NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Premeeting briefing

Nintedanib for treating idiopathic pulmonary fibrosis

This premeeting briefing presents:

- the key evidence and views submitted by the company, the consultees and their nominated clinical experts and patient experts and
- the Evidence Review Group (ERG) report.

It highlights key issues for discussion at the first Appraisal Committee meeting and should be read with the full supporting documents for this appraisal.

Please note that this document includes information from the ERG before the company has checked the ERG report for factual inaccuracies.

Key issues for consideration

Clinical effectiveness

Generalisability

- During the NICE appraisal of pirfenidone, clinical experts suggested that people with a percent predicted forced vital capacity (FVC) greater than 80% are not normally treated in current clinical practice. Is this true of current clinical practice?
- Are results from the nintedanib INPULSIS and TOMORROW trials generalisable to patients seen in clinical practice in England?
  - Excluded people with a percent predicted FVC below 50%.
  - People with a percent predicted FVC of more than 80% accounted for around 45% of patients in the model. Is this generalisable to the population that would be treated in clinical practice?

  - Included people aged at least 40 years, with a percent predicted diffusion capacity of the lung for carbon monoxide (DLCO), corrected for haemoglobin, of 30–79%.
  - Excluded people with a history of haemorrhagic central nervous system event, myocardial infarction or unstable angina; people with airway obstruction; and people receiving therapeutic anticoagulation.
- What is the relevant comparator for nintedanib based on current clinical practice?
  - For people with percent predicted FVC <50%?
  - For people with percent predicted FVC 50–80%?
  - For people with percent predicted FVC >80%?
- Is N-acetylcysteine (NAC) used, either alone or as triple therapy, to treat idiopathic pulmonary fibrosis?
- What are the most clinically meaningful endpoints in idiopathic pulmonary fibrosis?
- How is idiopathic pulmonary fibrosis diagnosed?

Disease progression

- Is there a relationship between the rate of change in FVC and improvements in clinically meaningful endpoints (mortality and morbidity)?
- How is disease progression defined in clinical practice?
- Is a 10 percentage-point reduction in percent predicted FVC a clinically meaningful change in lung function?
- Is 'exacerbation' (on which the model depends) a clearly defined event in idiopathic pulmonary fibrosis?
- Can lung function parameters (such as FVC) be used to identify subgroups of people who should be considered separately?
  - Is percent predicted FVC (<50%, 50–80%, >80%) used to distinguish between patients in clinical practice?

Effectiveness of nintedanib and place in the treatment pathway

- Is nintedanib more clinically effective than pirfenidone? More effective than placebo?
  - How does the survival of the 2 drugs compare?

- Is nintedanib likely to be used at the same point in the pathway as pirfenidone, or only when pirfenidone is contraindicated or not tolerated?
- Is there evidence to support the concomitant use of nintedanib and pirfenidone?
- Are the methods of the company's network meta-analysis base case and scenario analyses robust?
  - Is the company's choice to exclude some trials appropriate?
  - Is the potential for bias associated with the different length of follow-up between nintedanib and pirfenidone studies a key issue?
- In what setting should nintedanib be used (primary or secondary care, or specialist clinics)?
- Is the potential for bias associated with missing data estimation in the TOMORROW trial (last observation carried forward) a key issue?

Stopping rules

- The NICE technology appraisal guidance 282 states that treatment with pirfenidone should be discontinued if there is evidence of disease progression (an absolute decline in percent predicted FVC of 10% or more within any 12 month period). Is this followed in clinical practice?
- Do clinicians anticipate that it would be appropriate to apply stopping rules when treating people with nintedanib?

Cost effectiveness

Generalisability and robustness

- Does a model based on FVC and exacerbations reflect what happens in clinical practice?
- Is the population modelled by the company appropriate (FVC ≥50%, no upper limit)?
- The company and ERG base case ICERs do not account for stopping rules, is this appropriate?
- The company applied different scenarios from the network meta-analysis (fixed effect model) for each outcome. Is this appropriate?

Adverse events

- Is it appropriate for the company to have excluded diarrhoea as an adverse event from its economic model?
- Is the company assumption that adverse events last 1 year appropriate?

Exacerbations

- Did the company include appropriate exacerbation-related disutility? (sudden decline in health-related quality of life [−0.14] for 1st month, followed by a smaller and longer-lasting disutility [−0.078] in subsequent months).
- The model assumed that people who have experienced an exacerbation have the same risk of recurrent exacerbations as people that have not had an exacerbation. Is this appropriate?
- Do exacerbations impact the cost effectiveness results? Would changing the assumptions about exacerbations affect the results?

Survival

- The model is sensitive to the degree to which nintedanib (or pirfenidone) increases or decreases the risk of dying. The company and ERG base case use a loglogistic model to extrapolate overall survival data. Is this appropriate?

Cost-effectiveness results

- The ERG suggested that nintedanib and pirfenidone could be considered to be equally cost effective. Is this appropriate?

1 Remit and decision problems

1.1 The remit from the Department of Health for this appraisal was: to appraise the clinical and cost effectiveness of nintedanib within its licensed indication for treating idiopathic pulmonary fibrosis.

Table 1. Decision problem

	Final scope issued by NICE	Decision problem addressed in the submission	Comments from the ERG
Population	Adults with idiopathic pulmonary fibrosis	As in the scope	Appropriate
Intervention	Nintedanib	As in the scope (150 mg twice daily)	Appropriate
Comparators	• pirfenidone • best supportive care	• pirfenidone • best supportive care • N-acetylcysteine also included in the network meta-analysis, but not included in the company's model because the drug was not listed as a comparator in the final scope	Appropriate; reasonable to include N-acetylcysteine
Outcomes	• pulmonary function parameters • physical function • exacerbation rate • progression-free survival • mortality • adverse effects of treatment • health-related quality of life	As in the scope	Appropriate and clinically meaningful. The company did not state which specific pulmonary or physical function parameters are the most clinically meaningful or how these parameters or acute exacerbations predict patient prognosis. But the ERG agreed with the company's approach to defining loss of lung function in the network meta-analysis (see section 4.16)

2 The technology and the treatment pathway

2.1 Idiopathic pulmonary fibrosis is a chronic, progressive lung disease in which scarring (fibrosis) occurs. The cause of idiopathic pulmonary fibrosis is unknown although it is thought to be related to an abnormal immune response to an unknown cause. 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 death. The median survival for people with idiopathic pulmonary fibrosis in the UK is approximately 3 years from the time of diagnosis. The aim of treatment is to manage the symptoms and slow progression.

2.2 The severity of idiopathic pulmonary fibrosis is usually based on 2 measures of lung function:

- forced vital capacity (FVC): the maximum volume of gas that can be exhaled, from full inhalation, by exhaling as forcefully and rapidly as possible
- diffusion capacity of the lung for carbon monoxide (DLCO): measures how much oxygen travels from the lungs into the blood stream.

FVC and DLCO can be expressed as a percentage of the predicted normal value for a person of the same sex, age and height (DLCO is also corrected for haemoglobin). Lower percent predicted values indicate worsened health. Clinical trials generally define mild-to-moderate idiopathic pulmonary fibrosis as a percent predicted FVC of at least 50–55% and a percent predicted DLCO of at least 35–40%. A percent predicted FVC lower than 50%, and a percent predicted DLCO lower than 35%, indicates severe disease. The company suggested that a decline in percent predicted FVC of over 10% (absolute value) represents a clinically

important difference (based on published literature and consultation with clinical experts) (see section 3.4 for the opinions from professional group submissions).

2.3 NICE clinical guideline 163 on the diagnosis and management of suspected idiopathic pulmonary fibrosis recommends that best supportive care (including symptom relief, management of co-morbidities, withdrawal of therapies suspected to be ineffective or causing harm and end of life care) should be offered to people from diagnosis and be tailored according to disease severity, rate of progression and the person's preference. If pharmacological treatment is considered appropriate, the guideline recommends use of pirfenidone if a person's forced vital capacity (FVC) is between 50% and 80% of their predicted value in line with recommendations in NICE technology appraisal guidance 282 (a review of this guidance is in progress, expected publication May 2016). Treatment with pirfenidone should be discontinued if there is evidence of disease progression (an absolute decline in percent predicted FVC of 10% or more within any 12 month period). Lung transplantation is an option if there are no contraindications.

Table 2. Technology

	Nintedanib	**Pirfenidone**
Marketing authorisation	Indicated in adults for the treatment of idiopathic pulmonary fibrosis	Indicated in adults for the treatment of mild to moderate idiopathic pulmonary fibrosis
Mechanism of action	Tyrosine kinase inhibitor; targets 3 growth factor receptors involved in idiopathic pulmonary fibrosis	Immunosuppressant; anti-inflammatory and antifibrotic effects
Posology and method of administration	Oral, 150 mg twice daily. If the 150 mg dose is not tolerated, 100 mg dose twice daily is recommended. Source: summary of product characteristics	Oral, three 267 mg capsules three times daily (total of 2403 mg/day across nine capsules). Initiated with one 267 mg capsule, three times daily, on days 1–7. Titrated to recommended dose over 14 days. Source: summary of product characteristics

	Nintedanib	Pirfenidone
Cost	£2151.10 for 60 x 150mg capsules [taken from company submission, confirmed on MIMS online June 2015]. This equates to a daily cost of £71.70 (2 capsules /day). A confidential patient access scheme makes nintedanib available to the NHS at a lower cost	£501.92 for a 63-capsule pack, £2007.70 for a 252-capsule pack, £2151.10 for a 270-capsule pack [BNF edition 69].This equates to a daily cost of £71.70 (9 capsules/day). A confidential patient access scheme makes pirfenidone available to the NHS at a lower cost

See summary of product characteristics for details on adverse reactions and contraindications.

3 Comments from consultees

3.1 Patient experts reported that idiopathic pulmonary fibrosis is a distressing, progressive illness that limits physical activity because of breathlessness. Coughing spasms can be debilitating and uncontrollable. People with idiopathic pulmonary fibrosis often experience anxiety and depression because there is no cure for the disease, and may become isolated from family and friends because they avoid situations where they might be embarrassed by their symptoms.

3.2 Patient experts reported that the outcomes most important to people with idiopathic pulmonary fibrosis, in the absence of a cure, are slowed disease progression and reduced cough and breathlessness. Professional groups suggested that the most clinically relevant endpoints are preserved quality of life, reduced morbidity and mortality, and fewer acute exacerbations.

3.3 A number of consultees noted that the only treatment for idiopathic pulmonary fibrosis is pirfenidone, which patient experts stated only 35% of people are able to take because of prescribing restrictions (based on forced vital capacity [FVC]) and intolerance to the side effects (such as nausea, lethargy, loss of appetite, weight loss, liver toxicity and photosensitive skin rash). Professional groups noted that although lung transplant is an option for people with idiopathic pulmonary fibrosis, many are ineligible because of advanced age and co-morbidities. Comments from patient experts indicated that the majority of people with idiopathic

pulmonary fibrosis would find the balance of nintedanib's benefit-to-harm acceptable, but that the gastrointestinal side effects of the drug may worsen the quality of life of people with very advanced disease. Professional groups reported concerns that nintedanib may impair wound healing or increase the risk of aneurysm (although the summary of product characteristics does not mention this).

3.4 Submissions from professional groups highlighted challenges in diagnosing idiopathic pulmonary fibrosis, distinguishing subgroups according to severity and measuring treatment benefit. They suggested that if nintedanib is recommended by NICE, guidance should be included on the criteria for diagnosing patients with idiopathic pulmonary fibrosis suitable for this treatment. However, some consultees advised that applying eligibility criteria and stopping criteria based on only a single physiological parameter (such as FVC) may be inappropriate, because idiopathic pulmonary fibrosis is a heterogeneous condition with considerable variation in prognosis and rates of disease progression. Professional groups agreed that an absolute decline in percent predicted FVC of over 10% during a 24 week period reflects disease progression and predicts poor prognosis. However, some suggested that it is unclear whether modifying the rate of change in FVC, or any other surrogate endpoints, translates to improvements in clinically meaningful endpoints (mortality and morbidity). This suggests that it would not be feasible to use lung function parameters to identify subgroups of people who should be considered separately because they have a different prognosis or different capacity to benefit from treatment. Some consultees disagreed, stating that a change in FVC is reflective of a change in mortality risk.

3.5 Professional groups noted an unmet need in people with idiopathic pulmonary fibrosis and co-existing emphysema, who have a better prognosis and are likely to benefit from treatment. Because these people usually have an FVC greater than 80% of the predicted normal value they are ineligible to receive pirfenidone under current NICE guidance. Professional groups also reported geographic variation in the availability

of specialist services for Interstitial Lung Disease and the time taken for people to access them and receive treatment. They agreed that diagnosis of and treatment for idiopathic pulmonary fibrosis should be provided in a specialist or tertiary setting.

3.6 Professional groups were unsure where nintedanib would fit in the treatment pathway. That is, which of pirfenidone or nintedanib would be used first, and whether the drugs could be used sequentially or concomitantly. But they noted that nintedanib would provide a valuable option for people who are ineligible for pirfenidone. They reported that most specialist Interstitial Lung Disease centres have experience of using nintedanib and have systems for managing its side effects.

3.7 Professional groups submitted varied opinions about whether the clinical trials of nintedanib reflected current UK practice. It was suggested that:

- The clinical trials represented a group with fewer co-morbidities than people in clinical practice, because of substantial exclusion criteria (such as a history of haemorrhagic central nervous system event, myocardial infarction or unstable angina; people with airway obstruction; and people receiving therapeutic anticoagulation).
- The pragmatic design of the nintedanib clinical trials makes them generalisable to clinical practice.
- The accuracy of the diagnoses of idiopathic pulmonary fibrosis may have been limited, because they were made at a district general hospital rather than a specialist centre.

4 Clinical-effectiveness evidence

Overview of the clinical trials

4.1 The clinical evidence for nintedanib in idiopathic pulmonary fibrosis came from 3 randomised, double-blind trials, comprising 2 phase III trials (INPULSIS-1 [n=513] and INPULSIS-2 [n=548]) and a phase IIb trial (TOMORROW [n=428]). All 3 trials investigated the effect of treatment

with nintedanib for 52 weeks in adults aged 40 years or older with idiopathic pulmonary fibrosis. The trial included only people with a forced vital capacity (FVC) of at least 50% of the predicted normal value and a diffusion capacity of the lung for carbon monoxide (DLCO) of 30–49% of the predicted normal value at baseline. The trial excluded people if they had a history of a haemorrhagic central nervous system event (previous 12 months); myocardial infarction (previous 6 months); gastrointestinal bleeding or ulcers, major injury or surgery (previous 3 months); or unstable angina (previous month). People were also excluded if they had an inherited predisposition to bleeding, or required full-dose anticoagulation or antiplatelet therapy.

- INPULSIS-1 and INPULSIS-2 had near-identical methodology (differing only in the countries involved) and both had an additional 28-day follow-up period. Patients received nintedanib 150 mg twice a day (n=309 in INPULSIS-1 and n=329 INPULSIS-2) or placebo (n=204 in INPULSIS-1 and n=219 INPULSIS-2). If the person deteriorated or had an acute exacerbation, the INPULSIS trials allowed concomitant use of a range of medications (including azathioprine, cyclophosphamide, cyclosporine, N-acetylcysteine, and prednisone). The company stated that the background medications used closely resembled “best supportive care” in idiopathic pulmonary fibrosis.
- The phase II (dose-finding) TOMORROW trial compared 4 doses of nintedanib: patients were randomised to either nintedanib 50 mg once a day (n=86), nintedanib 50 mg twice a day (n=86), nintedanib 100 mg twice a day (n=86), nintedanib 150 mg twice a day (n=85), or placebo (n=85). The trial allowed concomitant treatment with prednisone.

4.2 The trials were conducted in Australia, Europe, North America and Asia. The TOMORROW trial recruited from 92 sites, including 5 in the UK, across 25 countries. The INPULSIS trials recruited from 205 sites, including 8 in the UK (7 of which were in England, n=45) across 24 countries. The company indicated that baseline characteristics were generally similar in the different arms of the trials (see tables 16 and 17 on

pages 60–61 of the company submission). The mean percent predicted FVC at baseline was approximately 80% in all 3 trials. The company submission stated that all randomised patients were included in the intent-to-treat analysis. The trial publication stated analyses were performed for randomised patients who received at least 1 dose of study medication (approximately 99.5% of the randomised population).

4.1 The primary outcome of all 3 trials was the annual rate of decline in FVC (expressed in mL per year). The key secondary outcomes in the INPULSIS phase III trials were time to first acute exacerbation, and absolute change from baseline in Saint George's Respiratory Questionnaire total score at 52 weeks. The Saint George's Respiratory Questionnaire is a healh-related quality of life instrument measuring frequency and duration of symptoms (including cough, sputum production and breathlessness), activities that cause breathlessness or are limited because of breathlessness, and impacts on daily life (including changes to employment and adverse effects of treatment). Of note, the definition of acute exacerbations varies across trials for idiopathic pulmonary fibrosis. The INPULSIS trials defined acute exacerbations as events meeting all of the following criteria:

- unexplained worsening or development of dyspnoea (previous 30 days)
- new pulmonary infiltrates (on chest x-ray, high resolution computed tomography, or both) or parenchymal abnormalities with no pneumothorax or pleural effusion since the last visit
- exclusion of any known causes of acute worsening (infection, heart failure, pulmonary embolism, acute lung injury).

Outcomes were assessed at weeks 0, 2, 4, 6, 12, 24, 36 and 52 in all 3 trials. EQ-5D data were collected at weeks 0, 12, 24 and 52 of the INPULSIS trials.

ERG comments

4.2 The ERG considered that the trials were generally of good methodological quality. The ERG reported some differences in baseline characteristics between UK patients in INPULSIS-1 and the overall population in INPULSIS-1 (see page 29 of the ERG report), but stated that it was not clear whether these were significant (INPULSIS-2 enrolled no UK patients and the company did not provide analyses of UK patients from TOMORROW). The ERG stated that the inconsistency in reporting of the intent-to-treat analyses (section 4.2) was unlikely to bias results because of the high proportion of randomised patients who received medication.

4.3 The ERG highlighted that the 3 nintedanib trials enrolled people with a percent predicted FVC of at least 50% and therefore do not provide evidence for people with more severe disease, which they considered to limit the evidence base.

Clinical trial results

4.4 The key outcomes from the 3 nintedanib trials, and from a pre-planned pooled analysis of the 2 INPULSIS trials, are presented in Table 3. The annual rate of decline in FVC with nintedanib was approximately half that of placebo; this difference was statistically significant ($p<0.05$) in all 3 trials. Fewer people randomised to nintedanib treatment died compared with placebo in all 3 trials, but these differences were not statistically significant. The treatment effects for health-related quality of life (measured using the St. George's Respiratory Questionnaire) and time to first acute exacerbation were inconsistent across the trials:

- In INPULSIS-1, there were no significant differences between nintedanib and placebo for these outcomes (Table 3)
- In INPULSIS-2 the differences showed a benefit in favour of nintedanib treatment, and were statistically significant (Table 3)
- The pooled analysis of the 2 trials showed a non-significant trend in favour of nintedanib for these outcomes (Table 3).

The company noted that the INPULSIS trials were not powered to detect the effect of nintedanib on acute exacerbations.

Table 3. Clinical trial outcomes from INPULISIS-1, INPULSIS-2, and the TOMORROW trials (shaded cells reflect significant differences in outcomes)

Study	Annual rate of FVC decline (mL/yr)	FVC responders[a]	≥1 acute exacer-bation in 52 weeks[b]	SGRQ score (change from baseline)[c]	Death (all cause)
INPULSIS-1 (source: company submission table 19)					
Nintedanib 150 mg twice/day	−114.7	218/309 (70.6%)	19/309 (6.1%)	4.34	13/309 (4.2%)
Placebo	−239.9	116/206 (56.9%)	11/206 (5.4%)	4.39	13/206 (6.4%)
HR/MD/OR (95% CI) p value	MD: 125.3 (77.7, 172.8) p<0.0001	OR: 1.91 (1.32, 2.79) p=0.0007	HR: 1.15 (0.54, 2.42) p=0.67	MD: −0.05 (−2.50, 2.40) p=0.97	HR: 0.63 (0.29,1.36) p=0.29
INPULSIS-2 (source: company submission table 19)					
Nintedanib 150 mg twice/day	−113.6	229/331 (69.6%)	12/331 (3.6%)	2.80	22/331 (6.7%)
Placebo	−207.3	140/220 (63.9%)	21/220 (9.6%)	5.48	20/220 (9.1%)
HR/MD/OR (95% CI) p value	MD: 93.7 (44.8, 142.7) p=0.0002	OR: 1.29 (0.89, 1.86) p=0.18	HR: 0.38 (0.19, 0.77) p=0.005	MD: −2.69 (−4.95, −0.43) p=0.02	HR: 0.74 (0.40, 1.35) p=0.30
Pooled analysis (INPULSIS-1 and INPULSIS-2) (source: nintedanib Summary of Product Characteristics tables 2–6)					
Nintedanib 150 mg twice/day	−113.6	447/640 (70.1%)	31/640 (4.9%)	3.53	35/640 (5.5%)
Placebo	−223.5	256/426 (60.5)	32/426 (7.6%)	4.96	33/426 (7.8%)
HR/MD/OR (95% CI) p value	MD: 109.9 (75.9, 114.0) p<0.0001	OR: 1.58 (1.21, 2.05) p=0.0007	HR: 0.64 (0.39, 1.05) p=0.08	MD: −1.43 (−3.09, 0.23) p=0.09	HR: 0.70 (0.43, 1.12) p=0.70
TOMORROW (source: company submission table 18 and clarification response, nintedanib Summary of Product Characteristics tables 2–6, Richeldi et al. 2011 NEJM 365: 1079)					
Nintedanib 150 mg twice/day	−60	NR	2/86 (2.3%)	0.66	7/86 (8.1%)
Placebo	−190	NR	12/87 (13.8%)	5.46	9/87 (10.3%)
HR/MD/OR (95% CI) p value	MD: 131 (27, 235) p=0.01	NR	HR: 0.16 (0.04, 0.71) p=0.005	MD: −6.12 (−10.57, −1.67) p=0.007	NR p>0.05

[a] FVC responders: people with absolute decline in percent predicted FVC <10% at 52 weeks
[b] Investigator-reported acute exacerbations (according to the criteria described by the trial protocol); hazard ratio is based on analysis of time to first event
[c] A positive change in SGRQ represents worsened health-related quality of life
Key: CI, confidence interval; FVC, forced vital capacity; HR, hazard ratio; MD, mean difference; NR, not reported; OR, odds ratio; SGRQ, St. George's Respiratory Questionnaire
Source: See Appendix A for extracts from the Summary of Product Characteristics

4.5 The company presented the results of a pre-planned pooled subgroup analysis of people with a percent predicted FVC of more than 70% at baseline, compared with people with a baseline percent predicted FVC of 70% or lower (but greater than 50%), using combined data from the INPULSIS trials. During clarification, the company stated that it chose these FVC values for consistency with a subgroup analysis of the phase II TOMORROW trial. The company justified this choice by noting that there are no accepted criteria to define mild, moderate or severe idiopathic pulmonary fibrosis. It highlighted that the collective guideline from the American Thoracic Society, the European Respiratory Society, the Japanese Respiratory Society, and the Latin American Thoracic Association (ATS/ERS/JRS/ALAT, 2011) did not propose a formal staging system to classify disease severity. The company conducted its subgroup analyses on the primary and key secondary endpoints: annual rate of decline in FVC, time to first exacerbation and change in St. George's Respiratory Questionnaire total score from baseline. FVC declined more slowly in people randomised to nintedanib compared with placebo in both subgroups. There company did not present results, but stated that there were no statistically significant differences in the outcomes by subgroup.

4.6 The company also did post-hoc subgroup analyses using pooled INPULSIS data, for:

- people with a baseline percent predicted FVC of more than 80%, compared with people with percent predicted FVC of 80% or lower (described in the company response to clarification, numerical results not presented)
- people with a baseline percent predicted FVC of more than 90%, compared with people with percent predicted FVC of 90% or lower (numerical results presented on page 66 of the company submission)
- people with or without emphysema at baseline (described on page 66 of the company submission, numerical results not presented).

Again, the company did not present results, but stated that there were no statistically significant differences in the primary or key secondary outcomes by subgroup. The company did not present a rationale for the subgroup analysis in people with or without emphysema. A professional group submission suggested that people with co-existing emphysema have a slightly better prognosis (see section 3.5).

ERG comments

4.7 Based on consultation with a clinical expert, the ERG concluded that the outcomes presented by the company were appropriate and adequately defined. But they were concerned that the company did not fully explain how pulmonary and physical function parameters or acute exacerbations predict patient prognosis, meaning that it was unclear which specific outcomes were the most clinically meaningful. A clinical adviser to the ERG suggested that progression-free survival is the most clinically important outcome, but that there is not consensus on this. The ERG also noted that the St. George's Respiratory Questionnaire is not a disease-specific measure for idiopathic pulmonary fibrosis, so it may not fully reflect changes in health-related quality of life in this population.

4.8 The ERG suggested that the company's approach to estimate missing data for the intention to treat analysis of secondary outcomes in the TOMORROW trial (last observation carried forward) could potentially bias the trial outcomes in favour of nintedanib. The ERG also noted an imbalance in drop-outs between the placebo and nintedanib arms which the company did not discuss in its submission. The ERG was uncertain whether the primary analysis used in the INPULSIS trials to account for missing data was appropriate because the company did not explain their assumptions.

4.9 The ERG considered that a subgroup analysis using a percent predicted FVC threshold of 80% would have better informed the efficacy of nintedanib in different populations than the company's pre-planned analysis using the 70% threshold. This was based on clinical advice that a

percent predicted FVC greater than 80% indicates mild disease, 50–80% indicates moderate disease and lower than 50% indicates severe disease. The ERG was unable to locate the published reference for the company's post-hoc analysis using the 80% threshold (which the company did not provide in its submission) and therefore was not able to comment on it.

Meta-analyses

4.10 As there are no head to head trials comparing nintedanib with pirfenidone, the company did a systematic review and identified the following 8 randomised placebo-controlled trials to include in a network meta-analysis:

- 3 trials of nintedanib compared with placebo (described in section 4.1): INPULSIS-1, INPULSIS-2 and TOMORROW
- 5 trials of pirfenidone compared with placebo: SP2, SP3, CAPACITY-1, CAPACITY-2 and ASCEND.

The company also included 4 trials of N-acetylcysteine in their network meta-analysis. However, the company did not include this evidence for N-acetylcysteine in its cost-effectiveness model (and is not presented in this document), because the drug was not listed as a comparator in the final scope. The company did not generally include outcomes data for unlicensed doses of nintedanib and pirfenidone, which were investigated in the comparator arms of some trials. However, the company considered that the low dose of pirfenidone used in 2 Japanese studies (SP2 and SP3) was comparable to the UK licensed dose, because of the lower average weight of people in the Japanese studies. The company included the results from the Japanese studies in their base-case network meta-analysis and tested the effect of excluding them in sensitivity analyses.

4.11 The base case network meta-analysis presented by the company included all relevant data. However, the company also presented a range of sensitivity analyses, which differed across outcomes. These scenarios

excluded trial data according to, for example, differences in patient characteristics, potential biases, or to limit data to phase III trials. For each scenario, the company presented results from a random effects and a fixed effect model. The company used data from the fixed effect model in the economic model. The company included all evidence for overall survival in the model; for acute exacerbations it excluded 3 pirfenidone trials in Japanese populations ('scenario 3'), and for decline in lung function it excluded the ASCEND study of pirfenidone by King et al because of heterogeneity ('scenario 2'). These results are presented in Table 4. Full results for all outcomes in the base-case analysis are presented in section 4.10 (pages 115–139) of the company submission; Appendix B of the company submission presents the results of all sensitivity analyses. The results for overall survival for nintedanib and pirfenidone were comparable, and neither drug showed a significant difference in mortality compared with placebo. The base-case analysis of acute exacerbations also showed comparable results for nintedanib and pirfenidone, but the company reported uncertainty in the results which it attributed to heterogeneity in the Japanese trials of pirfenidone (SP2 and SP3). After excluding these trials from the network meta-analysis ('scenario 3' of the sensitivity analyses for this outcome), the results showed fewer acute exacerbations with nintedanib than pirfenidone. However, the company noted that these results should be interpreted with caution because the definition of acute exacerbations varied across the trials (pages 110–111 of the company submission describes the definitions). The company's analysis of loss of lung function (defined by the company as an absolute decline in percent predicted FVC of over 10%) gave similar results for nintedanib and pirfenidone using the base case network meta-analysis. After excluding a trial from the network meta-analysis because of heterogeneity ('scenario 2' of the sensitivity analyses for this outcome), the results showed a statistically significant reduction in loss of lung function with nintedanib compared with placebo. The company did not report results of significance testing for the comparison of nintedanib with pirfenidone.

Table 4. Network meta-analysis of efficacy outcomes (fixed effect model): scenarios used in the company cost-effectiveness model (shaded cells reflect significant differences in outcomes; results of significance testing not reported for comparison of nintedanib with pirfenidone)

Comparison	Median odds ratio (95% CI)		
	Overall survival[a] (NMA base case: all evidence)	**Acute exacerbations** (NMA scenario 3: excluded heterogeneous trials)	**Loss of lung function**[b] (NMA scenario 2: excluded heterogeneous trial)
Nintedanib compared with placebo	0.70 (0.45, 1.10)	0.56 (0.35, 0.89)	0.54 (0.42, 0.69)
Pirfenidone compared with placebo	0.70 (0.46, 1.05)	1.01 (0.22, 4.50)	0.69 (0.47, 1.00)
Nintedanib compared with pirfenidone[c]	1.00 (0.55, 1.85)	0.56 (0.12, 2.68)	0.78 (0.49, 1.22)

[a] Assessed as a dichotomous outcome
[b] Defined as an absolute decline in percent predicted FVC of over 10% by the end of the study follow-up
[c] Results of significance testing not reported
Key: CI, confidence interval; FVC, forced vital capacity; NMA, network meta-analysis
Source: table 49 of the company submission, tables 24 and 44 of appendix B to the company submission

4.12 The company included progression-free survival as an outcome in the network meta-analysis, but not in its model. The company did an indirect pairwise comparison of nintedanib and pirfenidone (through placebo), because data were only available from the INPULSIS trials and the CAPACITY trials. This gave an estimated hazard ratio for nintedanib of 1.00 (95% credible interval 0.71 to 1.39, p=0.98). The company noted that progression-free survival was not reported or defined in the same way across studies. To allow comparison between nintedanib and pirfenidone studies, the company adopted the definition of progression-free survival used in the pirfenidone CAPACITY studies to do a post-hoc analysis of individual patient data from the INPULSIS studies.

4.13 The company identified a recently published network meta-analysis comparing treatments for idiopathic pulmonary fibrosis (Loveman et al.

2015) and reported that the results of the comparison between nintedanib and pirfenidone were similar to the results from its network meta-analysis. The company attributed any minor differences in results to methodological differences between the 2 analyses. The company noted that, for the outcomes compared, neither Loveman et al. nor the results of its own meta-analysis showed significant differences between nintedanib and pirfenidone (see page 88 of the company submission).

ERG comments

4.14 The ERG stated that the company's network meta-analysis appeared to be of good methodological quality and included all relevant trials, although the description of the methods used was brief. The trials were all placebo-controlled randomised controlled trials and therefore all comparisons in the network meta-analysis were made through placebo; the ERG therefore believes that the company's inclusion of N-acetylcysteine in the network had little influence on the results for nintedanib and pirfenidone.

4.15 The ERG's key concerns with the network meta-analysis were:

- some outcomes were informed by only a small number of trials (for example, only the INPULSIS and CAPACITY trials informed the serious cardiac adverse event outcome)
- there is the potential for bias in favour of nintedanib because the company excluded studies in some scenarios
- the different length of follow-up between studies may have introduced bias in favour of nintedanib.

Although the company provided a rationale for excluding particular studies in the different network meta-analysis scenarios, they did not describe its overarching logic for using different scenarios across the outcomes. The ERG acknowledged that excluding this evidence provided a better reflection of the UK population. The other potential area for bias identified by the ERG was that data for nintedanib was based on a 52 week

follow-up whereas the pirfenidone trials had follow-up periods ranging from 36 weeks to 72 weeks. The ERG suggested that trials with a shorter follow-up might observe fewer negative outcomes (such as exacerbations, decline in lung function, deaths), because idiopathic pulmonary fibrosis is a rapidly progressing disease. The company acknowledge this issue but did not perform sensitivity analyses to explore the effect. Clinical advice to the ERG indicated that a difference of 20 weeks might be too short to observe a difference in loss of lung function or mortality.

4.16 The ERG noted that the company did not base the network meta-analysis for loss of lung function on the primary outcome of the nintedanib trials, but on a 10-point decrease in percent predicted FVC (absolute value). The ERG agreed with this definition of loss of lung function, and that this decrease represents a clinically important difference, based on the published literature provided by the company and the opinion of a clinical expert consulted by the ERG. However, they stated that this outcome may be subject to some bias because the analysis was based on post-hoc analyses of observed data with no imputation of missing information, and proportionally more patients in the nintedanib arm dropped out of the TOMORROW study compared with the placebo arm.

4.17 The ERG was concerned that the progression-free survival outcome may be subject to some bias because the company did not include data from the TOMORROW trial, and did not provide a rationale for this. The ERG could not verify the outcome data for progression-free survival because they came from unpublished post-hoc analyses. The company model did not include progression-free survival as an outcome.

Adverse effects of treatment

4.18 The company reported that the most frequent adverse events with nintedanib compared with placebo were gastrointestinal. These were of mild to moderate severity and rarely led to stopping treatment; diarrhoea was reported in over 60% of people receiving nintedanib in the INPULSIS trials (compared with 19% of people receiving placebo), and led to

stopping nintedanib in fewer than 5% of people. Other commonly reported adverse events in the INPULSIS trials (10% or greater) were nausea, nasopharyngitis, cough, bronchitis, upper respiratory tract infection, dyspnoea, decreased appetite vomiting and weight loss (results are presented in table 93 of the company submission). The proportion of people with serious adverse events, or who discontinued study drug because of adverse events, was similar between trials arms. However, compared with placebo, more people receiving nintedanib had myocardial infarctions (1.3% compared with 0.5% in INPULSIS-1, 0.9% compared with 0.5% in INPULSIS-2). Elevated liver enzymes were more common in people receiving nintedanib (13.6%, based on pooled data from the INPULSIS trials) compared with placebo. The company noted that elevations of liver enzymes were reversible and not associated with liver disease.

4.19 The company evaluated 4 safety outcomes in its network meta-analysis. It reported that, compared with those receiving placebo, people receiving nintedanib were more likely to have severe gastrointestinal events (odds ratio 2.24, 95% CI 0.98 to 5.11, p=0.055), stop study drug (odds ratio 1.14, 95% CI 1.01 to 1.86, p=0.014) and have adverse effects that led to stopping study drug (odds ratio 1.52, 95% CI 1.12 to 2.07, p=0.007). Results of the meta-analysis indicated that nintedanib was associated with fewer serious cardiac events than placebo and pirfenidone, but none of the odds ratios were statistically significant and the company reported wide confidence intervals. The analysis of serious gastrointestinal events showed that nintedanib was associated with more serious events than pirfenidone (odds ratio 3.96, 95% CI 1.18 to 14.51, p value not reported). The company reported that nintedanib was associated with lower discontinuation because of adverse events than pirfenidone (odds ratio 0.88, 95% CI 0.57 to 1.37, p value not reported).

Additional evidence from the summary of product characteristics

4.20 As noted in section 3.6, professional groups questioned whether people could use pirfenidone and nintedanib concomitantly. The company did not submit any evidence for concomitant use. The summary of product characteristics for nintedanib provides details of a parallel group study in 24 people with idiopathic pulmonary fibrosis in Japan; 13 patients received nintedanib and pirfenidone and 11 patients received nintedanib alone. The summary of product characteristics states that the benefit/risk of administering nintedanib with pirfenidone cannot be established because of the short duration and small population of the trial.

5 Cost-effectiveness evidence

Model structure

5.1 The company provided a Markov model to assess the cost effectiveness of nintedanib compared with pirfenidone or best supportive care in adults with idiopathic pulmonary fibrosis. The company modelled people with a percent predicted FVC of 50% or more, based on the population in the TOMORROW and INPULSIS phase III clinical trials of nintedanib (although the marketing authorisation does not have any restrictions related to FVC). The model used a lifetime horizon, with a cycle length of 3 months. The company discounted costs and benefits at 3.5% per year and applied a half cycle correction. The company stated that costs were from the perspective of the NHS and Personal Social Services.

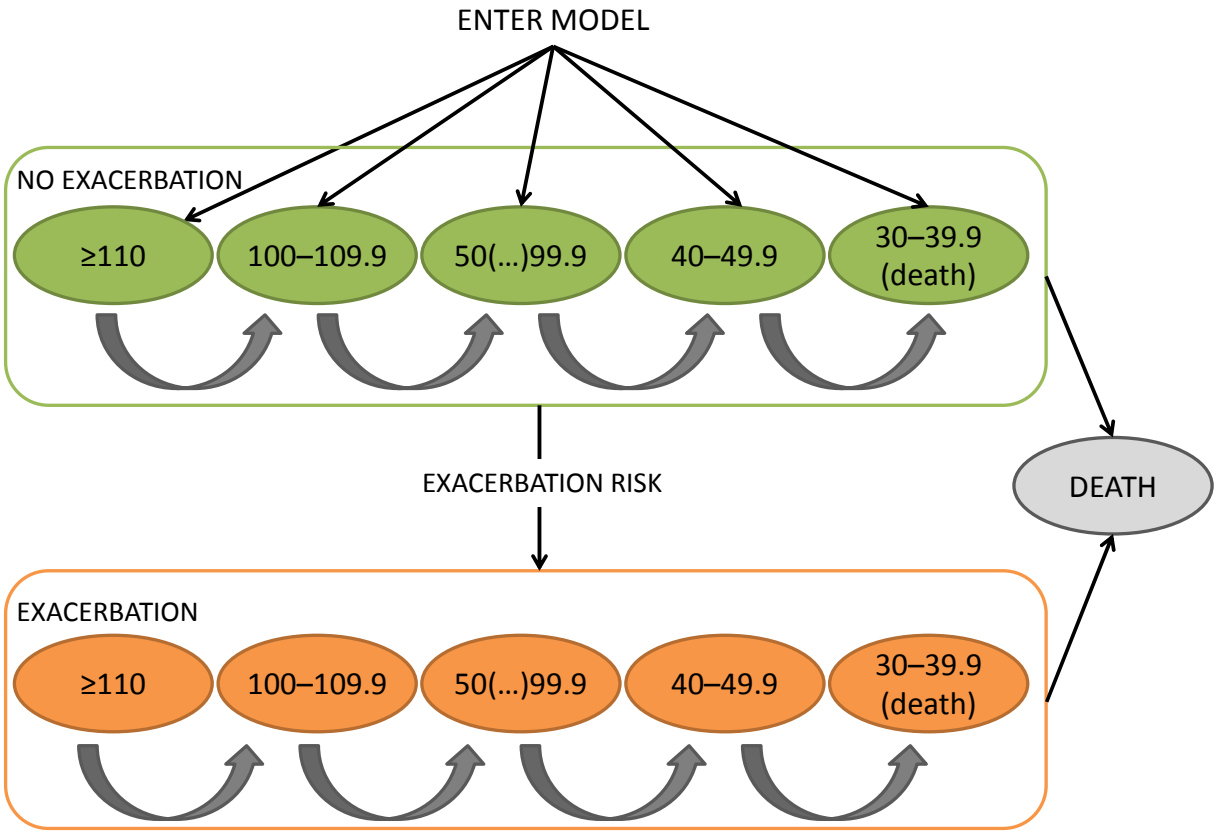
5.2 The health states in the model described the patient condition as a combination of 2 measures: percent predicted FVC (defined in approximately 10 percentage-point increments) and the occurrence of an acute exacerbation (Figure 1). People entered the model at different percent predicted FVC health states without exacerbation. Possible transitions were:

- death

- loss of lung function (representing disease progression, defined as a 10 percentage point decrease in percent predicted FVC)
- exacerbation
- loss of lung function combined with exacerbation
- remain in the same health state.

The company's rationale for choosing reduced FVC as a measure of disease progression was that FVC is a widely used measure of disease and a common endpoint in clinical trials in idiopathic pulmonary fibrosis. Once a person progressed to a health state with a lower percent predicted FVC it was not possible to return to a health state with an improved lung function (that is, a higher percent predicted FVC). Once an exacerbation occurred, people could not move back to a health state without exacerbation. Health states characterised by exacerbation had different health outcomes and costs than states without exacerbation; for example, an exacerbation changed the risk of disease progression (that is, loss of lung function). Because of lack of evidence on the incidence of recurrent events, the company assumed that people who experienced at least 1 exacerbation had the same risk of another exacerbation as people that had never had an exacerbation. Death could occur at any point in the model, or when a person's percent predicted FVC reduced to 39.9% or lower (the company assumed that this value represented an unsustainable level of lung function). Tables 159 and 160 of the company submission summarise the variables and assumptions that the company applied in its model.

Figure 1. Diagram of the company's model (figure 1 on page 19 of the company submission)



The number in each health state refers to percent predicted forced vital capacity (FVC).
The health state labelled "50(…)99.9" indicates 5 separate health states: 50–59.9, 60–69.9, 70–79.9, 80–89.9 and 90–99.9
Transition to the death health state is allowed from all health states in the model (arrows not displayed in the figure).

5.3 The company acknowledged that the population in its model differed from the population that receives pirfenidone in current clinical practice (current NICE guidance recommends pirfenidone only if a person's FVC is between 50% and 80% of its predicted value; in the model FVC was 50% or higher, with no upper limit). The company explained its rationale as follows:

- There are published observational data suggesting that pirfenidone is used in patients with a percent predicted FVC of more than 80%.

- The current NICE guidance for pirfenidone has been scheduled into NICE's work programme for reappraisal, to investigate the use of pirfenidone in patients with a percent predicted FVC greater than 80%.

ERG comments

5.4 The ERG stated that the methods applied in the company's economic analyses were generally appropriate and reported transparently. However, the ERG identified several areas where the company did not sufficiently justify its choice of parameters or explore uncertainty. The ERG conducted exploratory analyses (see sections 5.25–5.28).

5.5 The ERG suggested that the population in the company's model may not represent the population treated in the UK because it includes people with percent predicted FVC more than 80% (accounting for approximately 45% of patients in the model). The ERG noted that clinical advice during the pirfenidone appraisal suggested that this represents disease that is milder than would typically be treated in current UK practice.

5.6 The ERG suggested that a shorter time horizon of 30 years may have been sufficient (the company used 50 years), because the age of the population is generally above 60 years, with a median survival of approximately 3.5 years. It noted that the loglogistic overall survival model predicted that ~0.6% of patients would be alive at 30 years.

5.7 The ERG took clinical advice on the risk of recurrent exacerbation; the advisor acknowledged the lack of evidence in this area, but felt that people who have had one exacerbation were at higher risk of recurrent exacerbation compared to those who have not had any.

Model details

5.8 The company based baseline risk of mortality, disease progression (loss of lung function), and acute exacerbations on the placebo arm of nintedanib clinical trials (INPULSIS and TOMORROW). It based the efficacy of best supportive care on the placebo arms of the INPULSIS

trials. The company applied odds ratios from its network meta-analysis to the baseline risks to estimate the relative effectiveness and safety of nintedanib and pirfenidone compared with best supportive care (using all evidence for overall survival, 'scenario 3' for acute exacerbations and 'scenario 2' for decline in lung function; see section 4.11). To extrapolate data beyond what was available from clinical trials, the company fitted the following parametric models (see pages 164–180 and 268–274 of the company submission and pages 53–62 of the response to clarification for a detailed description and validation of this approach):

- a loglogistic model to estimate overall survival
- an exponential model to estimate the probability of exacerbation and discontinuing medication
- a logistic regression model to predict loss of lung function.

5.9 The company included adverse events in the model if they had a substantial impact on costs and quality-adjusted life years (QALYs), had an incidence of more than 5% or an incidence 1.5 times greater than the comparator arm. On this basis, serious cardiac events and serious GI events were included in the analysis. Gastrointestinal perforation (for nintedanib) and photosensitivity and rash (for pirfenidone) were included. The company assumed that liver enzyme elevations did not cause symptoms and therefore excluded them from the model. The company also excluded diarrhoea, although it was a common adverse event in the INPULSIS trials, because the condition was usually mild to moderate in severity and resulted in less than 5% of patients discontinuating treatment.

5.10 The company assigned utility values to each health state in the model defined by percent predicted FVC using EQ-5D data collected in the INPULSIS clinical trials (Table 5). The model also incorporated reductions in utility values (disutilities) from exacerbations and treatment-related adverse events, based on analysis of the INPULSIS trials and published data. Exacerbation-related and adverse event-related disutilities are

presented on tables 130 and 133 of the company submission, respectively. The company assumed that:

- health-related quality of life gradually declines as percent predicted FVC decreases
- acute exacerbations cause a sudden decline in health-related quality of life for 1 month following the event (a disutility of −0.14), followed by a recovery and smaller and longer-lasting disutility in the subsequent months (−0.078).

Table 5. Health-related quality of life in the company model: EQ-5D utility values for each health state (according to percent predicted forced vital capacity [FVC])

Percent predicted FVC (%)	Mean EQ-5D utility	Standard deviation
90 and above	0.8380	0.1782
80-89.9	0.8105	0.2051
70-79.9	0.7800	0.2244
60-69.9	0.7657	0.2380
50-59.9	0.7387	0.2317
40-49.9	0.6634	0.2552
Key: FVC, forced vital capacity **Source:** table 129 (page 190) of the company submission		

5.11 The company included the following costs in its model: drug treatments (including concomitant medications), adverse events, liver function tests, resource use (for drug acquisition, patient monitoring, treating acute exacerbations and adverse events), oxygen use, exacerbations, and end of life care. NHS reference costs were used to estimate the cost of nintedanib and pirfenidone, and the company assumed that there were no associated administration costs because they are oral treatments. The company did not include any costs for best supportive care, as the efficacy data were based on the placebo arms of the clinical trials (the company stated that the background medications in the trial closely resembled “best supportive care” in idiopathic pulmonary fibrosis). A detailed description of these costs is presented on pages 229–45 of the company submission (and summarised in table 159 of the submission).

ERG comments

5.12 The ERG suggested that the results of the company's cost-effectiveness analysis may have been biased, because the relative effectiveness of nintedanib and pirfenidone was informed by a different scenario analysis from the network meta-analysis for each outcome (see section 4.11 and 4.15). The ERG also suggested that the company did not sufficiently justify its use of results from fixed effect models, when clinical evidence suggested that random effects models performed better for acute exacerbations and serious cardiac events for the all evidence scenario of the network meta-analysis.

5.13 The ERG had concerns about the company using a loglogistic model to extrapolate overall survival data in the base case model (page 92 ERG report). After the company's response to clarification (see pages 53–62 of the response letter), the ERG were satisfied with the company's selection and validation of the loglogistic model (see pages 90–93 of the ERG report).

5.14 The ERG considered the company's approach for including adverse events in the economic model as reasonable and justified. But the ERG suggested that the company model overestimated disutilities for adverse events based on the following:

- adverse events in the company's model last 1 year; the ERG considered that the duration would be significantly shorter than this for gastrointestinal and skin disorders, based on published data
- the company used a disutility of −0.118 for gastrointestinal perforations, instead of the reported value of −0.025 from published data
- data from a long-term open label extension study of the CAPACITY trials (the RECAP study) suggest that the incidence of rash is lower than the estimates in the company model
- the company may have overestimated the incidence of photosensitivity, which the ERG suggested patients can prevent by avoiding sun exposure.

The ERG also noted that the disutility associated with new exacerbations described in the company submission (−0.14) did not match the disutility used in the model (0.0987).

5.15 The ERG had concerns about some of the costs in the model (see pages 103–4 of the ERG report), but the company's exploratory analyses (conducted in response to clarification) showed that the model results were not sensitive to alternative costs.

Company's base-case results and sensitivity analysis

5.16 In the company's comparison of nintedanib and pirfenidone using their list prices, nintedanib dominated pirfenidone (that is, nintedanib costs less and provides greater benefits). Detailed results for this pairwise comparison are presented on table 161 (page 261) of the company submission. The ICER for nintedanib (at list price) compared with best supportive care was estimated at £149,361 per QALY gained (see table 163 on page 263 of the company submission). When the ERG applied the confidential patient access schemes for nintedanib and pirfenidone, [redacted] Table 6 summarises the company's pairwise and incremental cost-effectiveness results.

Table 6. The company's base case cost-effectiveness results

	Total cost	Total QALYs	Inc. cost	Inc. QALYs	Pairwise ICER vs. BSC (£/QALY)	Incremental ICER (£/QALY)
Using list prices for nintedanib and pirfenidone[a]						
BSC	£25,359	3.27				
Pirfenidone	£87,479	3.62	£62,120	0.35	£176,081	dominated by nintedanib
Nintedanib	£85,088	3.67	−£2392	0.05	£149,361	£149,361
Using PAS for nintedanib and pirfenidone[b]						
BSC	£25,359	3.27				
Pirfenidone	[redacted]	3.62	[redacted]	0.35	[redacted]	[redacted]
Nintedanib	[redacted]	3.67	[redacted]	0.05	[redacted]	[redacted]

[a] **Source:** table 165 (page 266) of the company submission
[b] **Source:** table 46 (page 126) of the ERG confidential appendix
Key: BSC, best supportive care; ICER, incremental cost effectiveness ratio; Inc., incremental; PAS, patient access scheme; QALY, quality-adjusted life year

5.17 In the company's probabilistic sensitivity analysis of the base case, nintedanib dominated pirfenidone (see pages 279–285 of the company submission for full details). Results from probabilistic sensitivity analysis indicated that the probability of nintedanib being cost-effective compared with pirfenidone was 60% at any willingness to pay threshold (see figure 73 on page 283 of the company submission).

5.18 The company performed one-way sensitivity analyses to assess the uncertainty around the parameters and structural assumptions in the model (see pages 286–302 of the company submission for full methods and results). These included:

- performing tests around the 95% confidence interval values of all model parameters
- varying baseline risk of dying by changing the extrapolation method (from a loglogistic parametric model to either a Weibull parametric or Gompertz parametric model)

- varying baseline risk of having an acute exacerbation by replacing investigator reported estimates with estimates from an adjudication committee, or excluding the risk of recurrent exacerbation
- varying baseline risk of discontinuing medication by using registry data
- using different adverse event related disutility estimates
- applying different relative risks for patient outcomes based on the results of different scenarios from the network meta-analysis (that is, changing the studies include in the network meta-analysis; see section 4.11).

5.19 When comparing nintedanib and pirfenidone (using list prices for both), nintedanib dominated pirfenidone in all one-way sensitivity analyses. The company noted that the comparison between nintedanib and best supportive care was very sensitive to estimate of mortality risk associated with treatment. When the company applied the lower limit of the odds ratio comparing the effect of nintedanib and best supportive care on, the ICER for nintedanib was approximately £87,000 per QALY gained; by contrast, when the company applied the upper limit of the odds ratio, best supportive care dominated nintedanib. The company explained that this effect was because base case QALY gains were relatively low, meaning that small QALY changes have a big impact on the ICER. Changing the baseline survival risk (by using an alternative method of extrapolation) also had a significant impact on the model results; the ICER increased by approximately £91,000 when the company used a Weibull parametric model and by about £320,000 per QALY when it used a Gompertz parametric model (results are presented in table 191 on page 298 of the company submission).

ERG comments

5.20 The ERG criticised the company's deterministic one-way sensitivity analyses because the company did not adequately demonstrate the effect of varying a parameter for only 1 intervention at a time. The ERG explained that in most one-way sensitivity analyses the company adjusted

the baseline rates for only best supportive care, thereby applying most analyses to both comparator arms simultaneously (because the best supportive care rates are shared by both the nintedanib and pirfenidone arms). Therefore it was not surprising to the ERG that nintedanib remained dominant compared to pirfenidone in most one-way sensitivity analyses.

Company scenarios

5.21 The company also performed 4 scenario analyses:

- **Applying a stopping rule for pirfenidone:** people in the pirfenidone arm discontinued treatment and lost treatment effect when they had a decline in percent predicted FVC of 10% or more in 1 year. This is based on recommendations in current NICE guidance for pirfenidone; it was only relevant to the comparison of nintedanib with pirfenidone.
- **Applying a stopping rule for both pirfenidone and nintedanib:** people in both the pirfenidone and the nintedanib arms discontinued treatment when they had a decline in percent predicted FVC of 10% or more in 1 year. This scenario was performed in response to clarification.
- **Restricting the patient population (the 'ASCEND like' population):** the company restricted the modelled population based on selection criteria in the ASCEND clinical trial of pirfenidone, to better reflect patients treated in current clinical practice. For example, only people with a percent predicted FVC of 50–90% were included (see table 186 of the company submission for full details). This scenario is only relevant to the comparison of nintedanib with pirfenidone.
- **Using direct evidence for nintedanib:** the company used clinical trial data instead of results from its network meta-analysis to estimate the efficacy (survival, exacerbation and loss of lung function) and discontinuation rates. This scenario is only relevant to the comparison of nintedanib with best supportive care.

5.22 In the 3 scenario analyses relevant to the comparison with pirfenidone, nintedanib was no longer dominant. The ICERs for nintedanib compared with pirfenidone in the company scenarios (when list prices were used) ranged from £13,459 per QALY in the restricted 'ASCEND-like' population (full results are presented in table 188 on page 293 of the company submission) to £82,784 per QALY when the stopping rule was applied to people receiving pirfenidone. When the stopping rule was applied to both the nintedanib and pirfenidone arms, the ICER for nintedanib was £17,096 per QALY gained compared with pirfenidone (full results are presented on pages 69–71 of the company response to clarification).

5.23 In the 2 scenario analyses relevant to the comparison of nintedanib with best supportive care, the ICERs increased. Applying the stopping rule to the nintedanib arm increased the ICER for nintedanib by around £10,000 (to £159,829 per QALY gained; full results are presented on pages 69–71 of the company response to clarification). The use of direct evidence for nintedanib had a more substantial impact on the model results, increasing the ICER by around £45,000 (to £194,459 per QALY gained; full details presented on page 297 of the company submission).

ERG comments

5.24 The ERG suggested that the company may have changed too many model inputs in the 'ASCEND-like' population scenario analysis; as well as restricting the population, the company changed the rate of adverse events and replaced odds ratios in the model with relative risks and hazard ratios. The ERG did not consider this population to be fully representative of people treated in clinical practice in the UK.

ERG exploratory analyses

5.25 Because of concerns about the company's deterministic one-way sensitivity analyses (see section 5.20), the ERG performed tests around the 95% confidence interval values of the odds ratios associating nintedanib and the various outcomes. They tested the following model parameters reflecting the association between nintedanib and: overall

survival, risk of exacerbation, loss of lung function, risk of serious cardiac events and risk of serious gastrointestinal events (see table 42 on page 115 of the ERG report for full results). Nintedanib dominated pirfenidone in all of the ERG's sensitivity analyses using list prices, except when applying the upper limit of the odds ratio comparing the effect of nintedanib and pirfenidone on overall survival (1.095 instead of 0.70 as in the base case). This resulted in an ICER for nintedanib of £27,030 per QALY gained, compared with pirfenidone. The ICER for nintedanib compared with best supportive care varied between £143,279 and £155,751 per QALY gained in the ERG's sensitivity analyses using list prices, compared with £149,361 in the company base case. When the ERG applied the confidential patient access schemes for nintedanib and pirfenidone,

[redacted]

5.26 To further explore uncertainty in the model and address methodological concerns, the ERG conducted 5 scenario analyses (Table 7). Scenarios 2 and 3 of the ERG exploratory analyses used all evidence from the company network meta-analysis (and either the random effects or fixed effect model), thereby changing the odds ratios which informed the relative effectiveness of nintedanib and pirfenidone compared with best supportive care (Table 8).

Table 7. ERG scenario analyses

Scenario	Description	Rationale
1	Restricted population: percent predicted FVC = 50–79.0%	section 5.5 & 5.24
2	All evidence from NMA, fixed effect model	section 5.12
3	All evidence from NMA, random effects model	section 5.12
4	Disutility for new exacerbations = −0.14 (0.0987 in company model)	section 5.14
5	Lower disutility and shorter duration for photosensitivity and rash (data from RECAP study)	section 5.14
Key: FVC, forced vital capacity; NMA, network meta-analysis **Source**: pages 116–17 of the ERG report		

Table 8. Odds ratios from the company network meta-analysis used in the company base case model and the ERG exploratory analyses

Comparison with placebo	Median odds ratio (95% CI) used in model		
	Company base case (different NMA scenarios,[a] fixed effect)	ERG scenario 2 (all evidence from NMA, fixed effect)	ERG scenario 3 (all evidence from NMA, random effects)
Overall Survival			
Nintedanib	0.70 (0.45 to 1.10)	0.70 (0.45 to 1.10)	0.70 (0.25 to 2.02)
Pirfenidone	0.70 (0.46 to 1.05)	0.70 (0.46 to 1.05)	0.70 (0.32 to 1.87)
Acute exacerbations			
Nintedanib	0.56 (0.35 to 0.89)	0.56 (0.35 to 0.89)	0.47 (0.01 to 15.96)
Pirfenidone	1.01 (0.22 to 4.50)	0.59 (0.24 to 1.35)	0.37 (0.01 to 4.81)
Loss of lung function			
Nintedanib	0.54 (0.42 to 0.69)	0.54 (0.42 to 0.69)	0.54 (0.11 to 2.70)
Pirfenidone	0.69 (0.47 to 1.00)	0.55 (0.41 to 0.72)	0.54 (0.11 to 2.69)
Serious cardiac events			
Nintedanib	0.92 (0.53 to 1.63)	0.76 (0.45 to 1.27)	0.42 (0 to 21.16)
Pirfenidone	1.27 (0.66 to 2.49)	1.26 (0.65 to 2.49)	1.26 (0 to 459.98)
Serious gastrointestinal events			
Nintedanib	2.35 (1.05 to 5.88)	2.35 (1.05 to 5.88)	3.52 (0.08 to 429.92)
Pirfenidone	0.60 (0.23 to 1.45)	0.60 (0.23 to 1.45)	0.59 (0 to 178.99)
Discontinuation			
Nintedanib	1.42 (1.08 to 1.87)	1.53 (1.13 to 2.08)	1.50 (0.72 to 2.93)
Pirfenidone	1.34 (1.04 to 1.73)	1.73 (1.27 to 2.38)	1.78 (1.09 to 3.37)

[a] The company used the NMA base case (all evidence) for overall survival, serious gastrointestinal events, and discontinuation; NMA scenario 3 for acute exacerbations; NMA scenario 2 for loss of lung function and serious cardiac events
Key: CI, confidence interval; NMA network meta-analysis
Source: tables 49, 55, 61, 72, 78 and 84 of the company submission, tables 24 and 44 of appendix B to the company submission

5.27 The ERG scenarios all resulted in similar estimates; nintedanib dominated pirfenidone and was associated with an ICER greater than £146,000 per QALY compared with best supportive care in all of the ERG's individual scenario analyses when using list prices (Table 9). However, the amount by which nintedanib dominated pirfenidone was substantially reduced in scenarios 2 and 3 (when the ERG used odds ratios based on all evidence from the company network meta-analysis). When the ERG applied the confidential patient access schemes for nintedanib and pirfenidone (Table 10),

5.28 The ERG combined its preferred assumptions (scenarios 1,2,4 and 5) in an alternative base case (Table 9) and noted that the ICERs for nintedanib and pirfenidone compared with best supportive care were very similar to one another (approximately £150,000 per QALY). Pirfenidone was associated with only 0.008 more QALYs than nintedanib. Based on its alternative base case, the ERG suggested that there were no meaningful differences in cost effectiveness between nintedanib and pirfenidone because total costs, total life years and total QALYs were similar. The ERG suggested that the cost effectiveness of nintedanib compared with pirfenidone was largely driven by overall survival, which was approximately equal between the 2 drugs in the model. When the ERG applied the confidential patient access schemes for nintedanib and pirfenidone to its alternative base case (Table 10),

Table 9. Cost-effectiveness results from ERG scenario analyses and alternative base case (using list prices)

Treatment	Total costs	Total QALYs	Inc. costs	Inc. QALYs	Pairwise ICER vs. BSC (£/QALY)	Incremental ICER (£/QALY)
Company base case						
BSC	£25,359	3.27				
Pirfenidone	£87,479	3.62	£62,120	0.35	£176,081	dominated by nintedanib
Nintedanib	£85,088	3.67	−£2392	0.05	£149,361	£149,361
ERG Scenario 1: Restricted population: percent predicted FVC = 50–79.0%						
BSC	£27,960	3.06				
Pirfenidone	£90,164	3.39	£62,204	0.33	£184,829	dominated by nintedanib
Nintedanib	£87,987	3.45	−£2177	0.06	£153,582	£153,582
ERG Scenario 2: All evidence from NMA, fixed effect model						
BSC	£25,359	3.27				
Pirfenidone	£87,205	3.66	£61,666	0.39	£157,460	dominated by nintedanib
Nintedanib	£85,047	3.67	−£2158	0.01	£149,139	£149,139
ERG Scenario 3: All evidence from NMA, random effects model						
BSC	£25,359	3.27				
Pirfenidone	£87,045	3.68	£61,686	0.41	£152,191	dominated by nintedanib
Nintedanib	£84,972	3.68	−£2073	0.00	£146,860	£146,860
ERG Scenario 4: Utility decrements for new exacerbations = 0.14						
BSC	£25,359	3.26				
Pirfenidone	£87,479	3.61	£62,120	0.35	£176,908	dominated by nintedanib
Nintedanib	£85,087	3.66	−£2392	0.05	£148,820	£148,820
ERG Scenario 5: Lower disutility and shorter duration for photosensitivity and rash						
BSC	£25,359	3.27				
Pirfenidone	£87,381	3.64	£62,022	0.37	£168,022	dominated by nintedanib
Nintedanib	£85,087	3.67	−£2294	0.03	£149,361	£149,361
ERG alternative base case: combination of scenarios 1,2,4 and 5 (note order of incremental analysis)						
BSC	£27,960	3.0441				
Nintedanib	£87,941	3.4365	£59,981	0.3924	£152,861	£152,861
Pirfenidone	£89,984	3.4443	£2043	0.0078	£155,000	£263,051

Key: BSC, best supportive care; Inc., incremental; NMA, network meta-analysis; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio
Source: table 165 (page 266) company submission and table 44 (page 118) ERG report

Table 10. Cost-effectiveness results from ERG scenario analyses and alternative base case (using PAS prices for nintedanib and pirfenidone)

Treatment	Total costs	Total QALYs	Inc. costs	Inc. QALYs	Pairwise ICER vs. BSC (£/QALY)	Incremental ICER (£/QALY)
Company base case						
BSC	[redacted]	3.27				
Pirfenidone	[redacted]	3.62	[redacted]	0.35	[redacted]	[redacted]
Nintedanib	[redacted]	3.67	[redacted]	0.05	[redacted]	[redacted]
ERG Scenario 1: Restricted population: percent predicted FVC = 50–79.0%						
BSC	[redacted]	3.06				
Pirfenidone	[redacted]	3.39	[redacted]	0.33	[redacted]	[redacted]
Nintedanib	[redacted]	3.45	[redacted]	0.06	[redacted]	[redacted]
ERG Scenario 2: All evidence from NMA, fixed effect model						
BSC	[redacted]	3.27				
Pirfenidone	[redacted]	3.66	[redacted]	0.39	[redacted]	[redacted]
Nintedanib	[redacted]	3.67	[redacted]	0.01	[redacted]	[redacted]
ERG Scenario 3: All evidence from NMA, random effects model						
BSC	[redacted]	3.27				
Pirfenidone	[redacted]	3.68	[redacted]	0.41	[redacted]	[redacted]
Nintedanib	[redacted]	3.68	[redacted]	0.00	[redacted]	[redacted]
ERG Scenario 4: Utility decrements for new exacerbations = 0.14						
BSC	[redacted]	3.26				
Pirfenidone	[redacted]	3.61	[redacted]	0.35	[redacted]	[redacted]
Nintedanib	[redacted]	3.66	[redacted]	0.05	[redacted]	[redacted]
ERG Scenario 5: Lower disutility and shorter duration for photosensitivity and rash						
BSC	[redacted]	3.27				
Pirfenidone	[redacted]	3.64	[redacted]	0.37	[redacted]	[redacted]
Nintedanib	[redacted]	3.67	[redacted]	0.03	[redacted]	[redacted]
ERG alternative base case: combination of scenarios 1,2,4 and 5 (note order of incremental analysis)						
BSC	£27,960	3.0441				
Nintedanib	[redacted]	3.4365	[redacted]	0.3924	[redacted]	[redacted]
Pirfenidone	[redacted]	3.4443	[redacted]	0.0078	[redacted]	[redacted]

Key: BSC, best supportive care; Inc., incremental; NMA, network meta-analysis; PAS, patient access scheme; QALY, quality-adjusted life year; ICER, incremental cost-effectiveness ratio
Source: table 46, 49 & 50 (pages 126–31) ERG confidential appendix

5.29 The ERG noted inconsistencies between the results of the nintedanib model presented here and the pirfenidone model used in NICE technology appraisal guidance 282. The ERG's alternative base case using the nintedanib model estimated ICERs for both nintedanib and pirfenidone of approximately £150,000 per QALY gained compared with best supportive care (using list prices). Analyses using the confidential patient access schemes for nintedanib and pirfenidone [redacted] the pirfenidone model in TA282 estimated an ICER of £24,000 per QALY compared with best supportive care (including the confidential patient access scheme). The ERG was unable to fully compare the differences between the 2 models because of the confidential data in the pirfenidone appraisal.

5.30 The NICE technical team noted some key differences between the nintedanib model presented here and the pirfenidone company model used in NICE technology appraisal 282. For example, the pirfenidone model:

- was a microsimulation model using individual patient data (the nintedanib model was a cohort model)
- comprised 6 health states (rather than 19): alive (either hospitalised or not hospitalised), death from idiopathic pulmonary fibrosis (either in- or out-of hospital), and death unrelated to idiopathic pulmonary fibrosis (either in- or out-of hospital)
- measured treatment effect based on FVC and the 6-minute walk test distance; it did not include acute exacerbations as an outcome
- estimated utility values by mapping the St George's Respiratory Questionnaire scores predicted by the model onto the EQ-5D (using an algorithm that was estimated using data for patients with chronic obstructive pulmonary disease), rather than using EQ-5D data collected in clinical trials
- did not include disutilities from adverse events

- did not include the pirfenidone phase III trial ASCEND (results were not available at the time of the appraisal).

The Committee stated that the pirfenidone model generated survival rates that were subject to considerable uncertainty; the Committee was cautious in its interpretation of cost-effectiveness estimates produced using these data.

Innovation

5.31 The company provided justifications for considering nintedanib to be innovative:

- Treatment options for idiopathic pulmonary fibrosis are very limited.
- The marketing authorisation for pirfenidone is restricted to people with mild to moderate idiopathic pulmonary fibrosis. Current NICE guidance recommends pirfenidone only if the person has a percent predicted FVC between 50% and 80%. Nintedanib has a marketing authorisation for all adults with idiopathic pulmonary fibrosis, regardless of FVC value.
- Nintedanib has a unique mechanism of action (tyrosine kinase inhibitor) that simultaneously acts on 3 growth factor receptors.
- Nintedanib is taken as a single capsule twice daily, whereas pirfenidone is taken as 3 capsules 3 times daily,

5.32 A number of patient and carer organisation also considered nintedanib to be a new and innovative therapy for idiopathic pulmonary fibrosis, because it has a less burdensome dosing regimen and is associated with fewer adverse events than pirfenidone.

6 Equality issues

6.1 Evidence submissions from patient experts raised 2 potential equality issues for consideration. They noted that current NICE guidance for idiopathic pulmonary fibrosis restricts the use of pirfenidone to people with

FVC of 50–80% of their predicted value, and suggested that the drug is clinically effective in a broader population than this subgroup. They also suggested that NICE guidance should clearly define poor response to treatment.

7 Authors

Sophie Laurenson
Technical Lead

Melinda Goodall
Technical Adviser

with input from the Lead Team (Dr John Pounsford and Matthew Campbell-Hill).

Appendix A: European public assessment report

European Public Assessment Report (EPAR):

http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/003821/WC500182476.pdf

Summary of Product Characteristics (SmPC):

http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003821/WC500182474.pdf

Section 5.1 of the nintedanib SmPC (Pharmacodynamic properties – clinical efficacy)

The following tables from the nintedanib SmPC were used alongside the company submission to inform Table 3 of this document.

Table 2 [nintedanib SmPC]: Annual rate of decline in FVC (mL) in trials INPULSIS-1, INPULSIS-2 and their pooled data - treated set

	INPULSIS-1		INPULSIS-2		INPULSIS-1 and INPULSIS-2 Pooled	
	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily
Number of analysed patients	204	309	219	329	423	638
Rate[1] (SE) of decline over 52 weeks	−239.9 (18.71)	−114.7 (15.33)	−207.3 (19.31)	−113.6 (15.73)	−223.5 (13.45)	−113.6 (10.98)
Comparison vs placebo						
Difference[1]		125.3		93.7		109.9
95% CI		(77.7, 172.8)		(44.8, 142.7)		(75.9, 144.0)
p-value		<0.0001		0.0002		<0.0001

[1] Estimated based on a random coefficient regression model.
CI: confidence interval

Table 3 [nintedanib SmPC]: Proportion of FVC responders at 52 weeks in trials INPULSIS-1, INPULSIS-2 and their pooled data - treated set

	INPULSIS-1		INPULSIS-2		INPULSIS-1 and INPULSIS-2 pooled	
	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily
Number of analysed patients	204	309	219	329	423	638
5% threshold						
Number (%) of FVC responders[1]	78 (38.2)	163 (52.8)	86 (39.3)	175 (53.2)	164 (38.8)	338 (53.0)
Comparison vs placebo						
Odds ratio		1.85		1.79		1.84
95% CI		(1.28, 2.66)		(1.26, 2.55)		(1.43, 2.36)
p-value[2]		0.0010		0.0011		<0.0001
10% threshold						
Number (%) of FVC responders[1]	116 (56.9)	218 (70.6)	140 (63.9)	229 (69.6)	256 (60.5)	447 (70.1)
Comparison vs placebo						
Odds ratio		1.91		1.29		1.58
95% CI		(1.32, 2.79)		(0.89, 1.86)		(1.21, 2.05)
p-value[2]		0.0007		0.1833		0.0007

[1]Responder patients are those with no absolute decline greater than 5% or greater than 10% in FVC % predicted, depending on the threshold and with an FVC evaluation at 52 weeks.
[2]Based on a logistic regression.

Table 4 [nintedanib SmPC]: Frequency of patients with ≥ 10% absolute decline of FVC % predicted or death over 52 weeks and time to progression in trials INPULSIS-1, INPULSIS-2, and their pooled data - treated set

	INPULSIS-1		INPULSIS-2		INPULSIS-1 and INPULSIS-2 pooled	
	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily
Number at risk	204	309	219	329	423	638
Patients with events, N (%)	83 (40.7)	75 (24.3)	92 (42.0)	98 (29.8)	175 (41.4)	173 (27.1)
Comparison vs placebo[1]						
p-value[2]		0.0001		0.0054		<0.0001
Hazard ratio[3]		0.53		0.67		0.60
95% CI		(0.39, 0.72)		(0.51, 0.89)		(0.49, 0.74)

[1] Based on data collected up to 372 days (52 weeks + 7 day margin).
[2] Based on a Log-rank test.
[3] Based on a Cox's regression model.

Table 5 [nintedanib SmPC]: Frequency of patients with acute IPF exacerbations over 52 weeks and time to first exacerbation analysis based on investigator-reported events in trials INPULSIS-1, INPULSIS-2, and their pooled data - treated set

	INPULSIS-1		INPULSIS-2		INPULSIS-1 and INPULSIS-2 pooled	
	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily
Number at risk	204	309	219	329	423	638
Patients with events, N (%)	11 (5.4)	19 (6.1)	21 (9.6)	12 (3.6)	32 (7.6)	31 (4.9)
Comparison vs placebo[1]						
p-value[2]		0.6728		0.0050		0.0823
Hazard ratio[3]		1.15		0.38		0.64
95% CI		(0.54, 2.42)		(0.19, 0.77)		(0.39, 1.05)

[1] Based on data collected up to 372 days (52 weeks + 7 day margin).
[2] Based on a Log-rank test.
[3] Based on a Cox's regression model.

Table 6 [nintedanib SmPC]: All-cause mortality over 52 weeks in trials INPULSIS-1, INPULSIS-2, and their pooled data - treated set

	INPULSIS-1		INPULSIS-2		INPULSIS-1 and INPULSIS-2 Pooled	
	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily	Placebo	Ofev 150 mg twice daily
Number at risk	204	309	219	329	423	638
Patients with events, N (%)	13 (6.4)	13 (4.2)	20 (9.1)	22 (6.7)	33 (7.8)	35 (5.5)
Comparison vs placebo[1]						
p-value[2]		0.2880		0.2995		0.1399
Hazard ratio[3]		0.63		0.74		0.70
95% CI		(0.29, 1.36)		(0.40, 1.35)		(0.43, 1.12)

[1] Based on data collected up to 372 days (52 weeks + 7 day margin).
[2] Based on a Log-rank test.
[3] Based on a Cox's regression model.