):585-591. *(Guideline Ref ID THOMEER2008)*

127 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(4):449-455. *(Guideline Ref ID TOMIOKA2005)*

128 Trahan S, Hanak V, Ryu JH, Myers JL. Role of surgical lung biopsy in separating chronic hypersensitivity pneumonia from usual interstitial pneumonia/idiopathic pulmonary fibrosis. Chest. 2008; 134(1):126-132. *(Guideline Ref ID TRAHAN2008A)*

129 Vansteenkiste J, Verbeken E, Thomeer M, Van Haecke P, Eeckhout AV, Demedts M. Medical thoracoscopic lung biopsy in interstitial lung disease: a prospective study of biopsy quality. European Respiratory Journal. 1999; 14(3):585-590. *(Guideline Ref ID VANSTEENKISTE1999)*

130 Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. Thorax. 2012. *(Guideline Ref ID WELLS2012)*

131 Yamaguchi M, Yoshino I, Suemitsu R, Osoegawa A, Kameyama T, Tagawa T et al. Elective video-assisted thoracoscopic lung biopsy for interstitial lung disease. Asian Cardiovascular and Thoracic Annals. 2004; 12(1):65-68. *(Guideline Ref ID YAMAGUCHI2004)*

132 Yokoyama T, Kondoh Y, Taniguchi H, Kataoka K, Kato K, Nishiyama O et al. Noninvasive ventilation in acute exacerbation of idiopathic pulmonary fibrosis. Internal Medicine. 2010; 49(15):1509-1514. *(Guideline Ref ID YOKOYAMA2010)*

133 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. European Respiratory Journal : Official Journal of the European Society for Clinical Respiratory Physiology. 2010; 35(4):830-836. *(Guideline Ref ID ZAPPALA2010)*

134 Zielinski J. Long-term oxygen therapy in conditions other than chronic obstructive pulmonary disease. State-of-the-art conference on long-term oxygen therapy, part II. Respiratory Care. 2000; 45(2):172-177. *(Guideline Ref ID ZIELINSKI2000)*

135 Zisman DA, Schwarz M, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. New England Journal of Medicine. 2010; 363(7):620-628. *(Guideline Ref ID ZISMAN2010)*

16 Acronyms and abbreviations

ACA	Available case analysis
ATS	American Thoracic Society
CCA	Cost-consequences analysis
CEA	Cost-effectiveness analysis
CFA	Cryptogenic fibrosis alveolitis
CI	Confidence interval
CRP	Clinical, radiologic, physiological score
CT	Computed tomography
CUA	Cost-utility analysis
BAL	Bronchoalveolar lavage
DLCO	Carbon monoxide diffusing capacity
ERS	European Respiratory Society
EQ-5D	Euro quality of life – 5D
EAA	Extrinsic allergic alveolitis
FN	False negative
FP	False positive
FVC	Forced vital capacity
GORD	Gastro-oesophageal reflux disease
HAD	Hospital anxiety and depression
HR	Hazard ratio
HRQoL	Health-related quality of life
ICER	Incremental cost effectiveness ratio
IIP	Idiopathic interstitial pneumonia
ILD	Interstitial lung disease
INB	Incremental net benefit
IPF	Idiopathic pulmonary fibrosis
ITT	Intention to treat analysis
kPa	kilopascal
LTX	Lung transplantation

M/F	Male/ female
MDT	Multidisciplinary team
MID	Minimally important difference
N	Total number of patients
NNT	Numbers needed to treat
NPV	Negative predictive value
NR	Not reported
NSIP	Non-specific interstitial pneumonia
OLB	Open lung biopsy
Pa02	Partial pressure of oxygen in arterial blood
PAP	Pulmonary arterial pressure
PFS	Progression free survival
PFTs	Pulmonary function tests
PPV	Positive predictive value
PR	Pulmonary rehabilitation
QoL	Quality of life
QALY	Quality adjusted life year
RBILD	Respiratory bronchiolitis associated interstitial lung disease
RCT	Randomised controlled trial
RR	Relative risk
6MWD	Six minute walk distance
6MWT	Six minute walk test
TBB/ TBBX	Transbronchial biopsy
TN	True negative
TP	True positive
TCC	Transthoracic doppler echocardiography
TLCO	Transfer factor of the lung for carbon monoxide
SD	Standard deviation
SLB	Surgical lung biopsy
UIP	Usual interstitial pneumonia
VA	Alveolar volume

VATLB	Video assisted thoracic lung biopsy
VATS	Video assisted thoracic surgery
VC	Vital capacity

17 Glossary

Term	Definition
Abstract	Summary of a study, which may be published alone or as an introduction to a full scientific paper.
Acute IPF exacerbation	Unexplained worsening of dyspnoea within one month, evidence of hypoxia as defined by worsened or severely impaired gas exchange, new radiographic alveolar infiltrates and an absence of an alternative explanation such as infection, pulmonary embolism, pneumothorax, or heart failure. The term should not be used to describe simply deterioration in symptoms.
Acute respiratory failure	Type 1: Respiratory failure consists of hypoxia with a normal level of carbon dioxide (PaO_2 <8.0 kPa with $PaCO_2$ <6.5 kPa). Type 2: Respiratory failure consists of hypoxia and ventilatory failure (PaO_2 <8.0 kPa with $PaCO_2$ >6.5 kPa).
Algorithm (in guidelines)	A flow chart of the clinical decision pathway described in the guideline, where decision points are represented with boxes, linked with arrows.
Allocation concealment	The process used to prevent advance knowledge of group assignment in a RCT. The allocation process should be impervious to any influence by the individual making the allocation, by being administered by someone who is not responsible for recruiting participants.
Applicability	The degree to which the results of an observation, study or review are likely to hold true in a particular clinical practice setting.
Arm (of a clinical study)	Sub-section of individuals within a study who receive one particular intervention, for example placebo arm.
Association	Statistical relationship between two or more events, characteristics or other variables. The relationship may or may not be causal.
ATS/ERS consensus criteria	American Thoracic Society/European Respiratory Society criteria for the diagnosis and management of IPF based on evidence base and expert consensus.
Baseline	The initial set of measurements at the beginning of a study (after run-in period where applicable), with which subsequent results are compared.
Before-and-after study	A study that investigates the effects of an intervention by measuring particular characteristics of a population both before and after taking the intervention, and assessing any change that occurs.
Bias	Systematic (as opposed to random) deviation of the results of a study from the 'true' results that is caused by the way the study is designed or conducted.
Biopsy	Removal and examination, usually microscopic, of tissue from the living body, performed to establish precise diagnosis
Blinding	Keeping the study participants, caregivers, researchers and outcome assessors unaware about the interventions to which the participants have been allocated in a study.
Breathlessness	See 'dyspnoea'
Bronchoalveolar lavage	Procedure in which the bronchoscope is wedged in a sub-segmental bronchus and fluid (usually saline) is introduced in aliquots up to 240mls and removed again by suction, in order to sample the alveolar environment for infection and cell make up.

Term	Definition
Carer (caregiver)	Someone other than a health professional who is involved in caring for a person with a medical condition.
Case-control study	Comparative observational study in which the investigator selects individuals who have experienced an event (For example, developed a disease) and others who have not (controls), and then collects data to determine previous exposure to a possible cause.
Case-series	Report of a number of cases of a given disease, usually covering the course of the disease and the response to treatment. There is no comparison (control) group of patients.
Change in TLCO or DLCO (change in gas transfer)	Absolute or percent predicted change from baseline in lung function measured as diffusing capacity for carbon monoxide.
Clinical effectiveness	The extent to which an intervention produces an overall health benefit in routine clinical practice.
Clinical efficacy	The extent to which an intervention is active when studied under controlled research conditions.
Clinical evaluation	A review of the patient's history, findings on clinical examination and a review of the clinical investigations with the aim of refining the clinical diagnosis or management plan.
Clinical features	Particular aspects apparent in the history, examination or clinical investigations, which influence diagnostic or management decisions.
Clinician	A healthcare professional providing direct patient care, for example doctor, nurse or physiotherapist.
Cohort study	A retrospective or prospective follow-up study. Groups of individuals to be followed up are defined on the basis of presence or absence of exposure to a suspected risk factor or intervention. A cohort study can be comparative, in which case two or more groups are selected on the basis of differences in their exposure to the agent of interest.
Complications of IPF	Can include: pneumonia, pulmonary embolism, pneumothorax, pulmonary hypertension, acute coronary syndrome and lung cancer. People with IPF (compared to people without IPF of the same age) are more likely to develop these conditions.
Comparability	Similarity of the groups in characteristics likely to affect the study results (such as health status or age).
Computed tomography (CT or CAT scan) See HRCT	A radiological technique to image the thorax including the lungs. CT scans (with and without contrast) have become central to the diagnostic process for people with interstitial lung disease.
Concordance	This is a recent term whose meaning has changed. It was initially applied to the consultation process in which doctor and patient agree therapeutic decisions that incorporate their respective views, but now includes patient support in medicine taking as well as prescribing communication. Concordance reflects social values but does not address medicine-taking and may not lead to improved adherence.
Confident diagnosis	When other potential diagnoses have been excluded leaving the clinical with the view that the patient has idiopathic pulmonary Fibrosis. In the case of IPF a confident diagnosis usually applies to the degree of certainty that the diagnosis is IPF based on clinical features, CT scan finding and the histological assessment of a surgical biopsy if performed.
Confidence interval (CI)	A range of values for an unknown population parameter with a stated 'confidence' (conventionally 95%) that it contains the true value. The

Term	Definition
	interval is calculated from sample data, and generally straddles the sample estimate. The 'confidence' value means that if the method used to calculate the interval is repeated many times, then that proportion of intervals will actually contain the true value.
Confirmed idiopathic pulmonary fibrosis	A confident diagnosis of IPF on the basis of integration of clinical features, CT scan appearances and if required histological assessment of a surgical biopsy, by a multidisciplinary team.
Confirmed idiopathic pulmonary fibrosis	A confident diagnosis of IPF on the basis of a CT scan, histological assessment of a surgical biopsy, or multidisciplinary team.
Confounding	In a study, confounding occurs when the effect of an intervention on an outcome is distorted as a result of an association between the population or intervention or outcome and another factor (the 'confounding variable') that can influence the outcome independently of the intervention under study.
Consensus methods	Techniques that aim to reach an agreement on a particular issue. Consensus methods may use when there is a lack of strong evidence on a particular topic.
Control group	A group of patients recruited into a study that receives no treatment, a treatment of known effect, or a placebo (dummy treatment) - in order to provide a comparison for a group receiving an experimental treatment, such as a new drug.
Cost benefit analysis	A type of economic evaluation where both costs and benefits of healthcare treatment are measured in the same monetary units. If benefits exceed costs, the evaluation would recommend providing the treatment.
Cost-consequences analysis (CCA)	A type of economic evaluation where various health outcomes are reported in addition to cost for each intervention, but there is no overall measure of health gain.
Cost-effectiveness analysis (CEA)	An economic study design in which consequences of different interventions are measured using a single outcome, usually in 'natural' units (For example, life-years gained, deaths avoided, heart attacks avoided, cases detected). Alternative interventions are then compared in terms of cost per unit of effectiveness.
Cost-effectiveness model	An explicit mathematical framework, which is used to represent clinical decision problems and incorporate evidence from a variety of sources in order to estimate the costs and health outcomes.
Cost-utility analysis (CUA)	A form of cost-effectiveness analysis in which the units of effectiveness are quality-adjusted life-years (QALYs).
Cough	A sudden and repetitive reflex with the aim of clearing the large airways
Credible Interval	The Bayesian equivalent of a confidence interval.
Cryptogenic fibrosis alveolitis	A syndrome that encompasses a group of distinct interstitial lung diseases of unknown cause that often present with clinical features which resemble IPF. These include diseases several idiopathic interstitial pneumonias (IIP), including fibrotic non-specific interstitial pneumonia (NSIP).
Decision analysis	An explicit quantitative approach to decision making under uncertainty, based on evidence from research. This evidence is translated into probabilities, and then into diagrams or decision trees which direct the clinician through a succession of possible scenarios, actions and outcomes.

Term	Definition
Depression	A mood disorder characterised by one or more of the following - depressed mood, reduced interest in activities that used to be enjoyed, sleep disturbance, loss of energy, difficulty in concentrating, difficulty in decision making and suicidal thoughts or intentions.
Differential diagnosis	List of potential diagnoses that a clinician believes a patient may have on the basis of initial history, examination and clinical investigations. The differential diagnosis list is usually presented in decreasing order of likelihood of being correct.
Diffuse parenchymal lung disease	Synonymous with the term ''interstitial lung disease''.
Direct patient care	Any physical aspects of the healthcare of a patient, including treatments, self-care, and administration of medication.
Discounting	Costs and perhaps benefits incurred today have a higher value than costs and benefits occurring in the future. Discounting health benefits reflects individual preference for benefits to be experienced in the present rather than the future. Discounting costs reflects individual preference for costs to be experienced in the future rather than the present.
Disease progression	Evidence that the disease has advanced. Usually recorded objectively in terms of worsening lung function and/or increased extent of fibrosis on a CT scan, but also recorded subjectively on the basis of progression of breathlessness and/or cough.
DLCO	Lung function measure of gas exchange.
Dominance	An intervention is said to be dominated if there is an alternative intervention that is both less costly and more effective.
Drop-out	A participant who withdraws from a trial before the end.
Dyspnoea	The symptom reported by patients of shortness of breath.
Echocardiography	An ultrasound scan of the heart designed to detect structural abnormalities and assess function/functional impairment.
Economic evaluation	Comparative analysis of alternative health strategies (interventions or programmes) in terms of both their costs and consequences.
Effect (as in effect measure, treatment effect, estimate of effect, effect size)	The observed association between interventions and outcomes or a statistic to summarise the strength of the observed association.
Effectiveness	See 'Clinical effectiveness'.
Efficacy	See 'Clinical efficacy'.
Epidemiological study	The study of a disease within a population, defining its incidence and prevalence and examining the roles of external influences (For example, infection, diet) and interventions.
EQ-5D (EuroQol-5D)	A standardised instrument used to measure a health outcome. It provides a single index value for health status.
Evidence	Information on which a decision or guidance is based. Evidence is obtained from a range of sources including randomised controlled trials, observational studies, expert opinion (of clinical professionals and/or patients).
Exclusion criteria (clinical study)	Criteria that define who is not eligible to participate in a clinical study.
Exclusion criteria (literature review)	Explicit standards used to decide which studies should be excluded from consideration as potential sources of evidence.
Extended dominance	If Option A is both more clinically effective than Option B and has a

Term	Definition
	lower cost per unit of effect, when both are compared with a do-nothing alternative then Option A is said to have extended dominance over Option B. Option A is therefore more efficient and should be preferred, other things remaining equal.
Extrapolation	In data analysis, predicting the value of a parameter outside the range of observed values.
Extrinsic allergic alveolitis (EAA)	(synonymous with the term ''hypersensitivity pneumonitis'')
Fatigue	The symptom of tiredness, lethargy or exhaustion.
Follow up	Observation over a period of time of an individual, group or initially defined population whose appropriate characteristics have been assessed in order to observe changes in health status or health related variables.
Forced vital capacity (FVC)	A lung function test measuring the total volume of air that a person can exhale in a forced manner from their lungs after taking a full inspiration.
GDG Consensus (see informal consensus methods)	GDG Consensus may be used when there is a lack of strong evidence on a particular topic to reach an agreement for a recommendation.
Generalisability	The extent to which the results of a study based on measurement in a particular patient population and/or a specific context hold true for another population and/or in a different context. In this instance, this is the degree to which the guideline recommendation is applicable across both geographical and contextual settings. For instance, guidelines that suggest substituting one form of labour for another should acknowledge that these costs might vary across the country.
GRADE / GRADE profile	A system developed by the GRADE Working Group to address the shortcomings of present grading systems in healthcare. The GRADE system uses a common, sensible and transparent approach to grading the quality of evidence. The results of applying the GRADE system to clinical trial data are displayed in a table known as a GRADE profile.
Harms	Adverse effects of an intervention.
Health economics	The study of the allocation of scarce resources among alternative healthcare treatments. Health economists are concerned with both increasing the average level of health in the population and improving the distribution of health.
Health-related quality of life (HRQoL)	A combination of an individual's physical, mental and social well-being; not merely the absence of disease.
Heterogeneity or lack of homogeneity	The term is used in meta-analyses and systematic reviews when the results or estimates of effects of treatment from separate studies seem to be very different – in terms of the size of treatment effects or even to the extent that some indicate beneficial and others suggest adverse treatment effects. Such results may occur as a result of differences between studies in terms of the patient populations, outcome measures, definition of variables or duration of follow-up.
High resolution computed tomography scanning (see CT scan)	A radiological scan of the thorax including the lung in which the parameters are set to maximise the spatial resolution of the lung. This typically involves reconstructing the images to reflect a thin (1-2mm) slice of the thorax. Intravenous contrast enhancement is not usually required. However, with the introduction of multi-detector CT (MDCT) in recent years, high resolution images may be reconstructed from a standard volumetric data set
Hospitalisations due to all causes	An assessment of all admissions to hospital for people with IPF –

Term	Definition
	regardless of the main diagnoses leading to admission. Usually assessed in terms of number of admissions, number of recorded diagnoses during the admission and number of days spent in hospital.
Hospitalisations due to IPF	Number of hospital admissions in which the underlying diagnosis leading to admission is IPF or acute exacerbation of IPF.
Hospitalisations due to IPF complications (including IPF exacerbations)	Hospital admissions in which the underlying diagnosis leading to admission is IPF; these include acute exacerbation of IPF or a complication/co-morbidity relating to IPF.
Hypoxaemia	A deficiency of oxygen in the arterial blood. In physiological terms this is often defined as less than 8.0 kPa (see acute respiratory failure)
Hypoxia	A deficiency of oxygen in tissues.
Idiopathic interstitial pneumonia (IIP)	A general term used to describe interstitial lung diseases of unknown aetiology. The term encompasses IPF, non-specific interstitial pneumonia (NSIP) and others.
Idiopathic pulmonary fibrosis (IPF)	A progressive scarring disease of the lungs of unknown cause associated with characteristic clinical, CT and histological features.
Imprecision	Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of effect.
Inclusion criteria (literature review)	Explicit criteria used to decide which studies should be considered as potential sources of evidence.
Incremental analysis	The analysis of additional costs and additional clinical outcomes with different interventions.
Incremental cost	The mean cost per patient associated with an intervention minus the mean cost per patient associated with a comparator intervention.
Incremental cost effectiveness ratio (ICER)	The difference in the mean costs in the population of interest divided by the differences in the mean outcomes in the population of interest for one treatment compared with another.
Incremental net benefit (INB)	The value (usually in monetary terms) of an intervention net of its cost compared with a comparator intervention. The INB can be calculated for a given cost-effectiveness (willingness to pay) threshold. If the threshold is £20,000 per QALY gained then the INB is calculated as: (£20,000 x QALYs gained) – Incremental cost.
Indirectness	The available evidence is different to the review question being addressed, in terms of PICO (population, intervention, comparison and outcome).
Consensus methods	Techniques that aim to reach an agreement on a particular issue. Consensus methods may be used when there is a lack of strong evidence on a particular topic.
Intention to treat analysis (ITT)	A strategy for analysing data from a randomised controlled trial. All participants are included in the arm to which they were allocated, whether or not they received (or completed) the intervention given to that arm. Intention-to-treat analysis prevents bias caused by the loss of participants, which may disrupt the baseline equivalence established by randomisation and which may reflect non-adherence to the protocol.
Interstitial lung disease (ILD)	Synonymous with 'diffuse parenchymal lung disease'. A term that encompasses a variety of lung diseases of known and unknown cause and characterised by varying degrees of inflammation and fibrosis of the lung tissue. IPF is amongst the commonest of the ILDs.

Term	Definition
Intervention	Healthcare action intended to benefit the patient, for example, drug treatment, surgical procedure, psychological therapy.
Invasive ventilation	The process of additional mechanical ventilation via an airway adjunct such as an endotracheal tube or laryngeal mask.
Kappa statistic	A statistical measure of inter-rater agreement that takes into account the agreement occurring by chance.
Length of stay	The total number of days a participant stays in hospital.
Licence	See 'Product licence'.
Life-years gained	Mean average years of life gained per person as a result of the intervention compared with an alternative intervention.
Likelihood ratio	The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease. The likelihood ratio of a positive test result (LR+) is sensitivity divided by 1- specificity.
Linear decline in disease progression	An objective assessment of disease progression in terms of a progressive decline in lung function measurements (usually FVC or TLCO).
Local practice	The characteristics of clinical care in a particular centre.
Long-term care	Residential care in a home that may include skilled nursing care and help with everyday activities. This includes nursing homes and residential homes.
Loss to follow-up	Also known as attrition. The loss of participants during the course of a study. Participants that are lost during the study are often called dropouts.
Total lung capacity	The volume to which the lungs can be expanded with the maximum inspiratory effort
Lung transplantation	Replacement of a diseased lung with a donor lung, which may be a single or double lung transplant depending on whether 1 or 2 lungs are required.
Markov model	A method for estimating long-term costs and effects for recurrent or chronic conditions, based on health states and the probability of transition between them within a given time period (cycle).
Meta-analysis	A statistical technique for combining (pooling) the results of a number of studies that address the same question and report on the same outcomes to produce a summary result. The aim is to derive more precise and clear information from a large data pool. It is generally more reliably likely to confirm or refute a hypothesis than the individual trials.
MCID (minimal clinical important difference)	See MID
MID (minimal important difference)	The smallest difference in score in the outcome of interest which patients perceive as beneficial and which would mandate, in the absence of troubling side effects and excessive cost, a change in the patient's management.
Multidisciplinary Team (MDT)	A description of the full spectrum of healthcare workers that come together, usually in the form of regular clinical meetings, to care for people with IPF. Practically the MDT usually includes respiratory physicians, specialist nurses, histopathologists, radiologists,

Term	Definition
	administrative support and members of the palliative care team.
Multidisciplinary Team (MDT) consensus	A decision relating to disease diagnosis or management made by the MDT after review of available clinical information.
Multivariate model	A statistical model for analysis of the relationship between two or more predictor (independent) variables and the outcome (dependent) variable.
Negative predictive value (NPV)	A summary statistic usually used to describe the function of a diagnostic test the negative predictive value is the proportion of people with a negative test that our correctly diagnosed.
Non-specific interstitial pneumonia (NSIP)	A type of idiopathic interstitial pneumonia (IIP) with characteristic histological appearances. NSIP is often fibrotic and may be progressive and is a differential diagnosis for people with IPF.
Number needed to treat (NNT)	The number of patients that who on average must be treated to prevent a single occurrence of the outcome of interest.
Observational study	Retrospective or prospective study in which the investigator observes the natural course of events with or without control groups; for example, cohort studies and case–control studies.
Open lung biopsy	A surgical biopsy of the lung involving a thoracotomy. Historically this was the main procedure to obtain surgical lung biopsies – but increasingly surgical biopsies are being obtained via video assisted surgical procedures, which tend to be less invasive.
Opportunity cost	The loss of other healthcare programmes displaced by investment in or introduction of another intervention. This may be best measured by the health benefits that could have been achieved had the money been spent on the next best alternative healthcare intervention.
Outcome	Measure of the possible results that may stem from exposure to a preventive or therapeutic intervention. Outcome measures may be intermediate endpoints or they can be final endpoints. See 'Intermediate outcome'.
Oximetry	Pulse oximetry is the non-invasive measurement of the oxygen saturation of a person's haemoglobin usually via a finger or ear lobe sensor.
Oxygen assessment	The process of deciding when to prescribe oxygen to a patient and how much to give them.
Oxygen management	The process of monitoring a patient already receiving oxygen
Placebo	An inactive and physically identical medication or procedure used as a comparator in controlled clinical trials.
Positive predictive value (PPV)	A summary statistic usually used to describe the function of a diagnostic test the positive predictive value is the proportion of people with a positive test that our correctly diagnosed.
Power (statistical)	The ability to demonstrate an association when one exists. Power is related to sample size; the larger the sample size, the greater the power and the lower the risk that a possible association could be missed.
Performance on sub-maximal walk test (distance walked and lowest SaO_2)	Change from baseline in 6 minute walk distance and/or oxygen saturation.
Primary care	Healthcare delivered to patients outside hospitals. Primary care covers a range of services provided by general practitioners, nurses, dentists,

Term	Definition
	pharmacists, opticians and other healthcare professionals.
Primary outcome	The outcome of greatest importance, usually the one in a study that the power calculation is based on.
Product licence	An authorisation from the MHRA to market a medicinal product.
Progression-free survival (PFS)	The time elapsed between treatment initiation and disease progression [defined a priori] or death from any cause, with censoring of patients who are lost to follow-up.
Prospective study	A study in which people are entered into the research and then followed up over a period of time with future events recorded as they happen. This contrasts with studies that are retrospective.
Publication bias	Also known as reporting bias. A bias caused by only a subset of all the relevant data being available. The publication of research can depend on the nature and direction of the study results. Studies in which an intervention is not found to be effective are sometimes not published. Because of this, systematic reviews that fail to include unpublished studies may overestimate the true effect of an intervention. In addition, a published report might present a biased set of results (e.g. only outcomes or sub-groups where a statistically significant difference was found.
Pulmonary function tests (PFTs)	Clinical tests of lung volume and gas exchange.
Pulmonary hypertension	Raised pressure in the pulmonary arterial circulation.
Pulmonary rehabilitation	Multidisciplinary programme of care for patients with chronic respiratory impairment that is individually tailored and designed to optimise the individual's physical and social performance and autonomy.
Psychosocial health	A general term for the broad psychological and social aspects of health
P-value	The probability that an observed difference could have occurred by chance, assuming that there is in fact no underlying difference between the means of the observations. If the probability is less than 1 in 20, the P value is less than 0.05; a result with a P value of less than 0.05 is conventionally considered to be 'statistically significant'.
Quality of life	See 'Health-related quality of life'.
Quality-adjusted life year (QALY)	An index of survival that is adjusted to account for the patient's quality of life during this time. QALYs have the advantage of incorporating changes in both quantity (longevity/mortality) and quality (morbidity, psychological, functional, social and other factors) of life. Used to measure benefits in cost-utility analysis. The QALYs gained are the mean QALYs associated with one treatment minus the mean QALYs associated with an alternative treatment.
Randomisation	Allocation of participants in a research study to two or more alternative groups using a chance procedure, such as computer-generated random numbers. This approach is used in an attempt to ensure there is an even distribution of participants with different characteristics between groups and thus reduce sources of bias.
Rapid deterioration in IPF disease progression	Disease progression over a period of a few weeks. Rapid deterioration can be due to 'acute exacerbation of IPF', which has a specific definition that includes exclusion of known causes of deterioration in IPF, or may be due to one or more of the known complication of IPF.
Randomised controlled trial (RCT)	A comparative study in which participants are randomly allocated to intervention and control groups and followed up to examine

Term	Definition
	differences in outcomes between the groups.
RCT	See 'Randomised controlled trial'.
Relative risk (RR)	The number of times more likely or less likely an event is to happen in one group compared with another (calculated as the risk of the event in group A/the risk of the event in group B).
Reporting bias	See publication bias.
Resource implication	The likely impact in terms of finance, workforce or other NHS resources.
Respiratory bronchiolitis associated interstitial lung disease (RBILD)	A type of idiopathic interstitial pneumonia with characteristic CT and histological appearances. RBILD is often associated with smoking and may have a good prognosis.
Retrospective study	A retrospective study deals with the present/ past and does not involve studying future events. This contrasts with studies that are prospective.
Review question	In guideline development, this term refers to the questions about treatment and care that are formulated to guide the development of evidence-based recommendations.
Secondary outcome	An outcome used to evaluate additional effects of the intervention deemed a priori as being less important than the primary outcomes.
Selection bias	A systematic bias in selecting participants for study groups, so that the groups have differences in prognosis and/or therapeutic sensitivities at baseline. Randomisation (with concealed allocation) of patients protects against this bias.
Sensitivity analysis	A means of representing uncertainty in the results of economic evaluations. Uncertainty may arise from missing data, imprecise estimates or methodological controversy. Sensitivity analysis also allows for exploring the generalisability of results to other settings. The analysis is repeated using different assumptions to examine the effect on the results. One-way simple sensitivity analysis (univariable analysis): each parameter is varied individually in order to isolate the consequences of each parameter on the results of the study. Multi-way simple sensitivity analysis (scenario analysis): two or more parameters are varied at the same time and the overall effect on the results is evaluated. Threshold sensitivity analysis: the critical value of parameters above or below which the conclusions of the study will change are identified. Probabilistic sensitivity analysis: probability distributions are assigned to the uncertain parameters and are incorporated into evaluation models based on decision analytical techniques (For example, Monte Carlo simulation).
Severity of IPF	At the time of diagnosis and approximately 3 monthly time intervals thereafter, a combination of lung function tests and subjective descriptions of symptoms are used to determine the prognosis for people with IPF.
Significance (statistical)	A result is deemed statistically significant if the probability of the result occurring by chance is less than 1 in 20 ($p < 0.05$).
Six minute walk test (6MWT)	A lung function test in which a patient is asked to walk at a comfortable pace and is recorded under supervision and the total distance and the oxygen saturation determined.

Term	Definition
Specialist networks	Groups of hospital trusts that work together to provide care for people with IPF. In this instance there may be a central hospital with a full MDT and a number of peripheral centres linked to this.
Stakeholder	Those with an interest in the use of the guideline. Stakeholders include manufacturers, sponsors, healthcare professionals, and patient and carer groups.
Surgical lung biopsy	A biopsy obtained via a surgical procedure in which the lung is accessed from the skin surface rather than via the bronchial surface (i.e. this definition does not include bronchoscopy)
Survival rate	A summary statistic derived from following a cohort of people with IPF which can be reported in terms of deaths per person years – or more intuitively to clinicians in terms of 1 year, 5 year and median survival.
Suspected idiopathic pulmonary fibrosis	When IPF is included as part of the differential diagnosis.
Systematic review	Research that summarises the evidence on a clearly formulated question according to a pre-defined protocol using systematic and explicit methods to identify, select and appraise relevant studies, and to extract, collate and report their findings. It may or may not use statistical meta-analysis.
Time horizon	The time span over which costs and health outcomes are considered in a decision analysis or economic evaluation.
Time to disease progression	An outcome used mainly in clinical trials and allowing calculation of rates of disease progression for hypothesis testing.
Tissue sample	A sample (biopsy) obtained for diagnostic and prognostic purposes.
Transbronchial lung biopsy	A sample of lung tissue obtained from the bronchial surface using a bronchoscope. This amount of tissue obtained here is far less than that obtained in a surgical biopsy and this limits the usefulness of this test, but it is deemed a safer test than compared to surgical lung biopsy.
Transthoracic Doppler echocardiography (TCC)	Non-invasive ultrasound method used to estimate the pulmonary artery pressure
Treatment allocation	Assigning a participant to a particular arm of the trial.
TLCO	See DLCO.
Univariable	Analysis which separately explores each variable in a data set.
Usual interstitial pneumonia	The characteristic histological findings in people with IPF.
Utility	A measure of the strength of an individual's preference for a specific health state in relation to alternative health states. The utility scale assigns numerical values on a scale from 0 (death) to 1 (optimal or 'perfect' health). Health states can be considered worse than death and thus have a negative value.
Video assisted lung biopsy	A minimally invasive surgical lung biopsy technique.
Vital Capacity (VC)	A lung function measure which is the total volume of air that can be exhaled from the lungs after a full inspiration. In contrast to the FVC, the exhalation does not need to be forced. In people with IPF the VC and FVC are usually very close in value – but for people with emphysema the VC may be considerably higher than the FVC.