Pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282): A Single Technology Appraisal

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Sarah Davis, Senior Lecturer in Health Economics, ScHARR, University of Sheffield, Sheffield, UK
	Rachid Rafia, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Christopher Carroll, Reader in Systematic Review and Evidence Synthesis, ScHARR, University of Sheffield, Sheffield, UK
	Munira Essat, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Jean Sanderson, Research Associate in Statistics, ScHARR, University of Sheffield, Sheffield, UK
	Naila Dracup, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
	Dr Stephen Bianchi, Consultant in Respiratory Medicine, Sheffield Teaching Hospitals NHS Foundation
	Professor David Thickett, Consultant in Respiratory Medicine, University Hospitals of Birmingham NHS Foundation Trust and Chair in Respiratory Medicine, University of Birmingham
Correspondence Author	Sarah Davis, Senior Lecturer in Health Economics, ScHARR, University of Sheffield, Sheffield, UK
Date completed	Date completed (05/03/2016)

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 142/06/02.

ASCEND, CAPACITY 1 & 2 and SP3 at week 52 showed no difference between the pirfenidone and placebo group (OR: 0.90, 95% CI: 0.70 to 1.15, *p*-value not reported).

In the absence of head-to-head RCTs evaluating nintedanib against pirfenidone the company conducted a Bayesian NMA to perform an indirect treatment comparison. NMAs were conducted for 11 outcomes relevant to the decision problem and the results of four of these outcomes (overall survival [OS], PFS, time to treatment discontinuation and acute exacerbations) were used to inform the economic model. Based on the NMA, the treatment effects for pirfenidone were broadly similar to those for nintedanib for all outcomes, with the pairwise treatment effects indicating that neither treatment is statistically significantly more effective.

NMA of safety data indicated that pirfenidone is associated with reduced odds of diarrhoea compared to nintedanib, and increased odds of rash as compared to nintedanib. For discontinuation due to adverse events and serious cardiac adverse events, the treatment effects for pirfenidone are broadly similar to those for nintedanib, with the pairwise treatment effects indicating that neither treatment is associated with more adverse events.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The final selection of three trials (ASCEND, CAPACITY 1 and CAPACITY 2) for the main clinical efficacy review was considered to be appropriate by the ERG, as was the inclusion of the trials from Japan, SP3 and SP2, as supporting evidence. An additional relevant trial was also identified by the ERG and included as supporting evidence: this was a multicenter Chinese trial, which compared pirfenidone plus N-acetylcysteine (NAC) with placebo plus NAC in adult patients with mild or moderate IPF (Huang 2015). The ERG noted that there were between-trial differences across some baseline characteristics in the three key trials (ASCEND, CAPACITY 1 & 2), such as mean FVC or 6MWD at baseline, but subgroup analyses suggested that these and other variables did not influence treatment effect.

Overall, the ERG assessed the potential risk of bias in ASCEND and CAPACITY 1 & 2 to be low across most domains, with the exception of reporting bias and "other bias", which were judged to be "moderate", on account of inconsistency between some of the outcomes and analyses specified in the trial protocols and those presented in the CS, and the possible influence of uncontrolled variables such as rate of disease progression. The SP3, SP2 and Huang *et al.* (2015) trials were at a higher or more unclear risk of bias across many domains than the ASCEND and CAPACITY trials. These trials all evaluated lower doses of pirfenidone, which are licensed in Japan but not in the UK, applied different eligibility criteria and presented noticeable differences from the other three trials in some baseline characteristics of participants.

- the inadequacy of the partitioned survival approach when implementing the stopping rule
- the assumption that treatment effect is constant over the entire model duration
- the estimation of the treatment effect

The ERG further observes that under the company's base-case assumptions, there are discrepancies between the model's prediction of OS for people initiating BSC and the observed trial data for OS in patients who were randomised to placebo. The CS does not comment on these discrepancies and instead focuses on a comparison of the model prediction with registry data for patients receiving BSC, even though the registry data does not match the trial data for people randomised to placebo.

1.6 ERG commentary on the robustness of evidence submitted by the company

The ERG notes the following strengths and weaknesses in the evidence submitted by the company.

1.6.1 Strengths

- The CS reports a generally good quality systematic review of the RCT evidence.
- The three principal RCTs are generally at a low risk of bias.
- Generally, there are no major safety concerns, and some long-term safety evidence is available.
- *Evidence in the model for pirfenidone is based upon long-term data for people included in RECAP.*
- *Results from NMAs are used to inform the relative treatment effects for the comparators.*
- Whilst EQ-5D data were not directly available in the trials, SGRQ data from the trials were mapped onto the EQ-5D using a mapping algorithm developed in people with IPF.

1.6.2 Weaknesses and areas of uncertainty

- There is a moderate risk of reporting bias in the three key RCTs and unclear, moderate or high risk of bias across some domains in the three supporting RCTs.
- There are difficulties in controlling for the rate of disease progression among IPF trial participants, which might moderate outcomes, however the extent of this is unclear.
- The efficacy findings are not consistent across individual trials; one of the key trials reports no statistically significant treatment effect for pirfenidone compared with placebo on the primary outcomes measures relating to FVC or the secondary outcome of PFS.
- Individual trials do not report any statistically significant treatment effect compared to placebo for all-cause mortality; a statistically significant treatment effect is only observed when pooling or meta-analysing studies.

at baseline, FVC >80% at baseline). For the ITT population the ICERs incorporating the PAS ranged from £27,124 to £115,751. For the mild population (percent predicted FVC >80%) the ICERs incorporating the PAS ranged from £31,722 to £186,260. Fort the moderate population (percent predicted FVC 50 – 80%), the ICERs for pirfenidone versus BSC ranged from £27,432 - £104,915 when incorporating the PAS. Results incorporating the PAS for pirfenidone versus nintedanib in the moderate population are presented in the confidential appendix.

A key uncertainty in the company's model concerns the duration of the extrapolation of the treatment effect. As reported in the company's scenario analyses and the ERG's exploratory analyses, truncating the duration over which the treatment effect applies increases the ICERs for pirfenidone versus BSC. A further important limitation in the company's model relates to the implementation of stopping rules for pirfenidone and nintedanib. The inclusion of the stopping rule in the economic model lacks validity in that the modelled stopping rule impacts on costs but not health outcomes. The ERG considers that the analysis incorporating the stopping rule as implemented in the economic model provides a lower bound of the plausible ICER (i.e. most optimistic scenario).

4.1.5 Evidence synthesis

The synthesis for the review of clinical efficacy was a basic descriptive summary of the evidence from the five pirfenidone trials for the following outcomes: change from baseline in percent predicted FVC; all-cause and IPF-related mortality; PFS; acute exacerbation; hospitalisation; changes from baseline in 6MWD, the UCSD SOBQ and the SGRQ. This approach to evidence synthesis was neither described nor justified in the CS.[4]

Meta-analyses using both fixed and random effects models comparing pirfenidone with placebo were performed for selected outcomes and time-points, based on available trial data, and the methods used were described in the CS[4] (Section 4.9 and Appendix 9). Data were combined from CAPACITY 1 & 2[33, 36, 49] and ASCEND[33, 34] using the UK licence dosage (2,403mg/day) and from SP3[38] which uses a lower dosage (1,800 mg/day which is the licensed dose in Japan but is not a licensed dose in the UK). The company considered this to be appropriate as the dose by weight would be similar for all studies given the lower body weight of the Japanese population compared with the North American and European population. An NMA comparing effects across all treatments was also performed by the company. This is critiqued in Sections 4.6 and 4.7 of this report.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

4.2.1 Review of clinical efficacy (relevant pirfenidone RCT evidence)

The CS[4] provides a detailed description of trials identified by the company as satisfying the requirements of the final NICE scope,[3] i.e. pirfenidone compared with placebo or nintedanib (see Table 5). No trial compared pirfenidone with nintedanib. Five RCTs compared pirfenidone at various doses with placebo: ASCEND (Phase III),[34] CAPACITY 1 & CAPACITY 2 (Phase III),[49] SP3 (Phase III),[38] and SP2 (Phase II).[39] Three trials were international and multicentre (ASCEND[34] and CAPACITY 1 & 2[33, 36, 49]), although only CAPACITY 2 included any UK centres[35] (three of 110 centres across both CAPACITY trials).[49] The inclusion criteria in all three trials were adult patients with mild or moderate IPF based on percentage predicted FVC of ≥50% (in ASCEND[34] this had an upper limit of ≤90%). Two trials were conducted exclusively in Japan (SP3[38] and SP2[39]) and did not report baseline levels of FVC or VC. One trial was conducted in China and evaluated pirfenidone in combination with N-acetylcysteine (NAC). The trials varied in criteria relating to lung function, concomitant medications permitted for IPF, and the investigated doses of pirfenidone (for the purposes of this appraisal, ASCEND,[34] and CAPACITY 1 & 2[49] all evaluated the efficacy of the licensed dose of 2,403mg/d; the SP2,[39] SP3[38] and the Huang *et al.* [48] trial evaluated lower doses; the applicability of these lower doses to clinical practice in England and Wales is unclear.

Table 1: Characteristics of included pirfenidone RCTs (reproduced in part from CS,[4] Tables 10 and 15, pages 59 and 82)

Trial No. of patients	Location	Inclusion criteria		Exclusion criteria	Intervention and co-interventions (No. of patients)	Comparator (No. of patients)	Follow-up
		IPF diagnosis	Lung function parameters	Patient factors			
ASCEND (PIPF-016)[33, 34] n=555	International multi-centre	– Confident clinical and radiographic diagnosis of IPF, confirmed centrally with diagnosis of IPF >6 months but <48 months. – No improvement of IPF in preceding year. –	– FVC (% predicted value) 50-90% – DLco 30-90% – 6MWT ≥150 m	– Abnormal lab parameters – Obstructive airway disease – History of unstable /deteriorating cardiac or pulmonary disease – History of severe hepatic impairment/ end-stage liver disease/end-stage renal disease requiring dialysis	Pirfenidone 2,403mg/day (n=278) Concomitant treatment with any investigational drug for the treatment of IPF was prohibited. However, concomitant medications used in the treatment of IPF were permitted if given for a non-IPF indication and there was no clinically acceptable alternative.	Placebo (n=277)	52 weeks
CAPACITY 1 (PIPF-006)[36, 49] n=344	International multi-centre	– Confident clinical and radiographic diagnosis of IPF, confirmed locally (diagnosis previous 48 months) – No improvement of IPF in preceding year	– FVC (% predicted value) ≥ 50% – DLco ≥35% – FVC or DLco ≤90% – 6MWT ≥150 m		Pirfenidone 2,403mg/day (n=171) Concomitant treatments for IPF were prohibited, with exceptions of short courses of azathioprine, cyclophosphamide, corticosteroids, or acetylcysteine for protocol-defined acute exacerbation of IPF, acute respiratory decompensation, or progression of disease.	Placebo (n=173)	72 weeks
CAPACITY 2 (PIPF-004)[35, 49] n=435	International multi-centre		– FVC (% predicted value) ≥50% – DLco ≥35% – FVC or DLco ≤90%		Pirfenidone 2,403mg/day (n=174) Pirfenidone 1,197mg/day (n=87) As CAPACITY 1	Placebo (n=174)	72 weeks

IPF: Idiopathic Pulmonary Fibrosis; FVC: Forced Vital Capacity; DLco: Diffusing capacity of the lungs for carbon monoxide; 6MWT: 6-minute walking test

The exclusion of certain patients otherwise eligible for pirfenidone, based on co-morbidities, such as obstructive airways disease, must also be taken into account when judging the generalisability of the trials' findings.

The outcomes reported in the CS[4] are generally consistent with those that are listed in the final NICE scope.[3] The ASCEND,[34] CAPACITY[49] and Huang *et al*[48] trials use change from baseline in percent predicted FVC as an endpoint, while SP3[38] and SP2[39] use VC. The CS states that the decision to use VC in the SP3[38] and SP2[39] trials was dictated by the ATS international consensus statement published in 2000, which recommended measurement of VC.[54] The CS[4] did not state when the recommended measurement changed to FVC or provide any reference to substantiate the change. The CS[4] states that VC and FVC should be treated as comparable endpoints as there is little difference between VC and FVC in subjects without obstructive pathology. Whilst the clinical advisors to the ERG agreed with this statement, the ERG noted that the exclusion criteria for SP3[38] were not as explicit regarding the exclusion of patients with emphysema as the exclusion criteria for the other pirfenidone trials. Therefore, the ERG considers that the synthesis of VC data from SP3[38] with FVC data from the ASCEND[34] and CAPACITY trials[49] is questionable.

The outcomes listed in the trial protocols publicly-available from the clinical trials register (https://clinicaltrials.gov/ct2/home) are not entirely consistent with those reported in the CS.[4] For example, the principal efficacy outcome of "percent predicted FVC or death" does not appear in any protocol as a trial outcome but appears to describe the method used by the company in order to impute a FVC measurement for patients who have died (see clarification response[10], questions, A11 and A13). Furthermore, neither of the secondary outcomes of "treatment-emergent IPF-related mortality" nor the SGRQ was listed in the protocols, but both appear *post hoc* as outcomes in the CS[4] (as well as in the ASCEND[34] publication, but not in the CAPACITY trials' publication,[49] see Table 7).

The following outcome was listed in protocols but was not reported in the results for the CAPACITY 1 & 2[49] and SP3[38] trials: Change in Worst Oxygen Saturation by Pulse Oximetry (SpO_2) measurement observed during the 6-Minute Walk Test. The CAPACITY trial protocols[35, 37, 51] also listed lung transplantation as a secondary outcome, but this is not included as an outcome in the CS[4] (pages 53 and 66). The CS[4] lists fibrosis by use of high resolution computed tomography (HRCT) (see CS, Table 12, page 68) as an outcome, but this only appears to be used as a diagnostic criterion for IPF or as part of the definitions of acute exacerbations (see CS,[4] pages 104-105).

Definitions of outcomes are first provided under the trial results section of the CS[4] (Section 4.7, pages 90-113). The outcomes, and the definitions applied in each of the trials, taken from the CS and the

Table 2: Primary efficacy outcomes and measures in ASCEND, CAPACITY 1, CAPACITY 2, SP3 and SP2

Outcome	ASCEND[33, 34]	CAPACITY 1[36, 49]	CAPACITY 2[35, 49]	SP3[38]	SP2[39]
Protocol-listed outcome	Change in percent predicted FVC from baseline to week 52†	Mean and absolute change in percent predicted FVC from baseline to week 72		No protocols available	
Reported outcomes	Change in percent predicted FVC from baseline to week 52	Change in percent predicted FVC **or death** from baseline to week 52		Change in VC from baseline to week 52	Change in the lowest SpO_2 during 6MWT. Full definition given in Azuma, page 1041 Change in VC from baseline was listed as a secondary outcome
	Categorical decline of ≥10% in percent predicted FVC or death	Categorical decline of ≥10% in percent predicted FVC. This was listed as a secondary outcome in the protocols and publication, defined as "Categorical Assessment of Absolute Change in Percent Predicted Forced Vital Capacity (FVC) based on the change in baseline percent predicted FVC at week 72, patients were assigned to 1 of 5 categories: mild decline (<10% but >=0% decline), moderate decline (<20% but >=10% decline), severe decline (>=20% decline), mild improvement (>0% but <10% improvement), or moderate improvement (>=10% improvement). Those who died or had a lung transplant before week 72 were included in the severe decline category. The results indicate the number of patients who experienced a Categorical Change in Percent Predicted Forced Vital Capacity" [35, 36, 49]			
Magnitude of treatment effect	The magnitude of the treatment effect was estimated by comparing the distribution of patients in the pirfenidone group with those in the placebo group across two thresholds of change at week 52: an absolute decline of 10 percentage points in the percentage of the predicted **FVC or death**, or no decline in the percentage of the predicted FVC (King 2014, page 2085)[34]	Estimated by use of differences in treatment group means and categorical change in FVC (page 1763, Noble 2011)[49]			

† This outcome was not reported in the ASCEND publication; the data were only made available by Roche in the CS,[4] Table 20 and pages 93-94.

Table 3: Secondary efficacy outcomes and measures in ASCEND, CAPACITY 1, CAPACITY 2, SP3 and SP2

Outcome	ASCEND[33, 34]	CAPACITY 1[36, 49]	CAPACITY 2[35, 49]	SP3[38]	SP2[39]
All-cause mortality	Yes				
IPF-related death	Yes	Yes*		No	
Treatment-emergent IPF mortality	Yes. Defined as death occurring after randomisation and within 28 days of the last dose of the study drug (CS, page 96).[4] Listed only in the ASCEND NEJM protocol but reported for all mortality outcomes in ASCEND publication and separately, applied and not-applied, to all-cause and IPF-related mortality in the CAPACITY publication: appears to be a *post hoc* outcome measure.				
Progression-free Survival (PFS)	Defined in the CS (page 99)[4] as a confirmed ≥10% decline from baseline in %FVC, confirmed ≥50 m decline from baseline in 6MWD, or death	PFS is defined as the first occurrence of a 10% absolute decline from baseline in percent predicted Forced Vital Capacity, a 15% absolute decline from baseline in percent predicted hemoglobin(Hgb)-corrected carbon monoxide diffusing capacity (DLco), or death		Defined as VC decline of ≥10% or death. When the VC data could not be obtained due to worsening of respiratory symptoms, including acute exacerbation, the case was also classified as disease progression. (Taniguchi, page 822)[38]	No
Acute Exacerbations	Identified via a *post hoc* analysis of adverse events based on the MedDRA lower level term "acute exacerbation of IPF".(CS, page 104)	Definition not provided in clinical trials register protocols (where it is reported only as part of a composite measure*). CS (page 104)[4] defines this outcome as requiring all of the following within a 4-week interval: Worsening of PaO_2 (≥8 mm Hg drop from the most recent value); clinically significant worsening of dyspnoea; new, superimposed ground-glass opacities on HRCT in one or more lobes; all other cardiac, thromboembolic, aspiration, infectious processes ruled out		†Definition not provided in protocols. CS (page 104)[4] defines this outcome as requiring all of the following within a month: increase in dyspnoea; new, ground-glass opacities on HRCT in addition to previous honeycomb lesion; all oxygen partial pressure in resting arterial blood (PaO2) is lower by more than 10 Torr than previous one; exclusion of obvious causes, such as infection, pneumothorax, cancer, pulmonary embolism or congestive heart failure; the serum levels of CRP, LDH are usually elevated as well as serum markers of interstitial pneumonias, such as KL-6, Sp-A or Sp-D	†Definition not provided in protocols. CS (page 104)[4] defines this outcome as requiring all of the following: worsening, otherwise unexplained clinical features within 1 month; progression of dyspnoea over a few days to less than 5 weeks; new radiographic/HRCT parenchymal abnormalities without pneumothorax or pleural effusion (e.g., new, superimposed ground-glass opacities); a decrease in the PaO_2 by 10 mm Hg or more; exclusion of apparent infection based on absence of Aspergillus and pneumococcus antibodies in blood, urine for Legionella pneumophila, and sputum cultures

4.2.2 Results

Participants' baseline characteristics

More than 620 participants received the licensed 2,403mg/day dose during the three international RCTs compared with more than 620 control patients who received placebo in these trials. Another 322 participants received lower doses of pirfenidone in the CAPACITY 2,[49] SP2[39] and SP3[38] trials.

The final selection of three trials (ASCEND,[34] CAPACITY 1[36, 49] and CAPACITY 2[35, 49]) for the main clinical efficacy review was considered to be appropriate by the ERG. However, there are some between-trial differences across some baseline characteristics (see Table 8). The ASCEND trial[34] participants had a lower mean percentage predicted FVC (range across arms of 67.8-68.6) than the CAPACITY trials[49] (range across arms of 73.1-76.4) and lower pre-enrollment corticosteroid use (range across arms of 0.7%-2.2%) than the CAPACITY trials[49] (range across arms of 5.2%-12.9%). CAPACITY 1[49] participants had a lower mean 6MWD (range across arms of 378.0-399.1) than in ASCEND[34] and CAPACITY 2[35] (range across arms of 410.0-420.7), and there was a relatively lower proportion of patients in CAPACITY 2[35] requiring supplemental oxygen use (range across arms 14.0%-17.0%) than in ASCEND[34] and CAPACITY 1[49] (range across arms of 27.4%-28.1%). All of these variables, with the exception of corticosteroid use, are accepted potential treatment effect modifiers and therefore were the subject of subgroup analyses in the CS,[4] (Section 4.8, pages 114-117).

The ERG considers the relevance of the smaller SP3[38] and SP2[39] trials, which were conducted exclusively in Japan, to be more questionable. These trials evaluate lower doses of pirfenidone which are licensed in Japan but not in the UK, apply different eligibility criteria and present noticeable differences from the other three trials in some baseline characteristics of participants (see Table 9), for example, higher proportions of male participants (range across arms of 78%-94% for SP2[39] and SP3[38] compared with 68%-80% for ASCEND[33, 34] and CAPACITY 1 and 2[49]) and smokers (60%-86% compared with 58%-66%); higher mean percentages of predicted DLco compared with ASCEND[34] and the CAPACITY trials[49] (52.1-57.7 compared with 43.7-47.8), lower trial corticosteroid use (SP3[38] only, 4.8-10.9 compared with 21.0-36.5 in the ASCEND[34] and CAPACITY trials[49]), and smaller proportions having received surgical lung biopsies (21.0%-29.1% compared with 28.5%-55%, see Table 8).

Baseline data from participants on patient-reported outcome measures, such as the SGRQ and UCSD SOBQ, were not reported in the CS.[4]

The Huang *et al.* trial[48] comparing pirfenidone plus NAC with placebo plus NAC reported comparability between arms across all baseline characteristics except for smoking status.[48]

Table 4: Categorical analysis of change from baseline in percent predicted FVC or death (reproduced from CS,[4] Table 18)

Study	Time point	Treatment group	Decline ≥10% FVC or death, n (%)	No decline* in FVC, n (%)	*p*-value†
ASCEND[34]	52 weeks	PFN 2,403mg/day (N=278)	46 (16.5)	63 (22.7)	*p*<0.000001
		PBO (N=277)	88 (31.8)	27 (9.7)	
CAPACITY 1[49] §	72 weeks	PFN 2,403mg/day (N=171)	39 (22.8)	44 (25.8)	*p*=0.440
		PBO (N=173)	46 (26.6)	38 (22.0)	
CAPACITY 2[49] §	72 weeks	PFN 2,403mg/day (N=174)	35 (20.1)	42 (24.1)	*p*=0.001
		PBO (N=174)	60 (34.5)	24 (13.8)	
Pooled CAPACITY 1 & 2[49]	72 weeks	PFN 2,403mg/day (N=345)	74 (21)	86 (24.9)	*p*=0.003
		PBO (N=347)	106 (31)	62 (17.9)	

PFN: pirfenidone; PBO: placebo
** CAPACITY trials data not reported in original publication (Noble 2011[49]) but taken from respective CSRs*
†Rank ANCOVA (pirfenidone 2,403mg/day vs placebo). It is unclear if this p value relates to the "Decline or death" or the "No decline" comparison: the numbers in the CS, Table 18 refer to the "No decline" comparison in ASCEND (King 2014[34]), but the "Decline or death" comparison for the CAPACITY trials (Noble 2011[49])
§ Note: these data are from the original publication (Noble 2011[49]), which only reports decline of >10% FVC and not decline of >10% or death

A pooled analysis of ASCEND[34] (week 52) and CAPACITY 1 & 2[49] (week 48) reported a statistically significant difference in favour of pirfenidone compared with placebo in terms of those who had experienced a decline in FVC by ≥10% or had died (absolute difference: 10.0 [95% CI not reported], *p*<0.003), and reported a higher proportion of patients with no decline in percent predicted FVC for pirfenidone compared with placebo (24.9% versus 17.9%, *p*-value not reported). This analysis is described as "pre-specified" in the CS[4] (page 91), but this is inaccurate: there is no reference to this analysis for this outcome in the any of the ASCEND protocols,[33, 55] which only refer to pooling of these trials for mortality (see Section 5.4.2.3.2 in the protocols). The protocol that accompanied the ASCEND publication (Section 13.2, page 29) stated that, *"The clinical study protocol (dated 16 March 2011, section 5.4.2.1) describes a supportive analysis of FVC as the change from baseline to Week 52 in FVC volume (in mL) … A categorical analysis of relative change from baseline has been added"*.[55]

The ASCEND trial[34] reported that at 52 weeks there were fewer overall deaths and TE IPF-related deaths in the pirfenidone group than the placebo group, but these differences were not statistically significant (*p*=0.105 and *p*=0.226 respectively).

In the pooled analysis of CAPACITY 1 & 2[49] at 52 weeks, there were fewer overall deaths and TE IPF-related deaths in the pirfenidone groups compared with the placebo groups and this difference was statistically significant in both groups (p=0.047 and p=0.012 respectively).

In the pooled analysis of CAPACITY 1 & 2[49] at 72 weeks, there were fewer overall deaths and TE IPF-related deaths in the pirfenidone groups compared with the placebo groups. Overall, there was a 23% reduction in all-cause mortality versus placebo among patients treated with pirfenidone 2,403mg/day (HR=0.77; 95% CI: 0.47 to 1.28; *p*=0.315), a 38% reduction in IPF-related mortality (HR=0.62; 95% CI: 0.35 to 1.13; *p*=0.117) and a 35% reduction in TE all-cause mortality (HR=0.65; 95% CI: 0.36 to 1.16; *p*=0.141). However, none of these differences were statistically significant.

For TE IPF-related mortality, the HR between the pirfenidone and placebo groups at week 72 also favoured pirfenidone and was statistically significant (HR=0.48; 95% CI: 0.24 to 0.95; *p*=0.03, see Table 14).

There appears to be a markedly increased rate of mortality for the CAPACITY trials[49] between the data reported in the CS[4] for 52 weeks (Table 23, page 97) and the data reported in the publication for 72 weeks.[49] There is a substantial increase in all-cause mortality in the pirfenidone group, from 11 at 52 weeks to 27 at 72 weeks, compared with a much smaller increase in the placebo group from 22 at 52 weeks to 34 at 72 weeks (the *p*-values for the differences between groups are 0.047 and 0.315 for 52 weeks and 72 weeks, respectively). (see Table 14). In the same way, TE IPF-related mortality in the pirfenidone group increases from 4 deaths at 52 weeks to 12 deaths at 72 weeks in the pirfenidone group, and from 15 at 52 weeks to 25 at 72 weeks in the placebo group (*p*-values for the differences between groups are 0.012 and 0.030 for 52 and 72 weeks, respectively). No explanation is provided in the CS[4] for these relative increases in rates of mortality, particularly for the pirfenidone groups, between weeks 52 and 72 in the CAPACITY trials.[49]

In the pooled analysis of the data from 52 weeks for ASCEND[34] and CAPACITY 1 & 2[49] (required by the Food and Drug Administration (FDA)[57] and finalised as an analysis in the Statistical Analysis Plan only on 1st January 2014, according to the company's clarification response[10] (question A22), there were significantly fewer overall deaths (*p*=0.011) and TE IPF-related deaths (*p*=0.006) in the pirfenidone groups compared with the placebo groups.

Table 5: Post hoc analysis of data on hospitalisations in CAPACITY 1 & 2 (reproduced from CS, Table 28)[4]

Study arm	CAPACITY 1[49]		CAPACITY 2[49]		Pooled	
	PFN n=171	PBO n=173	PFN n=174	PBO n=174	PFN n=345	PBO n=347
Respiratory hospitalisations (RH)						
Number of patients with at least 1 RH	22 (12.9%)	23 (16.7%)	29 (16.7%)	29 (16.7%)	51 (14.8%)	52 (15.0%)
Number of RH	31	37	34	40	65	77
Mean length of RH (days)	8.5	17.3	7.6	12.1	8.0	14.6
Total number of days in hospital	264	640	259	484	522	1124
Average number of NRH days per patient	1.5	3.7	1.5	2.8	1.5	3.2
Non-respiratory hospitalisations (NRH)						
Number of patients with at least 1 NRH	37 (21.6%)	25 (14.5%)	35 (20.1%)	31 (17.8%)	72 (20.9%)	56 (16.1%)
Number of NRH	48	31	38	42	86	73
Mean length of NRH (days)	10.1	20.8	7.2	16.0	8.8	8.0
Total number of days in hospital	485	645	274	672	758	1317
Average number of NRH days per patient	2.8	3.7	1.6	3.9	2.2	3.8

PFN: pirfenidone 2,403mg/d; PBO: placebo

In SP2,[39] five patients in the placebo arm and none in the pirfenidone treatment were hospitalised due to exacerbations (Azuma 2005[39]). The company did not conduct a meta-analysis as data were only available for the CAPACITY trials.

4.2.2.6 Patient-Reported Outcomes (Quality of Life)

University of San Diego (UCSD) Shortness of Breath Questionnaire (SOBQ)

The ASCEND[34] and CAPACITY trials[39] reported this outcome. The CS[4] states (pages 111 and 112) that the SOBQ can be used to formulate clinically relevant inferences about IPF patients; that the total score in this questionnaire increases with increased dyspnoea, and an increment of 20 points is considered a clinically relevant threshold based on estimates of the minimal important difference for the USCD SOBQ that range from 5-11.[31] In ASCEND,[34] the proportion of patients with ≥20 point increase in shortness of breath as measured by SOBQ at week 52 was smaller in patients receiving pirfenidone than in those receiving placebo, but this difference was not statistically significant (p=0.1577, see Table 17).

Table 6: Serious treatment-emergent adverse events reported by ≥2 patients in CAPACITY 1 & 2 at 72 weeks[49]

	CAPACITY 1[49]		CAPACITY 2[49]		
Adverse event, n (%)	**PFN 2,403mg/d (n=171)**	**PBO (n=173)**	**PFN 2,403mg/d (n=174)**	**PFN 1,197mg/d (n=87)**	**PBO (n=174)**
Pneumonia	7 (4.1)	7 (4.0)	4 (2.3)	3 (3.4)	6 (3.4)
Respiratory failure	4 (2.3)	6 (3.5)	2 (1.1)	3 (3.4)	2 (1.1)
Angina pectoris			2 (1.1)	2 (2.3)	1 (0.6)
Atrial fibrillation	2 (1.1)	1 (0.6)	1 (0.6)	3 (3.4)	1 (0.6)
Coronary artery disease	6 (3.5)	0 (0)	0	3 (3.4)	2 (1.1)
Acute renal failure	2 (1.2)	2 (1.2)	1 (0.6)	2 (2.3)	0 (0)
Fall	2 (1.2)	1 (0.6)			
Hypotension	2 (1.2)	1 (0.6)			
Colitis	2 (1.2)	0 (0)			
Hip fracture	2 (1.2)	0 (0)			
Prostate cancer	2 (1.6)*	0 (0)			
Interverterbral disc profusion	2 (1.2)	0 (0)			
Liver test function abnormal	2 (1.2)	0 (0)			
Nephrolithiasis	2 (1.2)	0 (0)			
Sick sinus syndrome	2 (1.2)	0 (0)			
Pneumothorax			3 (1.7)	2 (2.3)	0
Pulmonary embolism			1 (0.6)	3 (3.4)	1 (0.6)
Syncope			3 (1.7)	1 (1.1)	1 (0.6)
Chest pain			3 (1.7)	0	0
Bladder cancer			2 (1.1)	0	0
Gastroesophageal reflux disease			2 (1.1)	0	0
Bronchitis	0 (0)	5 (2.9)†	2 (1.1)		2 (1.1)
Lobar pneumonia	-	-	2 (1.1)		2 (1.1)
Non-cardiac chest pain	-	-	2 (1.1)		2 (1.1)
Myocardial infarction	-	-	0 (0)		4 (2.3)†

* Male patients only †p<0.05

Adverse events leading to discontinuation of treatment

In the ASCEND trial,[34] the proportion of patients discontinuing treatment due to an AE was 14.4% (n=40) in the pirfenidone group and 10.8% (n=30) in the placebo group. The most common AE leading to treatment discontinuation was worsening IPF (1.1% [n=3] in the pirfenidone group versus 5.4% [n=15] in the placebo group), but again the caveats should be noted regarding the categorisation of this event as a safety outcome. The only other AEs leading to discontinuation of treatment in ≥1% of patients in the pirfenidone group were elevated hepatic enzymes levels, pneumonia, rash and decreased weight, which each occurred in 3 patients (1.1%).

In the CAPACITY trials,[49] treatment was discontinued due to AEs in 15% (n=51) of 345 patients in the pooled pirfenidone 2,403mg/day group compared with 9% (n=30) of 347 patients in the placebo group. The most common AE leading to discontinuation was worsening of IPF (3% in both groups). The other AEs leading to discontinuation of treatment in ≥1% of patients in the pirfenidone group were provided by the company in response to a request for clarification from the ERG (see clarification response addendum,[30] question A24). In CAPACITY 1,[49] these were elevated IPF (2.3% in each arm), photosensitivity, rash and respiratory failure, which each occurred in 2 patients (1.2%) in the pirfenidone trial arm but not at all in the placebo arm. In CAPACITY 2,[49] for the 2,403mg per day dose, these were elevated IPF (1.1% for pirfenidone versus 1.7% for placebo), bladder cancer (1.1% vs 0%), nausea (2.3% versus 0%) and rash (1.7% versus 0%). The following substantial laboratory abnormalities (Grade 4 or a shift of 3 grades e.g. from 0 to 3) occurred more frequently in the CAPACITY 1 and 2 pooled pirfenidone 2,403mg/day group compared with placebo: hyperglycaemia (1% [n=4] versus <1% [n=3], respectively); hyponatraemia (1% [n=5] versus 0%); hypophosphatemia (2% [n=6] versus <1% [n=3]); and lymphopenia (1% [n=5] versus 0). However, none were associated with clinically significant consequences. More patients in the pooled pirfenidone-treated group than in the pooled placebo group had elevations in alanine aminotransferase and aspartate aminotransferase of more than 3x the upper limit of normal (4% [n=14] versus <1% [n=2]). However, all reports were reversible and without clinical sequelae.

SP2[39] reported that 11 patients discontinued pirfenidone treatment, compared with 2 patients in the placebo arm, due to AEs.[39] The CS (page 172) stated that skin photosensitivity was the AE that was principally responsible for discontinuing or reducing pirfenidone dose; full data on AE discontinuations were provided in the publication:[39] the principal AEs affecting discontinuation from pirfenidone treatment were: photosensitivity (n=5); vomiting (n=1); fever (n=1); abnormality of hepatic function (n=1); dizziness (n=1); facial paralysis (n=1) and hepatoma (n=1). There were no instances of any of these events in the placebo arm.

Studies included in NMA

The company's systematic review identified 10 RCTs of reasonable methodological quality that compared pirfenidone, nintedanib, NAC, or triple therapy with placebo in patients with IPF. However, the company excluded two of the trials; SP2[39] (pirfenidone) and IFIGENIA[70, 71] (double and triple therapy) from the NMA. INFIGENIA[70, 71] was excluded from the NMA as the trial compares double and triple therapy, which are not comparators of interest for this appraisal. SP2[39] was excluded from the NMA as it was considered as an outlier by the NICE Appraisal Committee for the review of nintedanib (TA379)[12] and there was no useable data at one year as the trial was stopped early at 36 weeks. In addition, a non-valid primary end point, SpO_2, was used.

A total of eight studies were included in the company's NMA: ASCEND[34] (pirfenidone), CAPACITY 1[49] (pirfenidone), CAPACITY 2[49] (pirfenidone), SP3[38] (pirfenidone), INPULSIS 1[72] (nintedanib), INPULSIS 2[72] (nintedanib), TOMORROW[73] (nintedanib) and PANTHER[74, 75] (NAC and triple therapy). However, not all trials presented outcome data that could contribute to each NMA for all outcomes.

The ERG notes that although not in the final NICE scope,[3] the evidence network includes NAC and triple therapy. The trials of comparators contributing data to the NMA were all placebo-controlled RCTs and therefore all comparisons were made with placebo (see Figure 26). The ERG therefore believes that PANTHER[74, 75] has little influence on the NMA results for nintedanib and pirfenidone, and therefore data from PANTHER[74, 75] have been excluded from the additional analyses performed by the ERG in Section 4.8. In this section, only data from the trials of relevance to the decision problem are summarised.

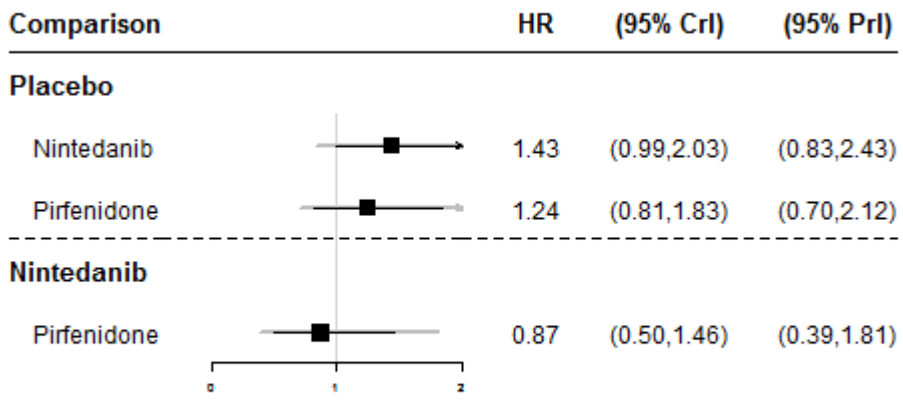
A summary of the design and study characteristics of the studies included in the NMA is provided in Table 34.

Table 7: Reported outcomes and definitions adapted from CS,[4] (including response from clarification question A14, and A17 and A32)[10]

Outcome	ASCEND[34]	*CAPACITY 1 & 2[49]	SP3[38]	INPULSIS 1&2[72]	TOMORROW[73]	PANTHER[74]
Study duration**	52 weeks	72 weeks	52 weeks	52 weeks	52 weeks	60 weeks (NAC), 32 weeks (Triple therapy)
Lung function						
Change in percent predicted FVC	Yes	Yes	Reported change in % predicted VC	Yes	Yes	Yes (NAC only)
Change from baseline in FVC (L)	Yes	Yes	Reported change from baseline in VC (L)	Yes	Yes	Yes
Categorical decline of ≥ 10% in percent predicted FVC	Yes	Yes	No	Yes	Not clearly defined, therefore excluded	Yes (NAC only)
Survival						
All-cause mortality	Defined as rate of death from any cause	Defined as OS	Number of deaths	Defined as OS	Deaths from any cause	
IPF-related death	Reported as treatment-emergent -IPF-related mortality and defined as deaths occurring between randomisation and within 28 days of last dose of study drug	Reported as IPF-related mortality and defined as deaths occurring between randomisation and within 28 days of last dose of study drug	No	Defined as death from respiratory cause		
PFS	Defined as a confirmed ≥10%	Defined as confirmed ≥10% decline in percent predicted	Defined as VC decline of ≥10% or death.)	No	Excluded as only reported the	Defined as decline of

Treatment effects are estimated as odds ratios (OR), and then converted to relative risks (RR) using the average rate in the placebo arms over all studies in the NMA for use in the cost effectiveness model (clarification response,[10] Appendix D). For the ERG base-case network the average rate of all-cause discontinuation for placebo was 0.17. The estimated treatment effect for nintedanib vs pirfenidone on the odds ratio scale was OR: 1.14 (1/0.87) which equates to a relative risk of RR: 1.11.

Figure 1: All cause discontinuation, ERG base-case network - HR, 95% CrI and 95% PrI



4.9 Conclusions of the clinical efficacy section

Five RCTs compared pirfenidone at various doses with placebo in adults with mild or moderate IPF: ASCEND (Phase III),[34] CAPACITY 1 & CAPACITY 2 (Phase III),[49] SP3 (Phase III),[38] and SP2 (Phase II).[39] Three trials were international and multicentre (ASCEND and CAPACITY 1 & 2[49]), although only CAPACITY 2[49] included any UK centres (three of 110 centres across both CAPACITY trials). One RCT compared pirfenidone plus N-acetylcysteine (NAC) with placebo plus NAC in Chinese adults with mild or moderate IPF: Huang *et al.* 2015.[48]

Overall, the ERG assessed the potential risk of bias in ASCEND[34] and CAPACITY 1 & 2[49] to be low across most domains, with the exception of reporting bias and "other bias", which were judged to be "moderate" on account of inconsistencies between some outcomes and analyses presented in the trial protocols, those presented in published manuscripts and those reported in the CS,[4] and the possible influence of uncontrolled variables such as rate of disease progression.

The SP3,[38] SP2[39] and Huang *et al.* (2015) trials[48] were at a higher or more unclear risk of bias across many domains than the ASCEND[34] and CAPACITY[49] trials. These trials all evaluate lower doses of

pirfenidone which are licensed in Japan but not in the UK, apply different eligibility criteria and present noticeable differences from the other three trials in some baseline characteristics of participants.

The final selection of three trials (ASCEND,[34] CAPACITY 1 and CAPACITY 2[49]) for the main clinical efficacy review was considered to be appropriate by the ERG. However, there are some between-trial differences across some baseline characteristics, such as mean FVC or 6MWD at baseline, but subgroup analyses suggested that these and other variables did not influence treatment effect. A *post hoc* pooled analysis of ASCEND[34] and CAPACITY 1 & 2[49] found no evidence of interaction between treatment for those patients with baseline FVC ≥ 80% predicted and those with FVC < 80% predicted.

The CS[4] reported three measures of lung function based on FVC: change from baseline in percent predicted FVC/VC; change from baseline in FVC/VC (ml); and relative proportions in each trial arm with FVC categorical decline of ≥10% percent predicted (this latter outcome measure included "death" in some analyses). The findings were not consistently statistically significant across trials for these outcome measures: ASCEND (52 weeks) [34] and CAPACITY 2 (72 weeks) [49] found statistically significant benefits for those on pirfenidone compared with those on placebo for mean change from baseline in percent predicted FVC (mean difference 4.78%; $p<0.001$ and mean difference 4.4%; relative difference 35.3%; 95% CI 0·7 to 9·1 $p=0.001$, respectively); but CAPACITY 1[49] found no statistically significant benefit for those on pirfenidone compared with those on placebo (absolute difference: 0.6%; relative difference: 6.5%; 95% CI -3·5 to 4·7 $p=0.501$). Pooled analyses of the CAPACITY trials[49] found statistically significant benefits for those on pirfenidone compared with those on placebo (absolute difference: 2.5%; relative difference: 22.8%; $p=0.005$). SP3[38] also reported statistically significant benefits for those on pirfenidone for change from baseline in percent predicted VC at 52 weeks ($p=0.044$); and change from baseline in VC (ml) ($p=0.042$). Huang *et al.* (2015)[48] reported a statistically significant mean change in FVC from baseline in favour of pirfenidone plus NAC compared with placebo plus NAC at 24 weeks ($p=0.02$) but not at 48 weeks ($p=0.11$). Meta-analyses of change in percent predicted FVC for CAPACITY 1 & 2[49] and ASCEND[34] and change in percent predicted VC for SP3,[38] suggested that pirfenidone reduces the decline in percentage predicted FVC compared with placebo up to 52 weeks (MD: 3.4, 95% CI: 1.87 to 4.94, *p*-value not reported). The meta-analysis also suggested that pirfenidone slows the rate of decline in FVC (MD: 0.12, 95% CI: 0.05 to 0.19, *p*-value not reported) up to 52 weeks.

In terms of decline in FVC by ≥10%, or death, ASCEND[34] reported a statistically significant difference in favour of pirfenidone compared with placebo at week 52 (absolute difference: 15.3 [95% CI not reported], $p<0.001$). For CAPACITY 1[49] the treatment effect at week 72 favoured pirfenidone

Limitations

The ERG notes that the main limitations of the company's meta-analysis relate to the following:

- Combining the 48-week outcome data from the CAPACITY trials[49] with the 52 week data from ASCEND[34] and SP3 trials.[38] Although the direction of effect for all analysed outcomes were the same for the 52 week and 72 week data, the magnitude of effect of pirfenidone was generally less at 72 weeks than 52 weeks.
- Inclusion of the SP3 trial[38] to assess the following outcomes: lung capacity (FVC/VC percentage predicted, FVC/VC (L)); PFS; acute exacerbation; and serious adverse events. SP3[38] used a lower dose (1,800mg/day) of pirfenidone, which is licensed in Japan but not in the UK, and included only Japanese patients. In contrast, the CAPACITY 1 & 2[49] and ASCEND[34] studies used licence doses of pirfenidone (2,403mg/day) and included people from Europe and the USA.
- Variation in outcome definitions used across the included trials for PFS, acute exacerbation, 6MWT, lung function and combining data of FVC with VC for lung function.

The NMA included trials were of different durations. CAPACITY 1 and 2[49] presented data at 72 weeks whilst the maximum follow up for the other studies (of interventions relevant to the scope) was at 52 weeks. Trials with a shorter follow-up might be expected to observe fewer negative outcomes and so in order to facilitate synthesis across trials, the NMA used data from CAPACITY 1 and 2[49] evaluated at an earlier follow up time of either 48 or 52 weeks (depending on the outcome). This is a valid approach for evaluating the treatment effects at a specific time point but means that the analyses did not make use of the full follow-up data available. Alternative methods that allow the incorporation of trials of different durations, whilst accounting for time effects, could have been used.

For time-to-event outcomes (all-cause mortality, PFS, IPF related mortality) the treatment effects are reported as HRs, which are time averaged estimates of treatment effect and under the assumption of proportional hazards should be constant over time. The CS[4] provided evidence to support the assumption of proportional hazards but, despite this, data at 52 weeks were used in the company's base-case NMAs rather than the full 72-week data. Although there is not enough evidence to reject the assumption of proportional hazards for the presented pirfenidone data, the ERG notes that treatment effects at 72 weeks were often substantially lower than those at 52 weeks. The company[4] reported that there was no evidence to support that proportional hazards hold for nintedanib in the long-term.

The company also described other potential sources of heterogeneity between trials, in terms of differences in outcome definitions and handling of missing data. Due to the limited number of studies

following price reduction.[92] Pirfenidone was dominated by nintedanib in the CDR for nintedanib (assumption of equal efficacy but nintedanib was less costly).[91] Finally, Loveman *et al*. (2014) reported that, at the list price, pirfenidone was dominated by inhaled NAC.[93, 94]

Quality assessment tables are presented in CS Appendix 19. Following quality assessment, the company reports that "*the CDRs provide only a brief summary of the cost effectiveness results and therefore score poorly against most areas of the Drummond quality assessment check list*" (see CS page 194) and have limited relevance to the UK. The ERG considers this to be justified but raises attention to particular comments expressed during these assessments [91, 92] that are relevant for this appraisal including: (a) the uncertainty around the duration of the treatment effect for pirfenidone and nintedanib against BSC; (b) the uncertainty around the relative effectiveness between pirfenidone and nintedanib, and; (c) concerns regarding the discontinuation rate and the assumption that the treatment effect remains following discontinuation.

The CS does not report results from the quality assessment for the previous model submitted to NICE[42] but does summarise some of the concerns expressed by the ERG[89] including the appropriateness of the model structure, comparators included and uncertainty around the clinical effectiveness of pirfenidone versus BSC. In Appendix 19 of the CS, the ERG observes that according to the company, the model that was previously submitted to NICE performed poorly against most areas of the Drummond quality assessment checklist[90] (did not conform to 17 criteria, conformed to 15 criteria and 4 criteria were non-applicable).

Finally, the company considered the Loveman study[93, 94] to be of high quality when assessed against the Drummond quality assessment checklist but that the relevance to the UK is limited given: (a) the study did not include data from the ASCEND and INPULSIS trials; (b) the inclusion of a trial in severe IPF; (c) utility values were taken from a non-UK source; (d) efficacy data were taken from studies outside the UK, and; (e) "*for pirfenidone the data were taken from two Japanese studies and two multi-national studies (of which the UK was one country).*" The ERG notes that whilst the company appears to suggest that the inclusion of Japanese studies is a limitation in its systematic review, as described in Section 4.6, despite a request from the ERG, the company refused to exclude Japanese studies from the NMA.

5.1.4 Conclusions of the review of published cost-effectiveness studies

The CS draws some conclusions regarding the quality of the included studies, comments on the applicability of the studies to the decision problem for this appraisal and tabulates the ICERs reported. Whilst the ERG is generally satisfied with the cost-effectiveness review presented by the company, the ERG considers the decision to exclude the model used for the nintedanib submission[26] from the cost-effectiveness review to be questionable. The ERG observes that the population entering the model resembles the population included in the INPULSIS and TOMORROW trials which consisted of people with a percent predicted FVC >50% at baseline and therefore consists of people considered to have mild to moderate IPF which is relevant for this submission. The ERG further notes that whilst people included in the nintedanib trials had milder disease compared with the population included in the pirfenidone trials (approximately 45% had a FVC >80% compared with approximately 25% in the pirfenidone trials), an analysis is conducted for an ASCEND-like population (defined as FVC 50-90% predicted, $FEV_1/FVC \geq 0.8$).[12, 26] The ERG considers that this study should have been included in the company's systematic review in addition to the original CDR for pirfenidone for consistency. The nintedanib model uses a cohort state transition approach whereby people entering the model progress through a series of health states defined by roughly 10 point percent predicted FVC intervals. EQ-5D scores were taken directly from the INPULSIS trials. In this assessment, pirfenidone was dominated by nintedanib when the stopping rule was applied to both or none of the interventions in people with a percent predicted FVC <80% at baseline (including the price discount for both interventions).

The ERG further notes that the CS does not provide any conclusions regarding the cost-effectiveness of pirfenidone compared with BSC or nintedanib based on this review of published cost-effectiveness analyses.

In summary, the ERG notes some inconsistencies in the company's review and considers that it is challenging to compare results from the different models given the differences in model structure, assumptions, data used and the existence of confidential price discounts.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

This section presents a summary description of the model submitted as part of the CS. ERG comments are provided directly after each aspect of the model is described.

5.2.1. Consistency of the CS with the requirements set out in the NICE reference case

As part of its submission to NICE, the company submitted a fully executable health economic model programmed in Microsoft Excel alongside a written description of the methods and results. A revised version of the model was submitted in response to the clarification questions from the ERG. The

Evidence on resource use and cost	Resource use estimates for routine management are based on telephone discussion with UK clinical experts. Hospitalisation data are based on estimates from pirfenidone trials. Unit costs are taken from NHS reference costs. Drug costs in the main CS are based on list prices (results which incorporated the PAS for nintedanib are reported in a confidential appendix). Costs of end of life care were taken from the literature.
Time horizon	Lifetime
Discount rate	3.5% per year for both costs and QALYs
Equality considerations	No weighting has been applied to QALYs

BSC – best supportive care; ITT – intention to treat; FVC – Forced vital capacity; QALY – quality-adjusted life year; IPF- idiopathic pulmonary fibrosis

[a] defined in the trial as symptom relief, pulmonary rehabilitation, management of comorbidities, and end of life care, including oxygen therapy

The population entering the company's model reflects the population included in the CAPACITY[49] and ASCEND trials.[34] Similarly, the intervention and associated treatment regimen assumed in the economic model reflects the regimens used in the Phase III trials.[34,49] The intervention consists of pirfenidone (267mg capsules, given orally), given as three 267mg capsules, three times a day, giving a total of 2403mg/day; before adjustments for dose reductions and interruptions. In the company's base-case, people initiating pirfenidone are assumed to discontinue treatment at the rate observed in the CAPACITY 1 & 2, ASCEND and RECAP studies; these discontinuation rates are not adjusted to reflect the implementation of the stopping rule in the base-case. The stopping rule defined by NICE which formed the basis for the positive recommendation for pirfenidone[2] and nintedanib[12] is however applied to nintedanib in the company's base-case and only in a scenario analysis for pirfenidone.

5.2.1.1. ERG comments on the population described in the CS and included in the company's model

The ERG is satisfied that the population and subgroups addressed by the company are largely in line with the final NICE scope.[3] In the CAPACITY/ASCEND trials,[34,49] which formed the main basis of the evidence used in the economic model, individuals were eligible if they had a percent predicted FVC ≥50% and predicted diffusing capacity of the lungs for carbon monoxide (DLco) ≥ 35% (≥ 30% in the ASCEND trial). This is largely in line with the definition provided by NICE in the final scope[3] for mild-to-moderate IPF; defined as "*a FVC greater than or equal to 50% predicted and a diffusing*

as separate health states, but are instead assumed to be treatment-specific and are applied within each model cycle.

QALYs are calculated as a function of time spent in the pre-/post-progression states with different utilities applied in each state. Cost components include drug acquisition, costs associated with the management of the condition, adverse events, acute exacerbation and end of life.

It should be noted that within its submission, the company makes reference to three modelling approaches that have been used in IPF: (i) the micro-simulation model submitted during the first appraisal of pirfenidone[2] (submitted by InterMune); (ii) the state transition approach based on percent predicted FVC categories submitted as part of the nintedanib NICE appraisal,[12,26] and; (iii) the state transition approach published by Loveman *et al.* (2014)[93,94] which is based on four main health states (unprogressed IPF, progressed IPF, lung-transplant and dead). The company considers that the micro-simulation approach used in the previous NICE submission[97] and the approach used in the nintedanib NICE appraisal[26] add complexity and are difficult to parameterise and therefore are not appropriate.

5.2.2.1. ERG's comments on conceptual representation of the condition

The ERG has a number of concerns regarding the structure and logic of the company's model. These can be separated into four sets of issues: (i) the conceptual representation of the condition; (ii) the representation of the treatment pathway in IPF; (iii) the use of a partitioned survival model approach and HR, and; (iv) questionable structural assumptions.

The ERG considers that the company's model ignores a key facet of the disease: specifically that IPF is a progressive condition characterised by irreversible loss of lung function. The company's justification to use PFS in the model relies on three key sets of arguments: (i) findings from a review by Albera et al[99] which concluded that PFS could be deemed to be an appropriate endpoint in IPF trials; (ii) that this approach has been used in a previous economic evaluation,[94] and; (iii) the difficulty in parameterising a model based on percent predicted FVC (as used in the nintedanib appraisal[12,26]).

The ERG considers that whilst PFS could be considered as an appropriate endpoint in trials when evaluating the effect of an intervention in IPF, separating the natural history of IPF into two distinct consecutive phases (the presence/absence of progression) is overly simplistic and does not reflect the natural history of the condition or its progressive nature. This limitation is recognised in the CS (page 278) when results are compared against those generated during the original submission to NICE.[2] The company states that "*the impact on patient quality of life has been conservatively included for one progression alone in the updated model*" (see CS,[4] page 278). The CS therefore acknowledges that this simplification has the potential to bias the QALY gains estimated by the model. However,

- (Figure 34). A parallel plot of the log-cumulative hazards for BSC and pirfenidone would suggest that the assumption of proportional hazards is reasonable within the trial period. Upon inspection of Figure 34, this assumption is questionable.
- Finally, the ERG advises considerable caution in the interpretation of any comparisons made by the company between the pirfenidone arm of the CAPACITY/ASCEND/RECAP trials[34, 40, 49] and data from registries. The ERG considers that such analyses are inherently subject to considerable bias. In brief:
  a. Despite the attempt by the company to select and match individuals from registries to people enrolled in the ASCEND and CAPACITY trials,[34, 49] the survival of individuals from the registries is inconsistent with the OS of people initiating BSC observed in the clinical trials (see Figure 35). The ERG notes that the company does not comment on the discrepancies between the OS in people enrolled in the CAPACITY/ASCEND trials[34, 49] and people enrolled in registries who were treated with BSC.
  b. The long-term survival for pirfenidone is based on the RECAP trial (OLE study of ongoing pirfenidone treatment) which enrolled people with IPF who completed the final follow-up visit of the CAPACITY-trials and received ≥80% of the assigned study treatment. Clarifications were requested from the company regarding the rationale for excluding people from RECAP who received less than 80% of the assigned study treatment (see clarification response,[10] question B2). In response, the company stated that "*Patients using less than 80% of drug are considered to be non-compliant (a standard cut-off being used in many trials), and for this reason were not included in RECAP. Although RECAP was an open-label extension study, the standard compliance considerations were still applied.*" Consequently, the ERG considers that the exclusion of people who received less than 80% of the assigned study treatment could overestimate the survival for pirfenidone as only people that are considered to be compliant have been included in RECAP, thereby making comparison with long-term registries less relevant.
  c. Finally, whilst individuals from the registries were matched to people included in the clinical trials, the ERG notes some potential discrepancies in the inclusion criteria applied to the registry data which may bias the estimate in favour of pirfenidone. For instance, the company excluded individuals with a percent predicted FVC ≥90% (if DLco≥90%).Throughout the CS, the company discuss a potential link between FVC and mortality; thus, excluding people with a percent predicted FVC≥90% could underestimate the survival in individuals included in the registries. However, this exclusion criteria only resulted in the exclusion of 1 patient from INOVA and 1 patient from Euro IPF, so any bias introduced is likely to be minimal.

particularly given the similarities between health states between the Loveman et al[93, 94] model and the company's model. Therefore, the ERG cannot be certain that all relevant resource use data have been identified and presented in the CS.

The ERG is generally satisfied with the inclusion of drug acquisition costs in the company's model but notes following clarification that; dose interruptions and reductions for pirfenidone are calculated after titration and therefore exclude the first 2 weeks. The ERG considers that a more appropriate approach would have been to separate the costs for the first model cycle from those for subsequent cycles. This is amended in the ERG preferred-base-case.

The ERG notes that the daily cost of pirfenidone and nintedanib is equivalent when assuming the full indicated dose is taken (after the titration period for pirfenidone) and when using the current list price.[16] However, assuming the same daily costs for pirfenidone and nintedanib based on the average dose used in the pirfenidone trials implies the same impact of dose reductions/interruptions for pirfenidone and nintedanib. The ERG notes that the price structure for pirfenidone and nintedanib is different and that a dose reduction with nintedanib (for instance, from 150mg to 100 mg) would not be associated with a reduction in costs. The ERG observes that the INPULSIS trial[113] reported a compliance with nintedanib of 96.4 % whereas the mean dose applied in the model for pirfenidone is 87.6% of the indicated dose. Therefore, the ERG considers that assuming the same cost for pirfenidone and nintedanib is likely to favour nintedanib.

The company's base-case assumes no drug acquisition costs for BSC and/or concomitant medications. The ERG considers this to be inappropriate as within the trials, individuals received concomitant medications as part of BSC. This was included in the nintedanib submission at a cost of approximately £25 per model cycle calculated from the trial for both nintedanib and BSC.[26] However, the ERG notes that the impact of the ICER is likely to be minimal given that the cost will be applied to all arms.

The CS also reports that resources use estimates were derived from discussion with a panel of clinicians, although no details were provided in the CS. In response to a request for clarification (clarification response,[10] question B16), the company provided further details, stating that: "*One-to-one telephone interviews were conducted with the panel of UK clinical experts. Content of the earlier NICE manufacturer submission was discussed, along with how the approach employed to assess resource use in the earlier submission matched current clinical practice in IPF. Discussions accounted for the revised descriptions of the NHS Reference Cost list for 2014-15 compared to earlier years (e.g. revision of 'simple lung exercise function test' to 'field exercise test')."* Despite this additional clarification, the ERG considers the process used by the company to elicit resource use has

5.2.13. Model validation

The company reports two main methods of model validation:

- Comparison of the model predictions with results from previous evaluations,
- Validation of the long-term prediction of survival.

The CS provides a comparison of the model outcomes from its model with those from the company's submission, in the nintedanib appraisal,[26] and the company's submission in the previous appraisal of pirfenidone[2] (see Table 66).

Table 8: Comparison of LYs and QALYs – moderate population (reproduced from CS,[4] Table 122)

Outcome	NTB submission[26]			This submission			TA282	
	BSC	NTB	PFN	BSC	NTB	PFN	BSC	PFN
Total QALYs	3.27	3.67	3.62	3.15	3.77	4.46	3.18	4.30
Total LYs	4.36	4.86	4.86	4.33	5.30	6.47	4.40	5.96

Key: IPF, idiopathic pulmonary fibrosis; LY, life year; NTB, nintedanib; PFN, pirfenidone; QALY, quality-adjusted life year

The CS also provides a comparison of OS from their model compared with two studies (see Table 67) which uses observational data (both sources are described further in Table 59 of the CS). Fisher al (2015)[64] reports OS from a modelling study whereby the OS in patients initiating BSC is modelled from a log-normal distribution which is fitted to data from the US strand registry. The OS in patients initiating pirfenidone is modelled from a log-normal distribution which is fitted to data from the RECAP trial. The Roskell *et al.* study[66] is also a modelling study and uses data from the RECAP OLE for pirfenidone (Weibull distribution fitted to the KM). The survival in patients initiating BSC was taken the CPRD and included patients with a FVC > 50% only. A Weibull distribution was fitted to the CPRD data.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

This section summarises additional analyses undertaken by the ERG using the company's model as well as the development of an ERG-preferred base-case.

The ERG expressed a number of concerns regarding the model structure and parameterisation of the company's model. A key concern related to the lack of ability of the model to capture the progressive nature of IPF and inflexibility associated with the modelling approach chosen by the company (partitioned survival model) which meant that correlations between outcomes are not captured in the model. This is a concern as the modelled stopping rule impacts on costs but not health outcomes. Importantly, the company's model also relies on a strong assumption that the treatment effect estimated within the trials (up to 52 weeks) is maintained over the entire model's duration (34 years). Such extrapolation is questionable and subject to considerable uncertainty. This leads to discrepancies between the model-predicted OS and observed OS in people initiating BSC from the ASCEND/CAPACITY trial (see Figure 35).

Unfortunately, a number of the issues identified cannot be addressed by the ERG without major restructuring of the economic model. It should also be noted that changes to the model are challenging given the structure of the model whereby outcomes are disconnected from each other. The ERG is not able to adequately amend the implementation of the stopping rule within the company's existing model structure and thus, considers that any ICER generated in the scenarios using the stopping rule need to be interpreted with caution as they are likely to provide ICERs that are favourable to pirfenidone when compared against BSC.

The following analyses were undertaken by the ERG to inform its base-case:

1. **Using the ERG's preferred estimate of the treatment effect, which uses data up to 72 weeks, excludes SP3, and uses the CODA samples from the predictive distribution.** As described in Section 5.2.4.1, the ERG considered the treatment effect estimated at 72 weeks to be more appropriate and more consistent with the company's assumption of proportional hazards. Furthermore, the ERG considered that SP3 should be removed from the network as this trial was conducted in a Japanese population, a dose licensed in Japan but not in the UK was given and the HR was not directly available which could introduce a bias. Finally, the ERG considered that the CODA samples (from the predictive distribution) should be used instead of the median HR in order to properly capture the joint uncertainty in the effectiveness estimates, and therefore the results for this scenario are run probabilistically.
2. **Use of the Gompertz distribution for OS (rather than the Weibull).** As described in Section 5.2.3.1, the ERG considered the Gompertz distribution to provide a more plausible long-term extrapolation compared with the Weibull distribution.

7. those reported by the company in Table 19 of the response to clarification (see addendum to clarification response).
8. **Using compliance from INPULSIS for nintedanib.** Given the different price structure, the ERG considered that assuming the same impact of dose reductions/interruptions between pirfenidone and nintedanib is likely to be unfavourable to pirfenidone. Consequently, an analysis is conducted assuming a compliance of 96.4% for nintedanib based on data from the INPULSIS trial.[113]
9. **Corrections of errors in the economic model.** As part of the critical appraisal of the model, the ERG identified a series of minor programming errors which have been corrected. These are described in appendix 4.

The impact of each individual change is reported in Section 6 in addition to the ERG-preferred base-case which combines all these changes. For consistency, results are reported with and without the stopping rule (same assumption for both treatments). It should also be noted that the ERG-preferred base-case is presented as a range (most optimistic to most pessimistic scenario) given the uncertainty surrounding the extrapolation of the treatment effect.

5.4 Conclusions of the cost-effectiveness section

The company submitted a fully executable economic model as part of their submission to NICE. The analysis was undertaken from the perspective of the UK NHS and PSS over a lifetime horizon. The company's analysis is presented for three populations: (1) the ITT trial population, which is comprised of adults with mild to moderate IPF; (2) people with a percent predicted FVC > 80% at baseline (considered to be mild IPF), and; (3) people with a percent predicted FVC of 50 - 80% at baseline (considered to be moderate IPF). All three analyses include BSC as a comparator (defined in the trial as symptom relief, pulmonary rehabilitation, management of comorbidities, and end of life care, including oxygen therapy). Nintedanib is included as a comparator only in the analysis of people with a percent predicted FVC of 50 - 80% at baseline.

The analysis in the ITT population does not include nintedanib as a comparator as nintedanib is only a valid comparator for the subgroup of the ITT population with moderate IPF (percent predicted FVC of 50 - 80%). The ERG considers that it is more appropriate to conduct an economic analysis separately within the mild and moderate subgroups as the comparators vary by subgroup.

In the company's base-case, people initiating pirfenidone are assumed to discontinue treatment at the rate observed in the trials, hence no stopping rule is applied. The stopping rule defined by NICE which formed the basis for the positive recommendation for pirfenidone[2] and nintedanib[12] is however

Table 9: Summary of the impact of individual changes to the ICER for pirfenidone versus BSC[a] using the list price and mean parameter inputs (deterministic model)

	ITT population	People with a percent predicted FVC of 50 – 80%[a]	People with a percent predicted FVC > 80%
Company base-case	[redacted]	[redacted]	[redacted]
No stopping rule for nintedanib	[redacted]	[redacted]	[redacted]
Inclusion of stopping rule for pirfenidone	[redacted]	[redacted]	[redacted]
Treatment effect assumed to stop after 2 years	[redacted]	[redacted]	[redacted]
Gompertz distribution for OS	[redacted]	[redacted]	[redacted]
HRQoL capped at 1.0	[redacted]	[redacted]	[redacted]
Adjustment of HRQoL by age	[redacted]	[redacted]	[redacted]
End of life costs applied to death irrespective of causes	[redacted]	[redacted]	[redacted]
Pirfenidone dose titration	[redacted]	[redacted]	[redacted]
Nintedanib compliance taken from INPULSIS	[redacted]	[redacted]	[redacted]
Correction of errors	[redacted]	[redacted]	[redacted]

[redacted]