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.

Declared competing interests of the authors

None of the authors have any conflicts of interest to declare.

Acknowledgements

We would like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Davis S, Rafia R, Carroll C, Essat M, Sanderson J, Dracup N, Bianchi S, Thickett D. Pirfenidone for treating idiopathic pulmonary fibrosis: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2016.

Contributions of authors

Christopher Carroll and Munira Essat summarised and critiqued the clinical effectiveness data reported within the company's submission. Sarah Davis and Rachid Rafia critiqued the health economic analysis submitted by the company. Jean Sanderson critiqued the statistical analyses undertaken by the company. Naila Dracup critiqued the company's search strategy. Dr Stephen Bianchi and Professor David Thickett acted as clinical advisors to the ERG. All authors were involved in drafting and commenting on the final report.

Abbreviations

AEs	Adverse events
BSC	Best supportive care
CADTH	Canadian Agency for Drugs and Technologies in Health
COPD	Chronic obstructive pulmonary disease
CrI	Credible interval
CRD	Centre for Reviews and Dissemination
CS	Company's submission
CSR	Clinical study review
DLco	Diffusing capacity for carbon monoxide
ERG	Evidence Review Group
EQ-5D	EuroQoL five dimensions questionnaire
FVC	Forced vital capacity
HR	Hazard ratios
HRCT	High resolution computed tomography
HRQoL	Health-related quality of life
HTA	Health Technology Appraisal
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient data
IPF	Idiopathic pulmonary fibrosis
ITT	Intention to treat
KM	Kaplan-Meier
LOCF	Last observation carried forward
NAC	N-acetylcysteine
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
NMA	Network meta-analysis
PBO	Placebo
PBS	Pharmaceutical Benefits Scheme
PEY	Person exposure years
PFN	Pirfenidone
PFS	Progression-free survival
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial

SD	Standard deviation
6MWD	Six minute walking distance
SGRQ	St George's Respiratory Questionnaire
STA	Single Technology Appraisal
TA	Technology Appraisal
VC	Vital capacity
UCSD SOBQ	University of California San Diego Shortness of Breath Questionnaire

CONTENTS

TABLES

FIGURES

1 SUMMARY

1.1 Critique of the decision problem in the company's submission

The population addressed in the company's submission (CS) is adults with mild to moderate idiopathic pulmonary fibrosis (IPF), as specified in the final scope issued by the National Institute for Health and Care Excellence (NICE).

The Evidence Review Group (ERG) notes that patients included in the main clinical trials for pirfenidone, may not be wholly representative of the population likely to receive pirfenidone in clinical practice as real-life patients often have comorbidities, more severe disease, take concomitant medications and have a higher mortality risk compared with those patients enrolled within the clinical trials. Patients with obstructive airway disease were excluded from the clinical trials. However, clinical advisors to the ERG stated that patients with obstructive airway disease may be offered pirfenidone in current clinical practice, provided that they meet the treatment criteria laid out in technology appraisal (TA) 282.

The final NICE scope specified that if evidence allows, subgroup analysis by disease severity, defined by forced vital capacity (FVC) (such as above and below or 80% FVC) and/or diffusing capacity for carbon monoxide (DLco), should be considered. However, the CS states that available data only allowed subgroups by FVC to be assessed.

In the company's health economic analysis, the CS presents results for three populations: mild to moderate IPF (described as the intention to treat [ITT] population), mild IPF (percent predicted FVC >80%) and moderate IPF (percent predicted FVC of 50 – 80%). No subgroups results are presented by DLco status.

The intervention specified in the final NICE scope is pirfenidone and the comparators specified are best supportive care (BSC) and nintedanib. Nintedanib is only listed in the scope as a comparator for the subgroup of patients with a percent predicted FVC of between 50% and 80% as this is the population recommended for treatment in the NICE appraisal of nintedanib (TA379).

Within the economic analysis nintedanib and BSC have been included as comparators for the subgroup of patients with moderate IPF (percent predicted FVC of 50 – 80%) and BSC has been included as a comparator for the subgroup of patients with mild IPF (percent predicted FVC >80%). The ERG considers the comparators chosen for the mild and moderate subgroups to be appropriate.

For the economic analysis considering the ITT population, which includes patients with both mild and moderate IPF, only BSC is included as a comparator. The ERG does not consider this analysis to be

relevant to the decision problem as nintedanib is 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 general, the CS adequately addresses the range of outcomes specified in the final NICE scope. The majority of the outcomes were reported for both the direct comparison with placebo from the pirfenidone clinical trial programme and for the indirect comparison with nintedanib from the network meta-analysis (NMA).

The definition of progression-free survival (PFS) used across the pirfenidone trials was not consistent; however, where possible, individual patient data (IPD) were re-analysed to provide results based on a consistent definition. However, this could not be done for all of the trials which contributed to the NMA. The ERG considers that the NMAs which combined data from studies using different definitions should be interpreted with caution.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company submitted a systematic review of randomised controlled trial (RCT) evidence comparing pirfenidone with placebo in adults with mild or moderate IPF. The review identified three multi-centre international RCTs: ASCEND and CAPACITY 1 compared pirfenidone at the licensed dose of 2,403mg per day with placebo, whilst CAPACITY 2 compared pirfenidone at doses of 2,403mg per day and 1,197mg per day with placebo. It also identified two multi-centre Japanese RCTs, which compared lower doses of pirfenidone with placebo: SP3 evaluated pirfenidone doses of 1,800mg per day and 1,200mg per day and SP2 1,800mg per day only. The five trials included more than 1,700 patients with IPF. The ASCEND and SP3 trials had 52 weeks follow-up, the CAPACITY trials had 72 weeks follow-up, and the SP2 trial was terminated early at 36 weeks. The company critically appraised all five RCTs and assessed the overall risk of bias in all trials to be low.

The primary efficacy outcome for all of these trials, except SP2, was change from baseline in percent predicted FVC. The magnitude of treatment effect was also measured by mean change from baseline in FVC (ml) and the categorical outcome of a $\geq$10% decline in percent predicted FVC or death. These trials also reported all-cause and IPF-related mortality, PFS (using different definitions), 6-Minute Walking Distance (6MWD), DLco, and patient-reported outcomes, as measured by the University of San Diego Shortness of Breath Questionnaire (UCSD SOBQ) for dyspnoea, and the St George's Respiratory Questionnaire (SGRQ).

The company focused on the categorical outcome of a ≥10% decline in percent predicted FVC or death. For this outcome, the ASCEND trial reported a statistically significant difference in favour of pirfenidone compared with placebo at week 52 (absolute difference: 15.3 [95% Confidence Interval (CI) not reported], $p<0.001$), as did CAPACITY 2 at week 72 (absolute difference: 14.4 [95% CI: 7.4 to 21.3], $p=0.001$). CAPACITY 1 reported that there was no statistically significant difference between pirfenidone and placebo at week 72 (absolute difference: 3.8 [95% CI: -2.7 to 10.2], $p=0.440$). ASCEND also reported a significantly higher proportion of patients with no decline in percent predicted FVC (22.7% for pirfenidone versus 9.7% for placebo, $p<0.000001$), whilst CAPACITY 2 reported a higher proportion of patients with no decline in percent predicted FVC for pirfenidone compared with placebo (24.1% versus 13.8%) but did not report a *p*-value. CAPACITY 1 reported no statistically significant difference between pirfenidone and placebo on this outcome measure (25.8% versus 22%, *p*-value not reported). A meta-analysis of the ASCEND trial (52 weeks) and the CAPACITY trials (48 weeks) suggested that, compared with placebo, pirfenidone lowers the proportion of patients experiencing decline in FVC percent predicted of ≥10% (odds ratio [OR]: 0.50, 95% CI: 0.31 to 0.82, *p*-value not reported).

In terms of change from baseline in FVC, ASCEND (52 weeks) and CAPACITY 2 (72 weeks) found statistically significant benefits for those on pirfenidone compared with those on placebo (mean difference [MD] 4.78%; $p<0.001$ for ASCEND and absolute difference 4.4%; relative difference 35.3%; CI 0.7 to 9.1 $p=0.001$ for CAPACITY 2), whilst CAPACITY 1 found no statistically significant difference for pirfenidone compared to 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 found statistically significant benefits for those on pirfenidone compared with placebo (absolute difference: 2.5%; relative difference: 22.8%; $p=0.005$). SP3, which reported Vital Capacity (VC), rather than FVC, 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) ($p=0.042$). Meta-analyses of change in percent predicted FVC for CAPACITY 1 & 2 and ASCEND, and change in percent predicted VC for SP3, suggested that pirfenidone reduces the decline in percent 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.

There were fewer overall deaths or treatment-emergent IPF-related deaths in the pirfenidone than the placebo arms of the ASCEND and CAPACITY trials. These differences were not statistically significant in the ASCEND trial at 52 weeks (for all-cause mortality or treatment-emergent IPF-related deaths, $p=0.105$ and $p=0.226$, respectively), but were significant in the pooled analyses for the CAPACITY trials at 52 weeks (for all-cause mortality and treatment-emergent IPF-related deaths,

p=0.047 and *p*=0.012, respectively). There was a significant difference between groups for treatment-emergent IPF-related mortality in the pooled CAPACITY trials at 72 weeks (*p*=0.03). Meta-analysis of CAPACITY 1 & 2 and ASCEND compared with placebo, at 52 weeks, suggests that pirfenidone reduces all-cause mortality (HR: 0.52, 95% CI: 0.31 to 0.88, *p*-value not reported) and IPF-related mortality (HR: 0.37, 95% CI: 0.18 to 0.76, *p*-value not reported).

Four of the key trials reported data for PFS: ASCEND, CAPACITY 1 & 2 and SP3. The definitions of PFS varied across the trials, albeit with a common element of a confirmed ≥10% decline from baseline in percent predicted FVC or VC. ASCEND at 52 weeks (HR 0.57; 95% CI, 0.43–0.77, *p*=0.0001) and CAPACITY 2 at 72 weeks (HR 0.64; 95% CI, 0.44–0.95, *p*=0.023) found statistically significant benefits in terms of PFS for those on pirfenidone compared with those on placebo, whilst the treatment effect for CAPACITY 1 was not statistically (HR: 0.84; 95% CI, 0.58, 1.22, *p*=0.355). *Post hoc* pooled analyses of the CAPACITY trials found statistically significant benefits for those on pirfenidone compared with those on placebo (HR: 0.62; 95% CI: 0.51 to 0.76; *p*<0.0001). Meta-analysis of the four trials, ASCEND, CAPACITY 1 & 2 and SP3 showed pirfenidone improves PFS at 52 weeks (HR 0.63 95% CI, 0.53 to 0.74, *p*-value not reported).

The CS reported the findings from two sets of analyses for 6MWD. The ASCEND and CAPACITY trials all reported findings on the pre-specified outcome of mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo. ASCEND at 52 weeks (absolute difference: 26.7m; relative reduction: 44.2%; *p*=0.036) and CAPACITY 1 at 72 weeks (absolute difference: 31.8m; relative difference: not reported; *p*<0.001) both reported a statistically significant and clinically important difference between pirfenidone and placebo on this outcome, but CAPACITY 2 did not (absolute difference: 16.4m; relative difference: not reported; *p*=0.171). A pooled analysis of the CAPACITY trials at 72 weeks also reported a statistically significant and clinically important difference between pirfenidone and placebo on this outcome (absolute difference: 24m; relative difference: 31.2%; *p*=0.0009). Meta-analysis of CAPACITY 1 & 2 (data from week 48) and ASCEND (data from week 52) suggested that pirfenidone reduces the decline in 6MWD (MD: 22.9, 95% CI (10.58 to 35.23, *p*-value not reported). A *post hoc* categorical analysis based on a mean decline ≥50 m in 6MWD from baseline, or death found that there was a statistically significant difference between pirfenidone and placebo in ASCEND (52 weeks: absolute difference: 9.8%; relative reduction: 27.5%; *p*=0.04) and CAPACITY 2 (*p*=0.049), but that there was no statistically significant treatment effect for pirfenidone in CAPACITY 1 (*p*=0.10). A pooled analysis of the CAPACITY trials (72 weeks: absolute difference: 12.2%; relative reduction: 26%; *p*=0.001) also reported a statistically significant effect for pirfenidone compared with placebo for this categorical outcome.

All five included trials reported outcome data on acute exacerbations but used different definitions. The rates of acute exacerbation were higher in the ASCEND trial than in the CAPACITY trials, with higher incidence in the placebo group compared with the pirfenidone arms in the ASCEND and CAPACITY 2 trials: no *p*-values were reported. None of these three trials reported statistically significant treatment effects for this outcome measure. A meta-analysis of ASCEND, CAPACITY 1 & 2 and SP3 indicated a treatment effect in favour of pirfenidone, although the result was not statistically significant (OR 0.64 ,95% CI: 0.38 to 1.06, *p*-value not reported). CAPACITY 1 & 2 and SP2 also reported similarities in rates of hospitalisation (due to respiratory or non-respiratory causes) between the pirfenidone and placebo arms.

Neither ASCEND, CAPACITY 1 or CAPACITY 2 showed a statistically significant treatment effect compared to placebo, as assessed using the UCSD SOBQ or the SGRQ, although results of the meta-analysis suggest that pirfenidone is associated with a statistically significant reduction in USCD SOBQ compared with placebo. Four trials (CAPACITY 1 & 2, SP3, SP2) reported data on the change from baseline in DLco. The CAPACITY trials reported the change in percent predicted DLco, whilst SP2 and SP3 reported the mean decline (mL/min/mmHG). None of the trials showed a statistically significant treatment effect compared to placebo for this outcome measure.

The company submitted evidence from an ongoing, non-controlled, open-label extension (OLE) of the ASCEND and CAPACITY trials (RECAP, PIPF-012). The RECAP study is ongoing. The most recent data-cut was performed in June 2015 and the next data-cut is planned in June 2016. Survival data and time-on-treatment data were reported in the CS and were presented for patients who received pirfenidone 2,403mg per day from baseline onwards in CAPACITY and ASCEND, and through the RECAP extension period, for whom data are available through to 8.8 years. Information on survival of patients with IPF was also presented from six registries to explore the relative survival rates of trial patients receiving pirfenidone compared with these “matched” real-world patients receiving BSC. The company stated that results were similar to the comparisons reported for the trials.

The company submitted a review of evidence on the safety of pirfenidone in patients with mild or moderate IPF. The evidence presented was from the following trials: ASCEND, CAPACITY 1 & 2, SP3, SP2, RECAP and a final, non-controlled safety trial, PIPF-002. Adverse events of any intensity with the highest frequency across all trials were nausea, rash, dizziness, dyspepsia and anorexia, and these were all relatively frequent compared with placebo (no statistically significant *p*-values for between-group differences were reported, except for IPF). SP3 and SP2 also reported a very high frequency of photo-sensitivity (much higher than the CAPACITY trials). Similar, albeit slightly higher, frequencies of these and other adverse events were found in an integrated population from the RECAP extension study. Meta-analyses of treatment-emergent serious adverse events using data from

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, unlicensed doses of pirfenidone, applied different eligibility criteria and presented noticeable differences from the other three trials in some baseline characteristics of participants.

The ERG agreed with the findings reported for the FVC outcomes for individual trials and noted that the meta-analyses generated small differences compared with the pooled analyses. The ERG also noted that the findings for CAPACITY 1 differed from those reported for CAPACITY 2 and ASCEND. The additional RCT, Huang *et al.* (2015), 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). In response to an ERG request to explain the differences between the trials on this outcome, the company stated that "*the natural variability in rates of FVC percent predicted decline of this heterogeneous disease*" might explain differences in outcomes both within and across trials.

The ERG accepts that there were fewer overall deaths or treatment-emergent IPF-related deaths in the pirfenidone arms than the placebo arms of the ASCEND and CAPACITY trials and that, in some pooled analyses, these differences were statistically significant at the 5% level. However, the ERG noted that these differences were not statistically significant in the ASCEND trial at 52 weeks and most differences that were significant in pooled analyses of the CAPACITY 1 & 2 data at 52 weeks were no longer significant at 72 weeks. However, meta-analysis of CAPACITY 1 & 2 and ASCEND at 52 weeks did suggest that pirfenidone reduces all-cause mortality (HR: 0.52, 95% CI: 0.31 to 0.88, *p*-value not reported) and IPF-related mortality (HR: 0.37, 95% CI: 0.18 to 0.76, *p*-value not reported) compared with placebo. Sensitivity analysis of the three trials at 72 weeks gave similar outcomes in favour of pirfenidone for both all-cause mortality (HR:0.64, 95% CI: 0.41 to 0.99, *p*-value not reported) and IPF-related mortality (HR: 0.49, CI: 0.27 to 0.87, *p*-value not reported), but the reduction in mortality was lower at 72 weeks compared with 52 weeks. The ERG noted that there appears to be a markedly increased rate of mortality in the CAPACITY trials between the data reported for 52 weeks and for 72 weeks, the reasons for which are unclear. SP3, SP2 and Huang *et al.* (2015) all reported all-cause mortality and found no differences between the pirfenidone and placebo arms.

The results for PFS were consistent across trials and analyses demonstrated a beneficial effect on this outcome for pirfenidone compared with placebo. The exception, again, was the CAPACITY 1 trial, which reported that the difference between pirfenidone 2,403mg per day and placebo was not significant (p=0.355).

The results for 6MWD were consistent in terms of direction of effect (favouring pirfenidone) but statistical significance varied between trials and between 6MWD outcome measures. The ASCEND and CAPACITY trials all reported findings on the pre-specified outcome of mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo. ASCEND (absolute difference 26.7 p=0.036) and CAPACITY 1 (absolute difference 31.8 p<0.001) reported a statistically significant and clinically important difference between pirfenidone and placebo on this outcome, but CAPACITY 2

did not (p=0.171). A *post hoc* categorical analysis based on a mean decline ≥50 m in 6MWD from baseline, or death, found that there was still a statistically significant difference between pirfenidone and placebo in ASCEND (p=0.04), but treatment effect for CAPACITY 1 was not statistically significant (p=0.10) and the treatment effect for CAPACITY 2 was statistically significant (p=0.049). An additional small RCT of pirfenidone in combination with NAC in adults with mild and moderate IPF identified by the ERG (Huang *et al.* 2015) also reported no statistically significant effect for pirfenidone (plus NAC) compared with placebo (plus NAC) on 6MWD outcomes (p=0.43).

The ERG noted that pirfenidone does not have a significant treatment effect compared to placebo, as assessed by a number of other outcomes: rates of acute exacerbations; patient-reported outcomes as measured by the SGRQ; or DLco. For the UCSD SOBQ the treatment effects were not statistically significant for any of the individual trials, but results of the meta-analysis suggest that pirfenidone is associated with a statistically significant reduction in in USCD SOBQ compared with placebo.

The ERG noted how the effect of the, "*intrinsic variability in rates of FVC decline*" (Noble 2011) might explain differences in some outcomes across trials. Participants in the trials included in the CS were not stratified by rate of progression, so it is possible, for example, that the placebo arm might have had more participants with more rapidly progressing disease than the intervention arm. As a result, the true treatment effect of the intervention relative to placebo is uncertain. This could work either for or against the intervention.

A *post hoc* pooled analysis of ASCEND and CAPACITY 1 & 2 found no evidence for differential treatment effects according to disease severity, as assessed using three key efficacy outcomes; absolute ≥10% FVC decline, ≥50m 6MWD decline, and ≥20-point worsening of dyspnoea as measured by UCSD SOBQ. For these analyses disease severity was categorised according to baseline percent predicted FVC of ≤80% (moderate IPF) and >80% (mild IPF). In response to a clarification request from the ERG, the company also provided subgroup analyses according to disease severity for OS and PFS from the ASCEND and CAPACITY trials, although exact numbers within each subgroup in each trial arm were not reported. The findings did not suggest differential treatment effects according to disease severity for either outcome (as judged by the reported HR and 95% CI), however a treatment-by-subgroup interaction test was not reported so it is unclear if the difference between these subgroups was statistically significant.

The ERG noted that, overall, some adverse events (AEs) were frequent, especially nausea, rash, dizziness, dyspepsia, anorexia and photosensitivity, but that these were generally mild or moderate in severity. The ERG requested from the company more detailed data on serious adverse events and adverse events leading to discontinuation. The most frequently-reported serious adverse events in the

pirfenidone arms of the ASCEND and CAPACITY trials, other than worsening of IPF, were pneumonia, prostate cancer, angina pectoris, coronary artery disease, congestive cardiac failure, atrial fibrillation and pneumothorax. The AEs leading to discontinuation of treatment in ≥1% of patients in pirfenidone groups were pneumonia, rash, raised hepatic enzyme levels and decreased weight (in ASCEND), photosensitivity, rash and respiratory failure (in CAPACITY 1) and bladder cancer, nausea and rash (in CAPACITY 2). The majority of safety data were from trials with a follow-up of no more than 72 weeks, but the CS did present analyses that included more than 300 patients who had received pirfenidone for more than four years. However, the results for these patients were not presented separately. The ERG noted that the two ongoing studies to evaluate safety would address some outstanding issues: the non-randomised, non-controlled, OLE study that included a set of patients who completed either ASCEND, CAPACITY 1 or 2 (RECAP) and PIPF-002, an ongoing open-label compassionate-use study in US patients with either IPF or secondary pulmonary fibrosis.

The ERG considers that the NMA appears to be of good methodological quality, and the choice of random effects model was appropriate given the stated concerns in terms of heterogeneity between the studies. The ERG's key concerns were in the use of the earlier 52 week follow up data for key time-to-event outcomes (all-cause mortality and PFS), rather than the full 72 week data available, and the difference in the treatment effects observed at these two time points despite the claim of proportional hazards over both the observed and unobserved time period.

1.4 Summary of cost effectiveness evidence submitted by the company

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 National Health Service (NHS) and Personal Social Services (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 above 80% at baseline (considered to be mild IPF), and; (3) people with a percent predicted FVC of 50 to 80% at baseline (considered to be moderate IPF). Within all three analyses, comparators include BSC (defined in the trial as symptom relief, pulmonary rehabilitation, management of comorbidities, and end of life care, including oxygen therapy). Nintedanib is included only in the analysis of people with a percent predicted FVC of 50 - 80% at baseline; it is excluded from the analyses of ITT-trial population and people with a percent predicted FVC above 80% at baseline. In the company's base-case, people initiating pirfenidone are assumed to discontinue treatment at the rate observed in the trials; and therefore no stopping rule is applied in the base-case. The stopping rule defined by NICE which formed the basis for the positive recommendation for pirfenidone and nintedanib is however applied to nintedanib in the company's base-case. A scenario analysis is also presented where the stopping rule is applied to both nintedanib and pirfenidone.

Within the ITT-trial population (adults with mild to moderate IPF), the company's model estimates the ICER for pirfenidone against BSC to be [redacted] per QALY gained (probabilistic ICER = [redacted] per QALY gained) using the list price for pirfenidone.

Within the subgroup of people with a percent predicted above 80% at baseline (considered to be mild IPF), the company's model estimates the ICER for pirfenidone against BSC to be [redacted] per QALY gained (probabilistic ICER = [redacted] per QALY gained) using the list price for pirfenidone.

Within the subgroup of people with a percent predicted FVC of 50 - 80% at baseline (considered to be moderate IPF), the CS estimates that BSC provided the least number of QALYs, followed by nintedanib and pirfenidone. Using the company's model estimates, based on a fully incremental analysis, nintedanib is ruled out due to extended dominance. The company's model estimates the ICER for pirfenidone against BSC to be [redacted] per QALY gained (probabilistic ICER = [redacted] per QALY gained).

Based on the company model when incorporating the PAS for pirfenidone, the ICER for pirfenidone versus BSC was £21,387 per QALY in the ITT population and £24,187 per QALY in the mild subgroup (percent predicted FVC >80% at baseline) and £21,318 per QALY in the moderate subgroup (percent predicted FVC of 50 - 80% at baseline). The results for pirfenidone versus nintedanib when incorporating the nintedanib and pirfenidone PAS (moderate subgroup) are reported in the confidential appendix.

The company presented a series of scenario analyses. The ICERs were mostly sensitive to the assumption regarding the time horizon, the duration over which the treatment effect is assumed to remain constant, the parametric distributions for OS in people initiating pirfenidone, the treatment effects taken from the NMAs for OS, and the inclusion of stopping rules for pirfenidone and nintedanib.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The ERG critically appraised the company's health economic analysis and the model upon which this analysis is based. The ERG has a number of concerns regarding the structure and parameterisation of the company's model. These include:

- the inability of the model to capture the progressive nature of IPF
- the absence of a stopping rule for pirfenidone in the company's base-case
- 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 mortality outcomes; a statistically significant treatment effect is only observed when pooling or meta-analysing studies.
- The treatment effects for a number of clinically important and patient-reported outcomes were either not statistically significant (DLco and SGRQ) or did not meet the threshold for a clinically important difference (UCD SOBQ).
- It is unclear how long the treatment effect might be sustained.

- Simplification of a progressive disease into two discrete health states (pre- and post-progression) fails to capture the ongoing progressive nature of IPF and the impact of different levels of disease severity on quality of life and costs.
- The implementation of the stopping rule in the company's model lacks validity.
- There is uncertainty around the treatment effects due to the heterogeneity between trials included in the NMAs in terms of study duration, outcome definition and handling of missing data.
- The duration of extrapolation of the treatment effect is associated with considerable uncertainty.
- There are discrepancies between the modelled OS in people initiating BSC and the OS observed in the clinical trials.
- The treatment effects for the subgroup of people with a percent predicted FVC above 80% are uncertain.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

A number of analyses were undertaken by the ERG which informed the ERG's preferred base-case. The main changes within the ERG's preferred base-case were:

- use of treatment effects estimated from the NMA from the CODA samples of the predictive distributions, using data up to 72 weeks, excluding SP3
- exploration of different durations for the extrapolation of the treatment effect (2 years and entire model duration)
- use of the Gompertz distribution for OS (rather than the Weibull)
- capping utility estimates for individuals at a maximum of 1.0
- adjustment of utility by age
- inclusion of the costs associated with end of life care for all people irrespective of the cause of death
- amendments to dose reductions/interruptions assumed in the company's model for pirfenidone and nintedanib
- amendment of minor programming errors in the economic model

The ERG's preferred scenario led to a higher ICER for pirfenidone against BSC (approximately two-fold compared with ICERs reported by the company) for all three populations (ITT, FVC of 50 - 80% 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 £31,722 - £186,260 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).

2 BACKGROUND

Pirfenidone is licensed in the EU for the treatment of mild to moderate IPF in adults.[1] Pirfenidone was previously appraised as part of the NICE Single Technology (STA) process (TA282), with guidance issued in April 2013.[2] Pirfenidone was recommended as an option for treating idiopathic pulmonary fibrosis only if the person has a forced vital capacity (FVC) > 50% and ≤ 80% predicted and the company provides pirfenidone with the discount agreed in the Patient Access Scheme (PAS). The review of TA282 was prompted by publication of the ASCEND study.[3] This report provides a review of the company's submission (CS)[4] provided by the company for pirfenidone (including any additional material submitted by the company in response to clarification requests) during NICE's review of TA282.

2.1 Critique of company's description of underlying health problem

The ERG considers that in general the company's description of the underlying health problem is appropriate and relevant to the decision problem. The ERG notes that whilst the CS states that median 5-year survival is 20%, the source paper by Kim *et al.*, estimates median 5-year survival to be between 20% and 40%.[5] Kim *et al.* also state that survival estimates are dependent on whether survival is estimated from diagnosis, symptom onset or first radiographic abnormality.[5]

The CS states that current guidelines do not propose a formal staging system for classification of disease severity. Clinical advisors to the ERG agreed with the statement in the CS that using percent predicted FVC alone to define mild and moderate disease has the potential to misclassify patients for two reasons. Firstly, FVC can be elevated in patients with emphysema, which masks the impact of fibrosis on lung capacity. Secondly, the normal range for percent predicted FVC is 90% to 120%, so some patients who have an FVC of 80% may have lost a third of their baseline lung capacity and others may have only lost a tenth. Therefore, the same percent predicted FVC may result in a different severity of IPF symptoms being experienced in different individual patients. Clinical advisors to the ERG considered that whilst percent predicted FVC had been used to define severity in clinical trials, this measure was not widely used in clinical practice, except to implement the recommendations in TA282. They commented that carbon monoxide diffusing capacity of the lungs (DLco) is clinically more meaningful and that DLco is the primary measure used to determine eligibility for lung transplantation, as some patients can have very low DLco values that suggest lung transplantation would be beneficial whilst maintaining a percent predicted FVC value that, in isolation form other measures, would indicate mild disease.

Clinical advisors to the ERG agreed that the course of IPF is unpredictable and heterogeneous. They also agreed with the statement in the CS that a prior decline in lung function does not predict a future decline and they noted that this statement is also supported by an analysis by Schmidt *et al.* based on a retrospective analysis of pulmonary function tests from 734 patients recruited across 3 centres.[6]

2.2 Critique of company's overview of current service provision

The ERG considers that in general the company's overview of current service provision is appropriate and relevant to the decision problem. However, some additional clarification on the treatment pathway described in the CS is provided below.

Whilst the ERG agrees that N-acetylcysteine (NAC) is not an appropriate comparator for pirfenidone, NAC is currently used in some patients. Clinical Guideline 163 (CG163) recommends that patients should be advised that *"oral N-acetylcysteine is used for managing idiopathic pulmonary fibrosis, but its benefits are uncertain."* Clinical advisors to the ERG confirmed that it is used in some patients for symptom relief as part of BSC but NAC is not expected to be disease-modifying. There is also a recent clinical trial of NAC versus placebo on top of a background therapy of pirfenidone in both arms, which is yet to report in full, but conference abstracts reporting preliminary results suggest that the combination is generally well tolerated but does not provide additional benefit compared to pirfenidone alone.[7, 8]

Clinical advisors to the ERG also reported that a few patients are currently managed with prednisolone or azathioprine. Whilst these drugs are not recommended in CG163 to modify disease progression, their ongoing use in some patients is a possibility under recommendation 1.5.14 of CG163 which states, *"if people with idiopathic pulmonary fibrosis are already using prednisolone or azathioprine, discuss the potential risks and benefits of discontinuing, continuing or altering therapy."* However, the ERG recognises that the use of prednisolone and azathioprine is likely to be limited to a minority of patients and is not expected to be disease-modifying.

Whilst the ERG agrees that pirfenidone has been the standard of care for patients with moderate IPF since TA282 was published in 2013, the ERG notes that following the publication of TA379 in January 2016, nintedanib is likely to become part of the standard of care in the coming months. In Section 3.6 of the CS, which describes other (non-NICE) guidelines, it is stated that pirfenidone is recommended by the ATS/ERS/JRS/ALAT Clinical Practice Guideline. The ERG notes that nintedanib is also recommended in the same document with both treatments being recommended on the basis of the panel considering that both have 'moderate confidence in effect estimates'.[9]

Clinical advisors to the ERG also noted that now that there are two disease-modifying therapies available for patients with moderate IPF, it is possible that a second therapy may be used in patients who have failed to tolerate one therapy or who have progressed on one therapy but who still meet the starting criteria for the other therapy. Treatment sequences were not addressed in the original CS. Following a clarification request, the company acknowledged that it is possible that clinicians may

sequence pirfenidone and nintedanib within the moderate population (see clarification response,[10] question B7). However, the company went on to state that no sequencing studies exist or are anticipated to become available and it is unclear whether the efficacy would be different when used second-line, particularly given that there remains uncertainty regarding the exact mechanism of both pirfenidone and nintedanib. They conclude that any analysis of treatment sequences would be purely speculative in design.

Clinical stopping rules are applied for pirfenidone in TA282 and for nintedanib in TA379. Both sets of guidance recommend that treatment is discontinued if there is evidence of disease progression which is defined as a decline in percent predicted FVC of 10% or more within any 12 month period. The CS claims that the application of this stopping rule is complicated since progression with treatment does not always constitute treatment failure. This statement is supported by a *post hoc* analysis of data from the ASCEND and CAPACITY 1 and 2 trials which showed that patients who continued with pirfenidone following a ≥10% decline in percent predicted FVC, had a significantly reduced risk of the composite outcome of death or a further 10% decline in percent predicted FVC (p=0.032), compared to those who continued with placebo following a ≥10% decline in percent predicted FVC.[11] However, it should be noted that this *post hoc* analysis may be subject to potential bias as it was based on a small proportion of the trial population (3.9% [=24/623] of patients randomised to pirfenidone and 9.6% [=60/624] of those randomised to placebo) who had experienced a 10% decline in the first 3 or 6 months of the study and who had remained on treatment,[11] and therefore patient characteristics may not be balanced between the two groups being compared. Clinical advisors to the ERG reported that to their knowledge the stopping rule is being rigorously applied in clinical practice, but they agreed that the stopping rule is clinically problematic as a prior decline in lung function does not predict a future decline, and periods of stability can sometimes only be identified retrospectively. They also noted that in clinical practice the stopping rule is only applied to patients with a >10% FVC or >15% DLco decline over any 12 month period when the lung function decline has been confirmed as not being due to a temporary and reversible infection. There will therefore be patients who will either already be defined as having severe disease on DLco criteria who would have been offered therapy due to an eligible FVC measurement or will have developed a DLco <35% but if FVC remains between 50 and 80% will have treatment continued. Similarly there is no necessity to stop a patient's therapy if the FVC declines below 50% if the decline is less than 10% per year.

3 CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 1 summarises the population, intervention, comparators and outcomes specified within the company's decision problem. These are discussed and critiqued in the following sections.

Table 1: Summary of the decision problem (adapted from Table 1 of the CS)

	Final scope issued by NICE[3]	Decision problem addressed in the CS[4]	ERG comments
Population	Adults with mild to moderate IPF	Same as final scope issued by NICE	The population addressed in the CS is consistent with the population specified in the final scope.
Intervention	Pirfenidone	Same as final scope issued by NICE	The intervention in the CS is consistent with the population specified in the final scope.
Comparators	• Best supportive care • Nintedanib (only for people with a percent predicted FVC of 50 - 80%, subject to ongoing NICE appraisal)	Same as final scope issued by NICE	The ERG notes that guidance on the use of nintedanib is now published (TA379) and nintedanib is recommended for people with a percent predicted FVC of 50 - 80%.[12] Therefore its inclusion as a comparator in this subgroup is appropriate. NAC, prednisolone and azathioprine were not considered to be relevant comparators for pirfenidone. The ERG notes that these are used as part of BSC in some patients but they are not expected to be disease-modifying.
Outcomes	Outcome measures to be considered include: • Pulmonary function parameters • Physical function • Exacerbation rate • PFS	Same as final scope issued by NICE	In addition to the outcomes listed in the scope, data are also presented for hospitalisations and all-cause discontinuations.

	• Mortality • Adverse effects of treatment • Health-related quality of life		
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies should be taken into account.	Same as final scope issued by NICE	For those analyses which incorporated the nintedanib PAS, results are provided in a confidential appendix.
Subgroups to be considered	If evidence allows, subgroup analysis by disease severity, defined by FVC (such as above and below or 80% FVC) and/or diffusing capacity for carbon monoxide, will be considered	Same as final scope issued by NICE. Subgroup analysis by FVC and DLco status at baseline was investigated, but the available data only allowed FVC to be assessed and reported in this submission.	In the economic analysis, the CS presents results for three populations: • mild to moderate IPF (described as the ITT population)

			• mild IPF (percent predicted FVC >80%) • moderate IPF (percent predicted FVC > 50% and ≤ 80%).

3.1 Population

The population addressed in the CS is adults with mild to moderate IPF; this is in line with the final NICE scope. Harari and Caminati[13] describe how populations and outcomes compare between clinical trials and observational studies that describe real-life treatment. The studies described by Harari and Caminati include single centre studies, such as the UK named patient programme which existed prior to TA282,[14] and international collaborative registries, such as PASSPORT which included UK sites.[15] They conclude that although the profile of patients treated with pirfenidone seems to be quite similar all over the world, patients treated in real-life scenarios differ from those treated in RCTs as real-life patients often have comorbidities, more severe disease, take concomitant medications and have a higher mortality.[13]

In terms of the patients excluded from the three main trials, the ASCEND trial appears to have been more restrictive as it excluded a larger proportion of patients following screening, with only 36% of those screened undergoing randomisation. The proportions of screened patients included in CAPACITY 1 and CAPACTIY 2 were 61% and 56%, respectively. The ASCEND and CAPACITY 1 and 2 trials all excluded patients with obstructive pulmonary disease (asthma or COPD) and patients with significant comorbidities such as a history of unstable or deteriorating cardiac or pulmonary disease (other than IPF). However, clinical advisors to the ERG stated that they would still treat with pirfenidone if there was evidence of asthma or COPD, provided the patient met the treatment criteria specified in TA282 (i.e. a predicted FVC between 50% and 80%). They also stated that many of the patients treated in routine clinical practice had comorbidities. This suggests that the key clinical trials for pirfenidone excluded some patients who would be treated in clinical practice.

In terms of disease severity, the proportion of patients with mild IPF, (i.e. a percent predicted FVC above 80%) was around 25% according to the figures presented in the CS (see CS, page 114 and Table 67). Clinical advisors to the ERG commented that the proportion of patients with an FVC above 80%, in the absence of emphysema (which elevates FVC), varied somewhat across different areas of the UK but was more likely to be between 30% and 50%. It is therefore possible that the subgroup who present with mild IPF are under-represented within the trial populations. It was also noted that only one of the pirfenidone trials, CAPACITY 2, recruited patients from UK centres (3 of 110 centres were UK).

The final NICE scope also specifies that if evidence allows, subgroup analysis by disease severity, defined by FVC (such as above and below or 80% FVC) and/or diffusing capacity for carbon monoxide, should be considered. The statement of the decision problem (see CS, Table 1, page 18, reproduced in Table 1) states that subgroup analysis by FVC and DLco status at baseline was investigated, but the available data only allowed FVC to be assessed and reported in the CS. In the original CS, some subgroup analyses by percent predicted FVC were presented for a limited number of outcomes, but subgroup analyses were not presented for all the outcomes specified in the final NICE scope. In response

to a request for clarification from the ERG, the company provided additional subgroup analyses which examined subgroups defined by percent predicted FVC (> 80% versus ≤80%) for the outcomes of change in percent predicted FVC, overall survival, and PFS, as requested, and for two additional supportive outcomes (see clarification response,[10] questions A29 and A31).

In the economic analysis, the CS presents results for three populations: (1) mild to moderate IPF (described as the ITT population); (2) mild IPF (percent predicted FVC >80%), and; (3) moderate IPF (percent predicted FVC > 50% and ≤ 80%). No subgroups results are presented by DLco status.

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, than to consider the ITT population with nintedanib excluded as comparator.

3.2 Intervention

The intervention is pirfenidone, as per the final NICE scope. Pirfenidone is indicated in adults for the treatment of mild to moderate IPF.[1] The mechanism of action of pirfenidone has not been fully established, however, existing data suggest that pirfenidone exerts both antifibrotic and anti-inflammatory properties.[1]

The previous appraisal of pirfenidone for treating idiopathic pulmonary fibrosis (TA282) recommended pirfenidone as an option only in patients with a percent predicted FVC of between 50% and 80%, which is a subgroup of the population covered by its marketing authorisation.[2] It also recommended that treatment *"should be discontinued if there is evidence of disease progression (a decline in per cent predicted FVC of 10% or more within any 12 month period)."*[2]

The recommended daily dose of pirfenidone for patients with IPF is three 267mg capsules three times a day (a total of 2,403 mg per day).[1] The capsules are taken orally with food.[1] Does adjustments and treatment interruptions are allowed to manage adverse events with re-escalation to the recommended daily dose as tolerated.[1] Treatment with pirfenidone should be initiated and supervised by specialist physicians experienced in the diagnosis and management of IPF.[1] Pirfenidone is linearly priced, with pack size costs for 267mg capsules of £501.92 for 63, £2,007.70 for 252 and £2,151.10 for 270.[4] The cost per day for the licensed dose of 2,403mg per day is £71.70 at the list price.[16]

3.3 Comparators

The comparators listed in Table 1 of the CS are BSC and nintedanib. These comparators are consistent with those defined in the final NICE scope. Nintedanib is only a comparator for people with a percent predicted FVC of between 50% and 80%. This is appropriate as this is the population covered by the recommendation for nintedanib in TA379.[12] However, it should be noted that nintedanib is indicated in

"adults for the treatment of IPF",[17] and the restriction of nintedanib as a comparator to patients with a percent predicted FVC of between 50% and 80% is due to the treatment criteria defend in the TA379, which match those defined for pirfenidone in TA282. The stopping criteria for nintedanib in TA379 also match those for pirfenidone in TA282.

BSC is defined in the CS as information and support, symptom relief, management of comorbidities, withdrawal of therapies suspected to be ineffective or causing harm, end of life care, oxygen therapy and/or pulmonary rehabilitation. The ERG and its clinical advisors considered this to be an appropriate description of BSC in current UK practice. The clinical advisors to the ERG also noted that BSC may vary internationally, particularly in countries without universal access to healthcare, and therefore the BSC received by non-UK trial participants may not reflect UK current practice.

Clinical advisors to the ERG were also asked whether any other therapies are currently used in the UK. As discussed in Section 2, the clinical advisors to the ERG noted that NAC is used off-license in some patients for symptom relief as part of BSC, but that it is not expected to be disease-modifying. They also reported that a few patients are currently managed with prednisolone or azathioprine, but again these treatments are not expected to be disease-modifying. NAC, prednisolone and azathioprine were not considered to be relevant comparators for pirfenidone by the clinical advisors to the ERG and were not included as comparators in the final scope.[3] The ERG therefore agrees with the exclusion of NAC, prednisolone and azathioprine from the list of relevant comparators.

Within the economic analysis nintedanib and BSC have been included as comparators for the subgroup of patients with moderate IPF (percent predicted FVC of 50 – 80%) and BSC has been included as a comparator for the subgroup of patients with mild IPF (percent predicted FVC >80%). The ERG considers the comparators chosen for the mild and moderate subgroups to be appropriate.

For the economic analysis considering the ITT population, which includes patients with both mild and moderate IPF, only BSC is included as a comparator, even though nintedanib is a valid comparator for the subgroup with moderate IPF. 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, than to consider the ITT population with nintedanib excluded as comparator.

3.4 Outcomes

The outcomes reported in the CS match those described in the final NICE scope.[3] The outcomes presented it the CS are discussed in turn.

3.4.1 Pulmonary function

A number of pulmonary function measures are reported in the CS including;

- mean change in percent predicted FVC/VC from baseline,
- mean change in FVC/ VC (mL)
- decline of ≥10 % in percent predicted FVC
- Mean change in percent predicted DLco
- Mean change in DLco (mL)

Mean change from baseline in predicted FVC/VC and mean change in FVC/VC (L) were included as continuous outcomes in the NMA. A decline of ≥10% in percent predicted FVC was included as a binominal outcome in the NMA. Outcomes relating to DLco were only reported for the direct comparison of pirfenidone against placebo.

The pulmonary function outcome which forms the main focus of the submission is FVC. FVC is an accepted trial endpoint for IPF, and one that has been widely used in trials to date.[18,19] It is widely recognised that the change in FVC over time, rather than the absolute FVC, is the outcome of interest, and a change of ≥10% appears to be accepted as being sufficient to define a true change.[19]

The CS cites evidence to support the claim that FVC is a good surrogate for survival, with a ≥10% decline in percent predicted FVC having been shown to be predictive of higher mortality in a number of studies and smaller changes (5-10%) in percent predicted FVC having been shown to be predictive of mortality in a smaller number of more recent studies (see CS, page 201).

The CS cites one study showing that there is a moderate correlation between changes in percent predicted FVC and changes in a disease specific health-related quality of life (HRQoL) measure, (Spearman correlation coefficient of -0.32), but the correlation between absolute values for percent predicted FVC and HRQoL is weaker (Spearman correlation of -0.16).[20] The ERG notes that whilst some evidence on the validity of FVC as a surrogate for mortality and HRQoL is presented, a systematic search does not appear to have been conducted as other relevant papers presenting data on the correlation between FVC and HRQoL have not been summarised.[21,22] However, the ERG notes that in the appraisal of ninetadnib (TA282), the Appraisal Committee concluded that, *"although it had some limitations, percent predicted FVC is the most reliable and widely used measure of lung function in clinical practice."*[12]

Within the CS, data from trials which reported VC but not FVC have been combined with data from trials that reported FVC. This is justified in the CS by the statement that: *"…there is little difference between VC and FVC in subjects without obstructive pathology."* Clinical advisors to the ERG considered this statement to be reasonable. However, whilst the ASCEND and CAPACITY 1 and 2

trials and SP2 trials excluded patients with obstructive airway disease, the exclusion criteria for SP3 are not as clear regarding the exclusion of patients with COPD or emphysema (CS Appendices, Tables A5.1. to A5.4). Therefore, the ERG considers that the combination of VC data from SP3 with FVC data from the ASCEND and CAPACITY trials is questionable.

Although there are some data to suggest that DLco is a good prognostic indicator for mortality in IPF,[23, 24] it is not as well accepted as a clinical trial endpoint.[18, 19] Clinical advisors to the ERG agreed that DLco is harder to measure and is more variable than FVC. The variability of DLco has commonly been recognised to be as high as 15%,[19] whereas the minimal clinically important difference for FVC is reported to be between 2% and 6%.[20] The ERG therefore concludes that whilst DLco may provide important relevant information in clinical practice, it is reasonable for the CS to focus on FVC as the main measure of pulmonary function as it is more accepted as a reliable outcome in a clinical trial setting.

The clinical advisors to the ERG also noted that there is up to 10% variation in FVC testing in real-life clinical settings and therefore when using a >10% decline in FVC to define disease progression, this should not be based on a single FVC reading and any decline should be confirmed as not being due to a temporary and reversible infection.

3.4.2 Physical functioning

The measure of physical functioning reported is the 6 minute walking distance (6MWD). Results are reported both for the mean change in 6MWD from baseline and for a categorical analysis of change from baseline using a threshold of a decrement of ≥50m. Mean change in 6MWD from baseline was included as an outcome in the NMA but loss of ≥50m in 6MWD was only reported for the direct comparison of pirfenidone against placebo.

In the appraisal of nintedanib (TA282), the Appraisal Committee heard from clinical experts that the 6MWD was an unreliable measure.[12] However, in the previous appraisal of pirfenidone, the Committee accepted the use of 6MWD as a covariate to predict survival in the microsimulation model.[3] This opinion is supported by an analysis by du Bois 2011, which showed that a decrement in 6MWD of greater than 50 metres over 24 weeks was associated with a HR for overall mortality at 1 year of 4.27 (p=0.001) when compared with a decrement of less than 25 metres.[25] However, the statistical significance of a decrement of greater than 50 metres when compared to a decrement of between 25 and 50 metres was not demonstrated.[25] Therefore, a decrement of more than 50m in 6MWD may not result in a statistically significantly higher risk of mortality compared with a decrement of less that 50m in 6MWD. The same study also found moderate correlations between changes in 6MWD and changes in disease-specific HRQoL measures which were statistically significant.[25] The ERG notes that the CS

states that the minimal clinically important difference (MCID) for the 6MWD was estimated to be 24-45 metres and therefore differences in the proportions experiencing a decrement of ≥50m and mean differences in 6MWD of ≥50m are likely to be clinically significant.

3.4.3 Exacerbation rate

Acute exacerbations are reported, however, the CS states that the outcome was defined differently across the trials and was not collected systematically in all trials. Acute exacerbation rate was included in the NMA.

In the nintedanib appraisal (TA282), the Committee concluded that exacerbations are an important clinical event, but can be difficult to define, particularly in trials.[12] In the company's clarification response (see clarification response,[10] question A15), the company states that acute exacerbations are notoriously difficult to diagnose, there is no universally agreed definition, and exacerbations meeting the strict definitions employed in trials are rare (<1% in the nintedanib trials). Clinical advisors to the ERG believed that this is because the definitions of acute exacerbations used in trials generally require other causes of respiratory symptoms, such as infection, to be ruled out, but this is a very restrictive definition as it is very hard in practice to rule out infection as a cause. However, in clinical practice, patients experience periods of acute worsening of symptoms with breathlessness that needs treatment and these are recognised by clinicians as acute exacerbation even though they may not meet the strict criteria applied in the trials.

3.4.4 Progression-free survival

Progression-free survival (PFS) is reported as per the NICE scope, however as noted in the CS, the definition of PFS varied between studies. PFS was included in the NMA, but this involved combining data from trials which used different definitions. Where possible, the data available were re-analysed to provide estimates using a consistent definition (that used in the ASCEND trial), but this was not possible for all of the trials included in the NMA. The various definitions for which data are presented are summarised in.Table 2.

Table 2 Summary of definitions used for progression-free survival

Trial	Definition specified in the final trial protocol	Other definitions for which results are provided [a]
ASCEND	confirmed ≥10% decline from baseline in %FVC, or confirmed ≥50 m decline from baseline in 6MWT distance, or death	Definition used in CAPACITY trials Definition(s) [b] used in SP3 / PANTHER
CAPACITY 1 and 2 [c]	confirmed ≥10% decline in percent predicted FVC, or ≥15% decline in percent predicted DLco or death	Definition used in ASCEND Definition(s) [b] used in SP3 / PANTHER
SP3	decline of 10% or more in VC or death	
PANTHER	decline of 10% or more in FVC or death	
INPULSIS trials	None pre-specified	Definition from CAPACITY [d]
TOMORROW	None	None

[a] in the CS or in the company response to the clarification request (clarification response,[10] question A33)
[b] SP3 used VC and PANTHER used FVC but the description of the re-analysis using this definition in Table 14 of the response to the clarification request simply states FVC/VC
[c] The definition in the protocol for the CAPACITY trials was updated by a protocol amendment so the definition in the final protocol is recorded here
[d] taken from the nintedanib company's submission for TA282[26]

3.4.5 Mortality

A number of measures are reported for mortality including all-cause mortality, IPF-related mortality and treatment-emergent IPF-related mortality. All-cause mortality and IPF-related mortality were included in the NMA but treatment-emergent IPF-related mortality was only included in the direct comparison of pirfenidone against placebo.

In the CS, treatment-emergent mortality was defined as occurring between randomisation and 28 days after the last dose of study drug. Treatment emergent IPF-related mortality is defined as a secondary efficacy outcome in the ASCEND protocol. In the protocols for the CAPACITY trials, deaths are described as a safety outcome.[27] Definitions are provided for the terms 'treatment-emergent' and 'IPF-related', but treatment-emergent IPF-related mortality is not specifically defined as an outcome.[28, 29] The clinical advisors to the ERG considered that all-cause mortality was the most important outcome for patients with IPF.

3.4.6 Adverse events

Adverse events (AE) of treatment are reported from the pirfenidone clinical trial programme in Section 4.1.12 of the CS and the AE data for nintedanib applied in the model are described in Section 5.3 of the CS but AEs are not reported systematically for the nintedanib studies. For ASCEND and CAPCATIY 1 and 2, the AEs summarised in Tables 60 and 61 of the CS, were treatment-emergent AEs with 'treatment-emergent' being defined as occurring after first dose and within 28 days after the last dose of study treatment. Additional data on AEs that led to discontinuation were provided in response to a clarification request (see clarification response,[10] question A24). Additional data on treatment-emergent serious adverse events reported in ≥2 patients were also provided in response to a clarification request (see clarification response,[10] question A25 and clarification response addendum,[30] question A28). For SP2, published AE data were presented, however additional summaries on serious AEs and AEs that led to discontinuation could not be provided by the company due to restrictions on access to data from this study.

The NMAs reported in the CS did not include AEs; however, additional NMAs were presented in the clarification response (see clarification response,[10] question A39) for the AEs of diarrhoea, rash, discontinuation of treatment due to AEs and serious cardiac AEs. The ERG considers it reasonable for additional NMAs to be presented for diarrhoea, rash and serious cardiac AEs as these data are useful for informing the indirect comparison with nintedanib within the company's model.

3.4.7 Health-related quality of life

EQ-5D data were not collected in the CAPACITY 1 and 2 or ASCEND trials (see CS, page 224).

CAPACITY 1 and 2 measured HRQoL using the St George's Respiratory Questionnaire (SGRQ); these data are reported in the CS. The change from baseline in the total SGRQ score was also included as an outcome in the NMA. A recent article examining the psychometric properties of the SGRQ in patients with IPF concluded that whilst it was not developed specifically for use in patients with IPF, and further research is needed to confirm the SGRQ's utility in IPF, at present, *"the balance of data suggests that the SGRQ may be a suitable secondary endpoint for measuring HRQoL in therapeutic trials of IPF."*[21] The MCID for the SGRQ in patients with IPF is reported to be 7 for the total SGRQ score.[22]

CAPACITY 1 and 2 and ASCEND measured HRQoL using the University of California San Diego Shortness of Breath Questionnaire (UCSD SOBQ); these data are reported in the CS. The mean change in the UCSD SOBQ was also included in the NMA. The MCID for UCSD SOBQ in patients with IPF is reported to be in the range of 5 to 11 points.[31]

WHO QOL data were also collected in the CAPACITY studies (see CS, page 224), but the results are not presented in the section reporting HRQoL outcomes (CS, pages 109 to 112).

3.4.8 Additional outcomes not specified in the scope

All-cause discontinuations and hospitalisations are reported in addition to the outcomes specified in the final NICE scope. All-cause discontinuations were included as an outcome in the NMA but hospitalisations were not.

3.4.9 Inclusion of outcomes in the indirect comparison

The majority of the outcomes were reported for both the direct comparison with placebo from the pirfenidone clinical trial programme and for the indirect comparison with nintedanib from the NMA. Outcomes addressed in the submission but not included in the NMA were DLco, treatment emergent IPF-related mortality, categorical change in 6MWD (decline of more or less than 50m), and hospitalisations.

3.5 Other relevant factors

Patient Access Schemes were agreed for pirfenidone at the time of TA282 and for nintedanib at the time of TA379. In both cases, the technologies were recommended only when the technology is provided with the discount agreed in the PAS. The company submitted a revised PAS which was accepted by the Department of Health. Further details on the PAS can be found in the confidential appendix.

No equality issues were raised in the CS.

4 CLINICAL EFFECTIVENESS

This chapter presents a summary and critique of the reviews submitted by the company on the efficacy and safety of pirfenidone in adults with mild to moderate IPF. The critique was performed following the principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and checklist.[32]

4.1 Critique of the methods of review(s)

The CS[4] reports the methods and results of three separate reviews:

(i) A review of the efficacy evidence from RCTs (see CS, [4] Sections 4.1-4.10);

(ii) A review of the efficacy and safety evidence from non-randomised and non-controlled studies (see CS,[4] Section 4.11), and;

(iii) A review of safety evidence from RCTs and a non-randomised study (see CS,[4] Section 4.12).

Each review applied slightly different inclusion criteria depending on the intended analysis and the included study designs.

The main review of efficacy evidence from RCTs was a generally well-reported systematic review. Following a request for clarification from the ERG regarding certain process elements adopted by the company, the ERG considered the review to be generally sound (see clarification response,[10] questions A1-A7). The key trials were listed as ASCEND (Phase III),[33, 34] CAPACITY 1 & 2 (Phase III),[35-37] SP3 (Phase III),[38] and SP2 (Phase II).[39] All studies compared pirfenidone with placebo. The NMA included five additional relevant RCTs (further details are provided in Section 4.6).

The review of the efficacy evidence from non-randomised and non-controlled studies consisted of a single open-label, non-controlled extension study (RECAP),[40] which was designed to assess long-term safety with some efficacy outcomes listed as secondary outcomes, plus data from six registries. This review was not considered to be a systematic review because it was unclear how the evidence was identified, selected and relevant data extracted; no inclusion or exclusion criteria were provided; and a list of excluded studies or registries was not provided. Quality assessment of the RECAP study[40] was not performed by the company.

The review of the safety evidence was also not considered by the ERG to be a systematic review because it was unclear from the original submission how the included non-RCT evidence, RECAP[40], plus the addition of a new study, PIPF-002[41], were identified and selected, no detailed inclusion or exclusion criteria or details of data extraction were provided, and a list of potentially relevant excluded studies was not provided.

4.1.1 Searches

The company conducted a systematic literature review search for evidence on the comparative efficacy and safety of interventions in IPF in April 2015.

The ERG notes that the search strategy was developed using the PICOS (patient – intervention – comparator – outcome – study types) elements of the systematic review. The strategy was structured to search for the concepts:

1. Idiopathic pulmonary fibrosis AND randomised controlled trials

   OR

2. Pirfenidone

The following sources were searched:

- Cochrane Central Register of Controlled Trials (CENTRAL)
- Cochrane Database of Systematic Reviews (CDSR) Embase
- Database of Abstracts of Reviews of Effectiveness (DARE)
- Embase 1974 to 2015 November 16
- Health Technology Assessment (HTA) database
- Ovid MEDLINE® in-process and other non-indexed citations
- Ovid MEDLINE® 1946 to present
- European Respiratory Society congress abstracts
- British Thoracic Society congress abstracts
- American Thoracic Society conference abstracts
- World Association for Sarcoidosis and Other Granulomatous Disorders conference abstracts

Reference lists of identified relevant studies, papers and review articles were also hand-searched for potentially relevant additional studies that may have been missed in the database searches. Update searching or forward citation searching was not reported to have taken place.

The CS[4] reports that the following databases were omitted from the search, despite having been included in the searches for the original NICE Technology Appraisal guidance on Pirfenidone for treating idiopathic pulmonary fibrosis (2013)[42]:

- World Association for Sarcoidosis and Other Granulomatous Disorders conference abstracts
- Allied and Complementary Medicine Database (AMED)
- British Nursing Index (BNI)
- Health Management Information Consortium (HMIC)
- PsycINFO
- Journals@OVID Full text
- Cochrane Methodology Register
- NHS Economic Evaluation Database (NHS EED)
- About the Cochrane Collaboration

The CS[4] states that the reason for excluding these databases was because their focus was not considered appropriate for the objectives of this specific systematic review. The ERG agrees with the decision to omit these databases based on the specific focus of these databases and they are not amongst the minimum databases suggested by the NICE Guide to the Methods of Technology Appraisal 2013[43] or the Centre for Reviews and Dissemination (CRD) guidance.[44]

When attempting to reproduce and verify the company's searches, the ERG identified a number of potentially relevant studies that met the inclusion criteria via searching the Web of Science.[45-47]

The searches were limited to information published, added to the databases, updated or indexed from January 2011 onwards. This date limit was applied because the original InterMune NICE STA submission searches were conducted in October 2011.[27] However, the ERG noted that the approach to searching differed from the approach that was undertaken for the original submission, which searched for: IPF AND pirfenidone AND RCTs.

The searches were comprehensive and the reporting of the search strategies is clearly reported, reproducible and transparent. The ERG obtained a similar number of records when re-running the searches. The ERG did not identify any errors in the execution of the searches in relation to Boolean or database specific syntax operators and the translation of the strategy across all of the databases from Medline is consistent.

The ERG also re-ran an amended version of the search for pirfenidone on Medline and Embase changing the fields from .ti,ab,kf,rn (Medline) and .ti,ab,kw,rn (Embase) to the more sensitive .af search field. This did not however affect the results.

Additional studies that were published after the systematic searches had been conducted were included in the meta-analysis and the PRISMA chart states that 23 studies were identified via 'other sources' but does not specify the methods of retrieval of these studies. The company stated in their clarification response that the included studies that were published after the search date were obtained via 'internal analyses' (see clarification response,[10] question A2). However, the ERG would recommend update searching or forward citation searching in order to maximise the transparency of reporting and reduce the risk of confirmation bias.

The ERG found that, despite these omissions, the numbers of results retrieved by the company were in accordance with the results obtained when all terms were entered correctly and the searches were re-run by the ERG.

The CS[4] does not report whether a published search filter was utilised in order to identify RCTs of IPF. The search filter appears to be a slightly modified version of the Cochrane highly sensitive search filter. The company stated in their clarification response (see clarification response,[10] question A3) that this filter had been amended to increase the sensitivity of the search.[10, 44]

The reporting of the conference abstract searches in the European Respiratory Society (ERS) Annual Conference Abstracts contains information about how many retrieval hits were obtained and how many records were retrieved for further consideration. However, the searches conducted in the ERS Annual Congress and Conference advanced search feature only contains information about how many results were retrieved and not how many were considered.

The ERG queried the lack of searches for ongoing and unpublished clinical trials in research registers including the metaRegister of Controlled Trials, the EU Clinical Trials Register and the World Health Organization. The company agreed that this was an oversight and searched ClinicalTrials.gov, ICTRP AND PharmNet Bund and provided the results in Appendix A of their clarification response.[10]

The searches conducted by the ERG in research registers identified an additional, potentially relevant RCT published in 2015, which compared pirfenidone plus N-acetylcysteine (NAC) with placebo and NAC in adult patients with mild or moderate IPF (percent predicted FVC at baseline was 75.55±14.72 in the pirfenidone and NAC group and 79.07±18.25 in the NAC and placebo group) (Huang *et al.*

2015[48]). Details and results from this trial have therefore been reported as supporting evidence by the ERG. This was also identified in the additional references provided by the company after conducting searches of research registers as part of the company's response to the ERG's clarification questions.

The ERG's view is that it is likely that all relevant RCTs will have been identified from the searches described in the CS[4] and the company response to the clarification request.[10] The ERG obtained a similar number of records when re-running the searches. No search strategies were reported for AEs; however, the ERG believes that searching for pirfenidone as a standalone concept maximises the sensitivity of this search and would be likely to capture any potentially relevant information in relation to AEs.

4.1.2 Inclusion criteria

The inclusion criteria for the review of pirfenidone RCTs are described in Section 4.1 of the CS[4] (Table 8, page 53) and reproduced in Table 3. These criteria describe RCTs measuring the efficacy and safety of pirfenidone compared with nintedanib or BSC (placebo) in adult patients with mild or moderate IPF. The five RCTs satisfying these criteria are: ASCEND (Phase III),[34] CAPACITY 1 & 2 (Phase III),[49] SP3 (Phase III),[38] and SP2 (Phase II).[39] All of these trials compared pirfenidone with placebo. These RCTs included four different doses of pirfendione: 2,403mg per day, 1,197mg per day, 1,800mg per day and 1,200mg per day. The NMA to evaluate efficacy applied different criteria (see CS,[4] Table 38, page 123) and is covered in detail in Section 4.6 of this report.

The review of the efficacy evidence from non-randomised and non-controlled studies did not specify any inclusion criteria (see CS,[4] Section 4.1). This review reported a single open-label, non-controlled extension study, RECAP,[40] whose participants were recruited from the ASCEND[34] and CAPACITY trials.[49] Further evidence was reported from the Edinburgh registry, INOVA registry and the EuroIPF registry, as well as three additional, "supportive" registries: CPRD, Strand *et al*[41] and Kondoh *et al*.[50] According to the inclusion criteria outlined in Section 4.1 of the CS,[4] non-randomised studies were explicitly excluded. The search conducted for the clinical efficacy review would have enabled the identification of the RECAP[40] non-RCT, but it is unclear whether additional, relevant evidence might have been excluded.

The inclusion criteria for the review of safety evidence from RCTs and non-randomised studies were not specified. The safety review included the five pirfenidone RCTs from the main clinical efficacy review, as well as the non-randomised studies RECAP[40] and an additional non-randomised, non-controlled study, PIPF-02.[41] However, as noted above, the methods by which these non-randomised studies were identified and the criteria by which they were selected, and others were excluded, are not reported.

Table 3: Inclusion and exclusion criteria for the pirfenidone RCT direct comparison clinical efficacy systematic review (reproduced from CS, Section 4.1, Table 8, page 53,)[4]

	Inclusion criteria	Exclusion criteria
Population	Adults (aged 18 or older) with suspected or diagnosed IPF	Studies of children and young people <18 years Studies of people with a diagnosis of pulmonary fibrosis as a complication of either of the following: • Connective tissue disorders • A known exogenous agent (for example, drug induced disease or asbestosis)
Intervention	Pirfenidone	Any studies not containing pirfenidone
Comparators	Any comparator: • Best supportive care* (placebo) • Nintedanib	N/A
Outcomes	**Pulmonary function parameters** • Lung capacity (VC/FVC) • Categorical declines in FVC • Gas transfer (carbon monoxide diffusing capacity [DLco]) **Physical function** • Physical functioning (6MWD) **Exacerbation rate** • Hospitalisations • Acute exacerbations **Progression-free survival** **Mortality** • All-cause mortality • IPF-related mortality **AEs of treatment** **HRQoL** • St George's Respiratory Questionnaire (SGRQ) • University of California, San Diego (UCSD) Shortness of Breath Questionnaire (SOBQ) • EuroQoL five dimensions questionnaire (EQ-5D)	• Anticoagulation for the treatment of pulmonary hypertension • Treatment of lung cancer • Lung transplantation other than timing and referral
Study design	• Studies in humans • Phase II or III RCTs • Studies published as abstracts, conference presentations or press releases were eligible if adequate data were provided • Systematic reviews of RCTs**	• Cross-over RCTs
Language	No language limits	No language limits

*Best supportive care is defined as information and support, symptom relief, management of comorbidities, withdrawal of therapies suspected to be ineffective or causing harm, end of life care, oxygen therapy and/or pulmonary rehabilitation

**Systematic reviews were eligible for inclusion as a source of references to primary studies

4.1.3 Critique of study selection and data extraction

Following an ERG request for the company to clarify the processes undertaken, the ERG was satisfied that standard systematic review good practice was followed in study selection: relevant papers were independently selected for inclusion at title, abstract and full text stage by two reviewers, with any discrepancies between reviewers resolved through discussion or the intervention of a third reviewer (see clarification response,[10] question A4). In a first screen, "obviously irrelevant" studies were excluded by a single information specialist (see clarification response,[10] question A4).

No information was given in any of the reviews regarding the data extraction process (for example, the number of reviewers involved, or actions taken to minimise error). This was addressed in response to clarification requests from the ERG, in which the company detailed standard processes for data extraction in systematic review (see clarification response,[10] question A5). Data extraction was performed by one reviewer and independently checked for errors against the original trial report by a second reviewer. Any discrepancies were resolved through discussion or through the intervention of a third reviewer.

During the clarification stage, discrepancies and inadequacies in some of the numbers reported in the PRISMA flowchart were acknowledged and addressed by the company, and an updated PRISMA flowchart was provided (see clarification response,[10] question A6).

4.1.4 Quality assessment

For the review of clinical efficacy evidence, the company conducted a critical appraisal of the five pirfenidone trials using a version of the Cochrane risk of bias assessment tool (see CS,[4] Section 4.6 and Appendix 6). The process was conducted according to standard systematic review practice, by two reviewers working independently, with any discrepancies resolved by discussion or reference to a third reviewer (see CS,[4] Appendix 6). The CS concluded that all five trials were at "low risk of bias" across the domains assessed, although the adequacy of randomisation and blinding was assessed as "unclear" for the SP3 trial.[38]

The ERG accepts these assessments for the ASCEND[33, 34] and CAPACITY trials,[35, 37, 51] for the domains of selection bias (randomisation, allocation concealment); performance and detection bias (blinding); and attrition bias (drop-out, ITT analysis and management of missing data). However, the ERG disagrees with assessments regarding reporting bias and other types of bias, especially given the absence of adequate information concerning some analyses and some secondary outcomes in both the publicly-available protocols for each trial from the clinical trials register (https://clinicaltrials.gov/ct2/home) and those protocols made available alongside the final publications or provided by the company in response to requests by the ERG (see clarification responses[10], questions A8-A10). For example, the SGRQ

outcome measure that is reported in the CS[4] is absent from all forms of protocol, as well as the actual CAPACITY trials publication[35, 37, 51] (although this outcome is listed in the CSR).

The effect of the "intrinsic variability in rates of FVC decline" acknowledged in the CAPACITY trials' publication,[49] and the company's clarification response (question A26:[10] "the natural variability in rates of FVC percent predicted decline of this heterogeneous disease"), which might explain differences in outcomes both within and across trials, must also be taken into account as a potential moderator influencing results.

Overall, however, the ERG assessed the potential risk of bias in ASCEND[33, 34] and CAPACITY 1 & 2[35, 37, 51] to be low or low-to-moderate. The details of the ERG assessment are provided in Table 4.

The SP3[38] and SP2[39] trials, by contrast, are at a higher or more unclear risk of bias across many domains compared with the ASCEND[33, 34] and CAPACITY trials,[35, 37, 51] principally because of the inadequacy of the information contained within the published manuscripts and the protocols provided by the company in response to a request by the ERG (see clarification response,[10] question A12). These issues particularly affect selection, detection and attrition bias; the last named on account of the smaller sample sizes, the rates of attrition and the application of the last observation carried forward (LOCF) method to impute missing data, which might potentially overestimate treatment effect in a progressive disease such as IPF[51] (see Table 4).

Finally, the supporting trial reported by Huang *et al.*,[48] was generally found to be at moderate risk of bias across most domains as a result of the lack of detail within the available protocol and the publication.

Table 4: ERG risk of bias assessment (Cochrane tool): ASCEND, CAPACITY 1 & 2, SP3 and SP2

Risk of bias	ASCEND[33, 34]	CAPACITY 1[36, 49]	CAPACITY 2[33, 49]	SP3[38]	SP2[39]	Huang 2015[48]
Selection bias	LOW Randomisation codes were generated by computer with the use of a permuted-block design, and the study drug was assigned by means of an interactive voice-response system. Protocol: 3, page 38: Patients will be randomised at the Day 1 Visit (see Section 4.3.4.1) in a 1:1 ratio to receive either pirfenidone 2403 mg per day or placebo equivalent using an automated system. All randomisation codes will be generated by a statistician independent of the trial conduct.	LOW The randomisation code (permuted block design with five patients per block in study 004 and four per block in study 006) was computer generated, stratified by region, by an independent statistician. Study centres, using an interactive voice response system, assigned study drug bottles to patients. The independent statistician had no role other than assignation of the randomisation code and study drug bottle numbers. All personnel involved in the study were masked to treatment group assignment until after final database lock.		UNCLEAR A multicentre, double-blind, placebo-controlled, randomised phase III clinical trial, page.821 etc.; Eligible patients were allocated to three groups: high dose (1,800 mg/day), low dose (1,200 mg/day) and placebo, in a ratio of 2:1:2, respectively, with a modified minimisation method, including some random allocation based on biased coin design to balance baseline SpO_2, page 822): insufficient information and he protocol does not provide any specific information[52]	MODERATE Patients were randomly assigned into pirfenidone or placebo (2:1) groups using a modified permuted-block randomisation method with block sizes of six.(page 1042), but it is not stated who does this. Investigators? Independent body? The protocol does not provide any specific information [53]	MODERATE “Patients were randomly assigned into pirfenidone or placebo (1:1) groups using a modified permuted-block randomisation method with block sizes of 4”, but it is not stated who does this. Investigators? Independent body? The protocol does not provide any specific information (https://clinicaltrials.gov/ct2/show/record/NCT01504334)
Performance bias	LOW Publication main text provides no information; page 31 NEJM protocol: Patients will receive blinded study treatment from the time of randomization until the week 52 Visit. Page 37: There will be 270 capsules per bottle, which will be labeled for investigational use only. Pirfenidone 267-mg and placebo will be supplied in opaque, hard, white gelatin capsules that are visually indistinguishable. 3, page 38: Pirfenidone and placebo will both be supplied in capsules that are visually indistinguishable. Pirfenidone and placebo packaging and labeling will be identical. There was no evaluation of blinding.	LOW All personnel involved in the study were masked to treatment group assignment until after final database lock. Available information from publication and protocols is too limited to give this a “low risk of bias” assessment, but sufficient information was given in the CSR. There was no evaluation of blinding.		LOW Allocation and blinding covered in detail in the protocol, sections 14.3.1, 14.3.2, 14.3.3 and 14.3.5 [52]	LOW Allocation and blinding covered in detail in the protocol, sections 6.3.1, 6.3.2 and 6.3.4 [53]	MODERATE Matching placebo tablets, but no other details of blinding and no evaluation of blinding
Detection bias	LOW Central reviewers at Biomedical Systems, who were unaware of study-group assignments, evaluated all FVC results for adequacy and repeatability, according to the criteria of the American	LOW Mortality was pre-specified as an exploratory endpoint, and death related to idiopathic pulmonary fibrosis was assigned by investigators masked to assignment. ASCEND publication: “The primary cause of death and its relation to idiopathic pulmonary		LOW / MODERATE Protocol states that: “14.3.5 Blindedness will be maintained with respect to all study personnel except the study drug allocation manager” [52], but publication acknowledges limitation of ,“The lack of a central pathology review” (page 824); plus	UNCLEAR Protocol indicates that outcome assessors were unblinded: “6.3.4 Blindedness will be maintained with respect to all study personnel except the study drug allocation manager	MODERATE Protocol states that outcome assessors were blinded, but there are no details

Risk of bias	ASCEND[33, 34]	CAPACITY 1[36, 49]	CAPACITY 2[33, 49]	SP3[38]	SP2[39]	Huang 2015[48]
	Thoracic Society The primary cause of death and its relation to idiopathic pulmonary fibrosis were assessed in a blinded fashion by an independent mortality assessment committee in the ASCEND trial and by the site investigators in the CAPACITY trials[33, 36, 49]	fibrosis were assessed in a blinded fashion … by the site investigators in the CAPACITY trials" [33, 36, 49]		problems with un-validated measure of lowest SpO2 during the 6MET	and the efficacy and safety evaluation committee."[53]	
Attrition bias	LOW / MODERATE 522 patients (94.1%) completed the study: 261 patients (93.9%) in the pirfenidone group and 261 patients (94.2%) in the placebo group. Study treatment was discontinued prematurely in 55 patients (19.8%) in the pirfenidone group and in 39 patients (14.1%) in the placebo group. Adherence to the study treatment was high; 237 patients (85.3%) and 256 (92.4%) patients in the pirfenidone and placebo groups, respectively, received at least 80% of the prescribed doses of the assigned study drug.	LOW / MODERATE 409 (94%) of 435 patients in CAPACITY 2 and 322 (94%) of 344 in CAPACITY 1 completed the study. 109 patients (14%) discontinued treatment prematurely: 13 (15%), 30 (17%), and 18 (10%) in the pirfenidone 1197 mg/day, pirfenidone 2403 mg/day, and placebo groups, respectively in CAPACITY 2; and 31 (18%) and 17 (10%) in the pirfenidone and placebo groups, respectively, in CAPACITY 1		MODERATE 30%+ rate of attrition and LOCF used to impute missing data (for a progression disease, this might overestimate treatment effect) if patient data were available for 4 weeks after the baseline (page 823)	MODERATE 20%+ rate of attrition and LOCF used to impute missing data (for a progression disease, this might overestimate treatment effect) For missing values, the principle of last observation carry forward was adopted (page 1042)	UNCLEAR Up to 16% attrition, but it is not clear from the publication or protocol how missing data were managed
Reporting bias	MODERATE Two primary, five secondary outcomes – only the basic primary outcome listed in NCT protocol; others in NEJM protocol, but SGRQ not in any protocol; plus acute exacerbations / hospitalisations – are recorded at Follow-Ups, but not specified as outcomes. Only pre-specified analyses listed in protocols relate to mortality.	MODERATE All protocol outcomes listed in primary publication ,[49] but SGRQ was not in any protocol and was not reported in the primary publication, but was only mentioned in CSR; plus acute exacerbations / hospitalisations are only reported as part of the "Worsening of IPF" composite outcome		LOW All of the outcomes reported in the protocol (Shinogi 2006[52]) were reported in the publication	LOW All of the outcomes reported in the protocol [53] were reported in the publication	LOW All of the outcomes reported in the protocol (https://clinicaltrials.gov/ct2/show/record/NCT01504334) were reported in the publication
Other bias	UNCLEAR "Intrinsic variability in rates of FVC decline" acknowledged as potential moderator of results, and possible explanation for differences across trials in certain outcomes (Noble 2011, pages 1767 and 1768) [49] . This represents a potential uncontrolled moderator of outcomes. Claim that this is controlled for by FVC and DLco eligibility criteria (more severe and progressive population) is questionable.	UNCLEAR "Intrinsic variability in rates of FVC decline" acknowledged as potential moderator of results, and possible explanation for differences across trials in certain outcomes (Noble 2011, pages 1767 and 1768) [49] . This represents a potential uncontrolled moderator of outcomes		MODERATE Acknowledged issue: (page 824 Taniguchi):[38] A selection bias, as patients enrolled in this study needed to be able to perform the 6MET at baseline in accordance with the protocol; the results in this selected group of patients with mild functional impairment may not therefore be applicable to all patients with IPF	UNCLEAR Trial discontinued early due to excessive rates of exacerbations in the placebo arm, so outcomes etc. were not measured at all planned time-points, only at 6 and 9 months Substantial links to study sponsor.	MODERATE Difficulty controlling for natural variability in IPF disease and speed of progression: *post hoc* analyses excluding patients with most substantial decline, produced different findings Some links to industry

Risk of bias	ASCEND[33, 34]	CAPACITY 1[36, 49]	CAPACITY 2[33, 49]	SP3[38]	SP2[39]	Huang 2015[48]
	Composite outcomes do not appear under Outcomes in protocol – first appearance is under efficacy analyses, page 60, 5.4.2 Efficacy Analyses: 5.4.2.1 Primary Efficacy Outcome Variable and Analysis in protocol analysis plan and represents a modification from the CAPACITY trials[33, 36, 49] – it does not appear as an outcome in the protocol or publication			with varying degrees of pulmonary symptoms and functional impairment.	Per protocol drop-outs based on the outcome measure (page 1042)	

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 an unlicensed dosage (1,800 mg/day). 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 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 $\geq$50% (in ASCEND[34] this had an upper limit of $\leq$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 5: 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=277) 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=278)	52 weeks
CAPACITY 1 (PIPF-006)[36, 49] n=344	International multi-centre		– 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	– 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%		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

SP2[39] n=107	Japan, multi-centre	Confident clinical and radiographic diagnosis of IPF (as per guideline consensus)	– Adequate oxygenation at rest (PaO2 70 mm Hg) and SpO2 ≤ 90% during exertion	– Coexisting pulmonary hypertension, asthma, tuberculosis, sarcoid, bronchiectasis or respiratory infection; – Comorbid conditions including malignancy, severe hepatic, renal, Diabetes Mellitus or cardiac disease	Pirfenidone 1800mg/day (n=72) Concomitant prednisone ≤10mg/day was allowed. The following immunosuppressants or other anti-inflammatory/ antifibrotic drugs were not allowed: cyclophosphamide, azathioprine, methotrexate, d-penicallimine, cochicine, erythromycin, IFNs, N-acetylcysteine, cyclosporine, tacrolimus and other investigational drugs for IPF.	Placebo (n=35)	36 weeks (trial terminated early due to adverse events)
SP3[38] n=275*	Japan, multi-centre	Confident clinical and radiographic diagnosis of IPF (as per ATS/ERS guideline consensus) No decrease in symptoms during the preceding 6 months	– O2 desaturation of 5% between resting SpO2 and min SpO2 during 6 min exercise test (6MET) – SpO2 >85% during 6MET (air).	– Coexisting pulmonary hypertension, asthma, tuberculosis, sarcoid, bronchiectasis or respiratory infection; – Comorbid conditions including malignancy, severe hepatic, renal, Diabetes Mellitus or cardiac disease	Pirfenidone 1800mg/day (n=108) Pirfenidone 1200mg/day (n=55) Concomitant corticosteroid ≤10mg/day (as the prednisone equivalent) was allowed. However, concomitant immunosuppressants or other investigational drugs for IPF were not allowed.	Placebo (n=104)	52 weeks

Huang 2015[48] n=76	China, multi-centre	The diagnosis of IPF was in accordance with evidence-based guidelines for the diagnosis and management of IPF published in 2011.	– percentage of predicted forced vital capacity (FVC) of at least 45%, – percentage of predicted carbon monoxide diffusing capacity – (DLCO) of at least 30%, and PaO2 of at least 50mmHg when – the patient is at rest and breathing room air	– aggravated dyspnea during the preceding 6 months; – currently in a period of acute exacerbation of IPF (AEIPF); – fasting blood glucose level of more than 11.1 mmol/L – comorbid conditions including malignancy, bleeding tendency, severe hepatic dysfunction or renal or cardiac disease; – use of immune-suppressants, antifibrotic drugs	Pirfenidone 1800mg/day (n=38) All patients were treated with 600 mg of N-acetylcysteine (NAC) 3 times daily as a baseline treatment.	Placebo (n=38)	48 weeks

**8 patients were excluded after randomisation for being ineligible; IPF: Idiopathic Pulmonary Fibrosis; FVC: Forced Vital Capacity; DLco: Diffusing capacity of the lungs for carbon monoxide; 6MWT: 6-minute walking test; ATS/ERS: American Thoracic Society/European Respiratory Society; PaO2: Partial pressure arterial oxygen; SpO_2: Blood oxygen saturation level; 6MET: 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 and death" does not appear in any protocol but appears to be a *post hoc* composite efficacy outcome in the CS[4] (see Table 6), which according to the company was used 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

original protocols and publications, are summarised in Table 6 and Table 7. The Huang *et al.*, trial[48] has been omitted from these tables because it is being used as supporting evidence only.

Table 6: 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 **and death** from baseline to week 52	Change in percent predicted FVC **and 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	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 7: 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 protocols or publication (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 (≥ 5 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

Outcome	**ASCEND**[33, 34]	**CAPACITY 1**[36, 49]	**CAPACITY 2**[35, 49]	**SP3**[38]	**SP2**[39]
Hospitalisations	No	Non-respiratory and *respiratory hospitalisations. Only the latter was listed in the protocols.		No	Respiratory hospitalisations
6MWD (6-Minute Walking Distance Test)	Defined as the change from Baseline to week 52 in distance walked during the 6-Minute Walk Test as measured in metres (m).	Defined as the change from baseline to week 72 in distance walked during the 6-Minute Walk Test as measured in meters (m).		The change in the lowest SpO_2 during the 6MET (the original primary endpoint, which was altered after the study started but before un-blinding, Taniguchi, page 822)	No
FVC/VC	No	No		No	Yes
SGRQ (St. George's Respiratory Questionnaire)	No	Yes. Not listed in protocols and not reported in the primary publication: a *post hoc* outcome measure.		No	
Dyspnoea using UCSD SOBQ)	The SOBQ is used to assess shortness of breath with various activities of daily living (for example, brushing ones teeth or mowing the lawn). Patients rated the severity of their shortness of breath experienced on an average day during the past week on a 6 point scale (0 to 5),with 0= not at all breathless, 4= severely breathless and 5= Maximally or unable to do because of breathlessness			No	
Gas transfer (DLco)	Excluded from this trial, see Clarification response,[10] question A9	The change from baseline in Percent Predicted Hemoglobin (Hb)-Corrected Carbon Monoxide Diffusing Capacity (DLco) of the Lungs.			

**Listed under the Worsening of IPF outcome in the CAPACITY 1 and 2 protocols; †Tertiary outcomes: PFS and change in the lowest SpO_2 during the 6MET were the designated secondary outcomes*

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, unlicensed doses of pirfenidone, 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[36]) 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 8: Characteristics of participants in ASCEND and CAPACITY 1 & 2 (reproduced from CS,[4] Table 16, pages 84-85)

Baseline characteristic	ASCEND[33, 34]		CAPACITY 2[35, 49]			CAPACITY 1[36, 49]	
	PFN (n=278)	PBO (n=277)	PFN (n=174)	PFN (1,197mg/d) (n=87)	PBO (n=174)	PFN 2,403mg/day (n=171)	PBO (n=173)
Age, mean years ± SD	68.4 ± 6.7	67.8 ± 7.3	65.7 ± 8.2	68.0 ± 7.6	66.3 ± 7.5	66.8 ± 7.9	67.0 ± 7.8
Male, n (%)	222 (79.9)	213 (76.9)	118 (68)	65 (75)	128 (74)	123 (72)	124 (72)
Percentage of predicted FVC, mean % ± SD	67.8 ± 11.2	68.6 ± 10.9	74.5 ± 14.5	76.4 ± 14.4	76.2 ± 15.5	74.9 ± 13.2	73.1 ± 14.2
Percentage of predicted DLco, mean % ± SD	43.7 ± 10.5	44.2 ± 12.5	46.4 ± 9.5	47.2 ± 8.2	46.1 ± 10.2	47.8 ± 9.8	47.4 ± 9.2
Dyspnoea score, mean ± SD	34.0 ± 21.9	36.6 ± 21.7	NR	NR	NR	NR	NR
Mean 6MWD, m ± SD	415.0 ± 98.5	420.7 ± 98.1	411.1 ± 91.8	417.5 ± 112.8	410.0 ± 90.0	378.0 ± 82.2	399.1 ± 89.7
Supplemental O_2 use, n (%)	78 (28.1)	76 (27.4)	29 (16.7)	15 (17)	25 (14)	48 (28)	49 (28)
HRCT definite IPF, n (%)	266 (95.7)	262 (94.6)	159 (91)	83 (95)	164 (94)	149 (87)	158 (91)
Surgical lung biopsy, n (%)	86 (30.9)	79 (28.5)	86 (49)	32 (37)	85 (49)	94 (55)	94 (54)
Time since IPF diagnosis, years ± SD	1.7 ± 1.1	1.7 ± 1.1	1.3 ± 0.96	1.4 ± 1.16	1.4 ± 1.12	1.2 ± 1.09	1.1 ± 1.04
Former smoker, n (%)	184 (66.2)	169 (61.0)	110 (63)	57 (66)	114 (66)	112 (66)	101 (58)
Pre-enrolment corticosteroid use, n (%)	6 (2.2)	2 (0.7)	14 (8.0)	10 (11.5)	9 (5.2)	22 (12.9)	17 (10.0)
Concomitant corticosteroid use, n (%)	82 (29.5)	101 (36.5)	38 (21.8)	24 (27.6)	52 (29.9)	42 (24.6)	50 (29.0)

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

Table 9: Characteristics of participants in SP2 and SP3 (reproduced from CS,[4] Table 16, pages 84-85)

Baseline characteristic	SP3[38]			SP2[39]	
	PFN (1,800mg/d) (n=108)	PFN (1,200mg/d) (n=55)	PBO (n=104)	PFN (1,800mg/d) (n=72)	PBO (n=35)
Age, mean years ± SD	65.4 ± 6.2	63.9 ± 7.5	64.7 ± 7.3	64.0 ± 7.1	64.3 ± 7.6
Male, n (%)	85 (78.7)	47 (85.5)	81 (77.9)	62 (86.0)	33 (94.0)
Percentage of predicted VC, mean % ± SD	77.3 ± 16.8	76.2 ± 18.7	79.1 ± 17.4	81.6 ± 20.3	78.4 ± 17.2
Percentage of predicted TLC, mean % ± SD	73.2 ± 16.5	72.4 ± 15.6	75.2 ± 15.7	78.5 ± 17.9	73.9 ± 16.4
Percentage of predicted DLco, mean % ± SD	52.1 ± 16.8	53.6 ± 19.1	55.2 ± 18.2	57.6 ± 17.2	57.7 ± 13.8
Lowest SpO_2 during 6MWT, mean % ± SD	89.0 ± 2.3	88.8 ± 2.4	89.0 ± 2.0	87.1 ± 3.9	87.1 ± 4.2
Desaturation <88% during 6MWT, n (%)	34 (31.5)	19 (34.5)	24 (23.1)	NR	NR
Mean $P(A\text{-}a)O_2$ ± SD	18.4 ± 11.3	16.9 ± 9.6	17.4 ± 9.7	NR	NR
Percentage of predicted SpO_2, mean % ± SD	89.0 ± 2.3	88.8 ± 2.4	89.0 ± 2.0	NR	NR
Mean PaO_2 at rest, mmHg ± SD	79.8 ± 10.2	81.6 ± 8.4	81.0 ± 9.5	80.3 ± 7.7	82.0 ± 17.6
Mean VC, mL ± SD	2400.8 ± 638.4	2437 ± 684.8	2472.3 ± 698.9	NR	NR
Surgical lung biopsy, n (%)	26 (24.1)	16 (29.1)	28 (26.9)	15 (21.0)	8 (23.0)
IPF diagnosis, n (%) ≤1 year 1-3 years >3 years	 38 (35.2) 29 (26.9) 41 (38.0)	 20 (36.4) 13 (23.6) 22 (40.0)	 41 (39.4) 25 (24.0) 38 (36.5)	 20 (28.0) 17 (24.0) 35 (49.0)	 6 (17.0) 10 (29.0) 19 (54.0)
Former smoker, n (%)	81 (75.0)	33 (60.0)	70 (67.3)	57 (79.0)	30 (86.0)
Pre-enrolment corticosteroid use, n (%)	9 (8.3)	6 (10.9)	6 (5.8)	10 (14.0)	5 (14.0)
Concomitant corticosteroid use, n (%)	8 (7.4)	6 (10.9)	5 (4.8)	NR	NR

PFN: pirfenidone; PBO: placebo; mg/d: milligrams per day

Participant flow and numbers

The loss to follow-up in the three trials was reported in the participant flow figures in the CS (Section 4.5, pages 77-81),[4] which were reproduced from the original publications. The ASCEND[33, 34] and CAPACITY trials all reported two types of patient trial discontinuation. Some patients discontinued the trial due to AEs, being lost to follow-up, withdrawing themselves or being withdrawn by the clinician. These were designated as the "discontinued study" group and did not include patients who had died or underwent lung transplantation. A second group of patients discontinued study treatment, principally on account of AEs, but also due to reasons such as death and lung transplantation. These were designated as the "discontinued treatment" group. However, they were deemed to have completed the study and were included in the analysis. The ASCEND[34] and CAPACITY trials[49] therefore experienced only a small loss of patients to follow-up in terms of those who "discontinued the study": approximately 5%-8% in any arm (see Table 10), compared with between 22% and 37% for any arm in the SP3[38] and SP2[39] trials. However, the rate of attrition was substantially higher (up to 22%) in the ASCEND[34] and CAPACITY trials[49] for the "discontinued treatment" groups (see Table 10). The overall rate of attrition for participants who either "discontinued study" or "discontinued treatment" was between 23% and 29% in any arm of the ASCEND[34] and CAPACITY trials[49] (see Table 10). However, the rates of attrition were essentially similar across intervention and placebo arms.

The primary approach for managing missing values in the efficacy analysis in ASCEND and the CAPACITY trials was to use the ITT population (which consisted of all patients who signed the informed consent form and were randomised). Last observation carried forward (LOCF) was used in SP2 and SP3. The safety analysis population included all patients who signed informed consent and received any amount of study drug (see CS,[4] Table 13). In the analyses of mean change, missing values owing to death were assigned the worst possible outcome (e.g. FVC=0%). Missing values with reasons other than death were imputed as the average value for the three patients with the smallest sum of squared differences at each visit. For the ranked ANCOVA analyses, missing values owing to death were assigned the worst ranks, with early deaths ranked worse than later deaths.

Table 10: Patient loss to follow-up in trials

Trial	Follow-up	Arms	Baseline n	Completed study n (%)	Completed treatment n (%)	Completed study and treatment n (%)
ASCEND[33, 34]	52 weeks	PFN 2,043mg/d	278	261 (94)	223 (80)	206 (74)
		PBO	277	261 (94)	238 (86)	222 (76)
CAPACITY 1[36, 49]	72 weeks	PFN 2,043mg/d	171	158 (92)	137 (80)	124 (72)
		PBO	173	164 (95)	142 (82)	133 (77)
CAPACITY 2[35, 49]	72 weeks	PFN 2,043mg/d	174	161 (93)	136 (78)	123 (71)
		PFN 1,197mg/d	87	82 (95)	70 (80)	65 (75)
		PBO	174	166 (95)	143 (82)	135 (77)
SP3[38]	52 weeks	PFN 1,800mg/d	108	68 (63)	Not reported	Not reported
		PFN 1,200mg/d	55	40 (73)		
		PBO	104	73 (70)		
SP2[39]	9 months	PFN 1,800mg/d	72	56 (78)		
		PBO	35	27 (78)		

PFN: pirfenidone; PBO: placebo; mg/d: milligrams per day

There was only general consistency across trials in terms of the primary and secondary outcomes designated in protocols and reported in publications, so for this reason the efficacy results are structured by clinical area or outcome measure, reflecting the structure of the CS.[4]

4.2.2.1 Lung function

Change from baseline in percent predicted FVC/VC

This outcome was reported by four of the five trials: for FVC by ASCEND[33, 34] and CAPACITY 1 &2[33, 36, 49] and for VC by SP3.[38]

The protocol made publicly available in the clinical trials register reported the primary efficacy outcome in the ASCEND trial[33] as change in percent predicted FVC from baseline to week 52 (see Table 6). The protocol[55] that accompanied the publication stated (Section 13.2, page 29): *"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). Based on new findings from external sources, the analysis of FVC volume will be based on relative change (%) rather than actual volume (mL). A categorical analysis of relative change from Baseline has been added."* [55] The primary efficacy outcome in the protocols and publication for CAPACITY 1 & 2[35, 37, 51] was the change in percent predicted FVC from baseline to week 72.[49] In SP3,[38] the primary efficacy outcome reported was the change from baseline in VC in the pirfenidone 1,800mg per day group compared with the placebo group at 52 weeks.[38]

The ASCEND manuscript[34] did not report the change in percent predicted FVC, but this was reported in the CS,[4] principally to inform the NMA (see CS,[4] Table 20). At week 52, the mean difference in change from baseline in percent predicted FVC for pirfenidone 2,403mg per day compared with placebo was statistically significant in ASCEND[34] (mean difference 4.78%; $p<0.001$, see Table 11).

Table 11: Change from baseline in percent predicted FVC in ASCEND (reproduced from CS,[4] Table 20, page 94)

Study (source)	Treatment	Time point	Mean change from baseline	SE	Mean difference from PBO	*p*-value
ASCEND*[56] (Data on file[1])	PFN 2,403mg/day (n=278)	52 weeks	-6.17	0.875	4.781	<0.001
	PBO (n=277)		-10.95	0.877		
* The ASCEND manuscript did not report the change in percent predicted FVC but this was analysed to inform the NMA.						
[1] Roche 2016a[56]						

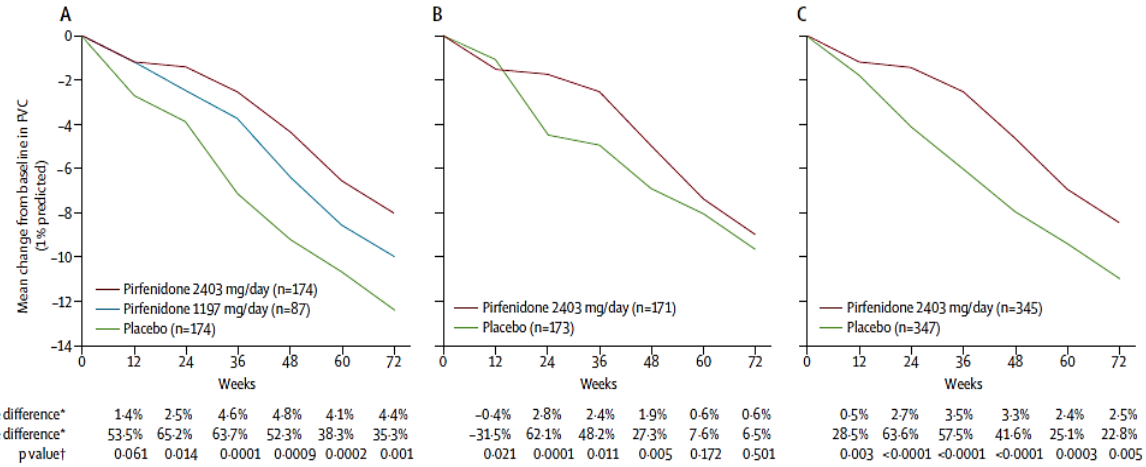
At week 72, the absolute difference in change in percent predicted FVC for pirfenidone 2,403mg per day compared with placebo in CAPACITY 1[49] was not statistically significant (absolute difference:

0.6%; relative difference: 6.5%; 95% CI -3·5 to 4·7, *p*=0.501), see Figure 1 reproduced from Noble 2011[49]).

At week 72, the absolute difference in change in percent predicted FVC for pirfenidone 2,403mg per day compared with placebo in CAPACITY 2[49] was statistically significant (absolute difference 4.4%; relative difference 35.3%; 95% CI 0·7 to 9·1, *p*=0.001). Outcomes in the pirfenidone 1,197mg/day group were intermediate to the pirfenidone 2,403mg/day and placebo groups.

At week 72, the absolute difference in change in percent predicted FVC for pirfenidone 2,403mg per day compared with placebo in a reported pooled analysis of the CAPACITY 1 & 2 trials[49] was statistically significant (absolute difference: 2.5%; relative difference: 22.8%; *p*=0.005, rank ANCOVA, see Figure 1 reproduced from Noble 2011[49]).

Figure 1: Change from baseline in percent predicted FVC in the CAPACITY 2 (A), CAPACITY 1 (B), and in the pooled population (C) (reproduced from Noble *et al.* 2011[49] and CS,[4] page 93)



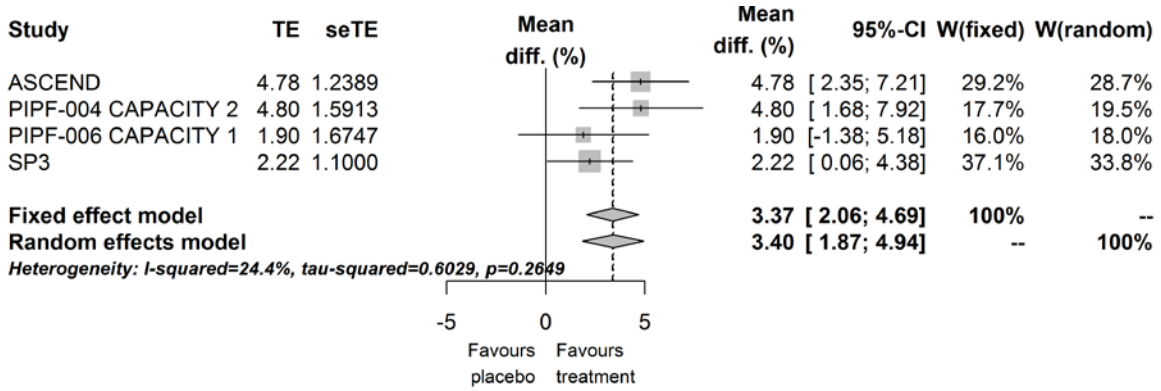
**Pirfenidone 2,403 mg/day versus placebo †Rank ANCOVA (pirfenidone 2,403mg/day vs placebo). 95% CIs were only calculated for absolute differences for the Week 72 time point in CAPACITY 2 (95% CI: 0.7-9.1) and CAPACITY 1 (95% CI: -3.5-4.7)*

At week 52, in SP3,[38] an analysis of the mean decline from baseline in percent predicted VC showed a significant treatment effect of pirfenidone 1,800mg/day compared with placebo, respectively: -2.91% ± 0.77 compared with -5.13% ± 0.78 (*p*=0.044, ANCOVA, see CS,[4] page 93)

The company conducted a meta-analyses using change in percent predicted FVC for ASCEND[33, 34] and CAPACITY 1 & 2[49] and change in percent predicted VC for SP3.[38] Both ASCEND[33, 34] and SP3[38] reported data at week 52, whilst the primary analysis in the CAPACITY trials[49] was at week 72. However, data at week 48 were used for the CAPACITY trials[49] to facilitate a like-for-like comparison between all four studies. The results of the meta-analysis are presented in Figure 2. The results suggest

that the decline in percent predicted FVC in patients receiving pirfenidone (2,403mg/day) was 3.4% less (95% CI: 1.87 to 4.94, *p*-value not reported) than in patients receiving placebo.

Figure 2: Forest plot of the mean difference in change from baseline in percent predicted FVC/VC (%) up to week 52 (reproduced from CS,[4] Appendix 9)



TE, Treatment effect; SE, Standard error

The ERG notes that both CAPACITY 1 and 2[49] report smaller treatment effects at week 72 (MD: 0.6 % in CAPACITY 1 and MD: 4.4% in CAPACITY 2) than at week 48. Selecting the 48 week data for inclusion in the meta-analysis therefore provides a larger estimate of overall treatment effect than would have estimated had the longer-term follow up data been used.

Mean change from baseline in FVC/VC (ml)

This outcome was reported by all five trials: FVC by CAPACITY 1,[49] CAPACITY 2[49] and ASCEND,[34] and VC by SP2[39] and SP3.[38]

Data from 48 weeks from the CAPACITY trials[49] were used in the NMA to allow comparison of studies across a similar time point (see Section 4.6), but the 72-week data are reported here.

All trials showed a statistically significant difference at the 5% level in favour of pirfenidone compared with placebo for change in FVC/VC, except CAPACITY 1[49] (absolute difference -5%; relative difference -1.4%; *p*=0.508). Detailed results are presented in Table 12.

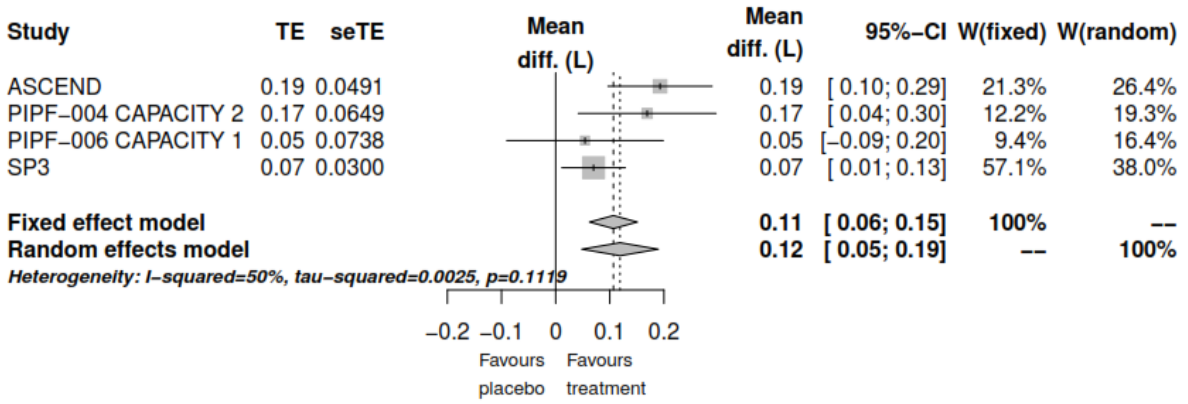
Table 12: Mean change from baseline in FVC/VC (ml) (reproduced from CS,[4] Table 21)

Study	Time point	Treatment group	Mean decline in FVC/VC	Difference, *p*-value†
ASCEND[34]	52 weeks	PFN 2,403mg/day (N=278)	FVC: 235 ml	Absolute difference: 193ml Relative difference: 45.1% *p*<0.001
		PBO (N=277)	FVC: 428 ml	
CAPACITY 1[49]	72 weeks	PFN 2,403mg/day (N=171)	FVC: 379 ml	Absolute difference: -5ml Relative difference: -1.4% *p*-value=0.508
		PBO (N=173)	FVC: 373 ml	
CAPACITY 2[49]	72 weeks	PFN 2,403mg/day (N=174)	FVC: 318 ml	Absolute difference:157ml Relative difference: 33% *p*-value=0.004
		PBO (N=174)	FVC: 475 ml	
SP3[38]	52 weeks	PFN 1,800mg/day (N=108)	VC: 90 ml	PFN 1,800 mg/day vs. PBO: Absolute difference: 70ml Relative difference: NR *p*=0.042
		PFN 1,200mg/day (N=55)	VC: 80 ml	
		PBO (N=104)	VC: 160 ml	
SP2[39]	9 months	PFN 1,800mg/day (N=72)	VC: 30 ml	Absolute difference: 100ml Relative difference: NR *p*=0.037
		PBO (N=35)	VC: 130 ml	

PFN: pirfenidone; PBO: placebo; NR: not reported
†Rank ANCOVA: ASCEND, CAPACITY 1 & 2 (pirfenidone 2,403mg/day vs placebo); SP2 and SP3 (pirfenidone 1,800mg/day vs. placebo)

A meta-analysis for change in FVC/VC (L) was conducted using data form ASCEND[34] and CAPACITY 1 & 2[49] (FVC (L)) and SP3[38] (VC (L)). Both ASCEND[34] and SP3[38] reported data for this outcome at week 52 and data at week 48 were used for the CAPACITY trials.[49] The meta-analysis suggests that on average, over 52 weeks, FVC in patients receiving pirfenidone (2,403mg/day) decline by 0.12L less than patients receiving placebo (95% CI: 0.05 to 0.19, *p*-value not reported), suggesting that pirfenidone slows the decline in lung function (see Figure 3). However, there was moderate heterogeneity between the trials (I^2=50%). In addition, the ERG noted that as with mean difference in change from baseline in percent predicted FVC (%), both CAPACITY 1[49] (MD: 0.005L) and CAPACITY 2[49] (MD: 0.16L) report smaller treatment effects at week 72 than that at week 48.

Figure 3: Forest plot of the mean difference in change from baseline in FVC/VC (L) up to week 52 (reproduced from CS,[4] Appendix 9)



TE, Treatment effect; SE, Standard error

FVC categorical decline of ≥10% percent predicted or death

This outcome was only reported for the ASCEND[34] and CAPACITY 1 & 2 trials.[49] The CS[4] states that a decline in percentage predicted FVC of ≥10% is a decrement that is recognised as clinically significant (see CS,[4] Section 4.7, page 90).

ASCEND[34] 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 at week 52 (absolute difference: 15.3 [95% CI not reported], $p<0.001$). ASCEND also reported a significantly higher proportion of patients with no decline in percent predicted FVC for pirfenidone compared with placebo (22.7% versus 9.7%, $p<0.000001$).[34]

CAPACITY 1[49] reported that there was no statistically significant difference between pirfenidone and placebo in terms of those who had experienced a decline in FVC by ≥10% at week 72 (absolute difference: 3.8 [95% CI: -2.7 to 10.2], $p=0.440$). CAPACITY 1[49] also reported no statistically significant difference in the proportion of patients with no decline in percent predicted FVC for pirfenidone compared with placebo (25.8% versus 22%, *p*-value not reported).

CAPACITY 2[49] 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% at week 72 (absolute difference: 14.4 [95% CI: 7.4 to 21.3], $p=0.001$). CAPACITY 2[49] also reported a higher proportion of patients with no decline in percent predicted FVC for pirfenidone compared with placebo (24.1% versus 13.8%, *p*-value not reported).

Table 13: Categorical analysis of change from baseline in percent predicted FVC or death (reproduced from CS,[4] Table 18)

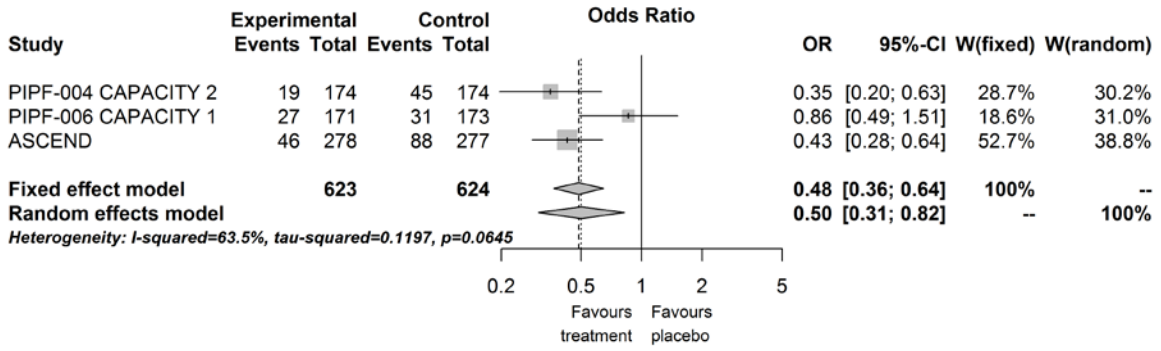
Study	Time point	Treatment group	Decline ≥10% FVC or death, n (%)	No decline* in FVC, n (%)	*p*-value†
ASCEND[34]	52 weeks	PFN 2,403mg/day (N=278)	46 (16.5)	63 (22.7)	*p*<0.000001
		PBO (N=277)	88 (31.8)	37 (13.7)	
CAPACITY 1[49] §	72 weeks	PFN 2,403mg/day (N=171)	39 (22.8)	44 (25.8)	*p*=0.440
		PBO (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
**Change in predicted FVC ≥10%; CAPACITY trials data not reported in original publication (Noble 2011[49])*
†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]

A meta-analysis was conducted using data from 52 weeks for the ASCEND trial[34] and 48 weeks from the CAPACITY trial.[49] The results suggested that compared with placebo, pirfenidone lowers the proportion of patients experiencing decline in FVC percent predicted of ≥10% (OR: 0.50, 95% CI: 0.31 to 0.82, *p*-value not reported, see Figure 4). However, heterogeneity between the trials (I^2=63.5%) was moderately high.

Figure 4: Forest plot of odds ratios for FVC categorical decline of ≥10% percent predicted up to week 52 (reproduced from CS,[4] Appendix 9)



The pooled analysis of the two CAPACITY trials[49] at week 72, showed a lower proportion of patients experienced a decline of ≥10% in percent predicted FVC in the pirfenidone 2,403mg/day group (21% compared with 31%, respectively *p*=0.003).[49]

4.2.2.2 Mortality

All-cause and IPF-related mortality

All five trials provided data on mortality, although none of the studies was powered to assess the effect of pirfenidone on this outcome. No definition of IPF-related mortality was provided in the CS[4] or in the relevant publications.

ASCEND[34] reported all-cause mortality and so-called treatment-emergent IPF-related mortality (i.e. defined as the time after randomisation until 28 days after the final dose of the study drug) at 52 weeks; and CAPACITY 1 & 2[49] reported all-cause mortality, treatment-emergent all-cause mortality, IPF-related mortality and so-called treatment-emergent IPF-related mortality for 52 and 72 weeks.

Details of the all-cause mortality and TE IPF-related mortality at the common time point of 52 weeks, as well as the evidence from CAPACITY 1 & 2[49] at 72 weeks, are presented in Table 14.

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] If one assumes that the reported "all-cause mortality" is actually "treatment emergent all-cause mortality" (these distinctions exist in the CAPACITY trial publication,[49]), then there is a substantial increase in death rates in the pirfenidone group, from 11 at 52 weeks to 19 at 72 weeks, compared with a much smaller increase in the placebo group from 22 at 52 weeks to 29 at 72 weeks (the p-values for the differences between groups are 0.047 and 0.315 for 52 weeks and 72 weeks, respectively). The numbers for non-treatment emergent all-cause mortality are higher (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=0.047) and TE IPF-related deaths (p=0.012) in the pirfenidone groups compared with the placebo groups.

SP3[38] and SP2[39] reported all-cause mortality; there was no significant difference between groups. SP3[38] reported three deaths, four deaths and four deaths in the high-dose (1,800mg/d), low-dose (1,200mg/d) and placebo groups respectively, at 52 weeks,[38] and SP2[39] reported one death in the placebo group only, at 9 months.[39]

Table 14: Mortality rates in the CAPACITY 1 & 2 studies at week 52 and week 72 and the ASCEND and pooled populations at week 52 (reproduced from CS,[4] Table 23, page 97 and Noble 2011[49])

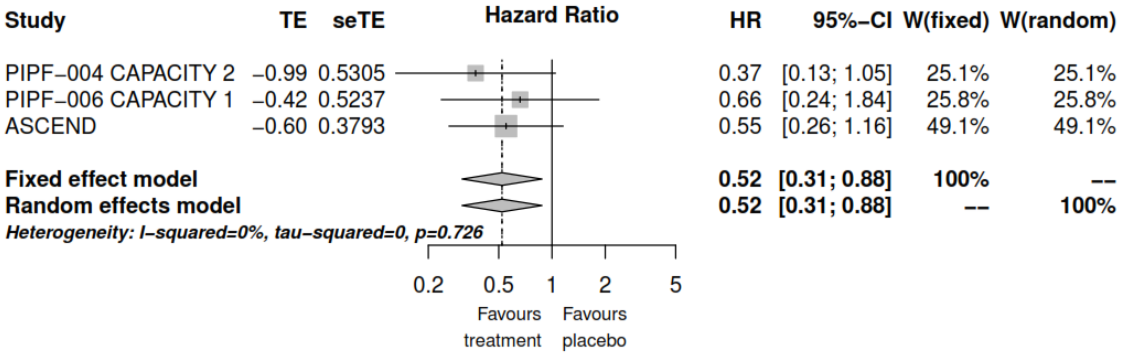
Patients	Time-point	PFN n (%)	PBO n (%)	HR (95% CI)*	*p*-value**
ASCEND[33, 34]	52 weeks	n=278	n=277		
All-cause mortality		11 (4.0)	20 (7.2)	0.55 (0.26, 1.15)	0.105
TE IPF-related mortality		3 (1.1)	7 (2.5)	0.44 (0.11, 1.72)	0.226
CAPACITY 1 & 2[49†]	52 weeks	n=345	n=347		
All-cause mortality		11 (3.2)	22 (6.3)	0.49 (0.24-1.01)	0.047
TE IPF-related mortality		4 (1.2)	15 (4.3)	0.27 (0.09–0.81)	0.012
All-cause mortality	72 weeks	27 (8)	34 (10)	0.77 (0.47-1.28)	0.315
IPF-related mortality		18 (5)	28 (8)	0.62 (0.35-1.13)	0.117
TE all-cause mortality		19 (6)	29 (8)	0.65 (0.36-1.16)	0.141
TE IPF-related mortality		12 (3)	25 (7)	0.48 (0.24-0.95)	0.030
Pooled data for ASCEND, [34] CAPACITY 1 & 2[49]	52 weeks	n=623	n=624		
All-cause mortality		22 (3.5)	42 (6.7)	0.52 (0.31–0.87)	0.011
TE IPF-related mortality		7 (1.1)	22 (3.5)	0.32 (0.14–0.76)	0.006

†Data in the CAPACITY 1 & 2 studies were censored at one year, but the 72-week data were published in Noble 2011[49]
**Cox proportional hazards model*
***Log-rank test (pirfenidone 2,403mg per day vs placebo)*
Abbreviations: PFN: pirfendione; PBO: placebo; TE- treatment-emergent

Meta-analysis was conducted using CAPACITY 1 & 2[49] and ASCEND[34] to assess the effect of pirfenidone on all-cause mortality; the trials reported HRs and the proportion of deaths. The company

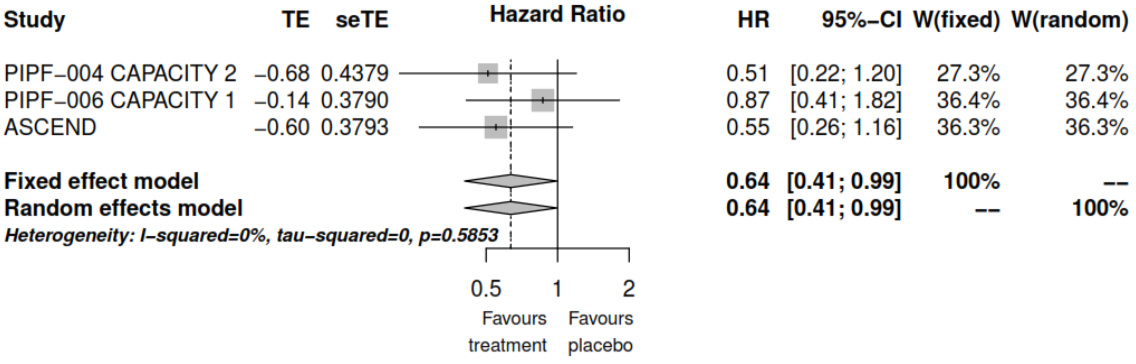
excluded SP3[38] from the analyses as it only reported the proportion of deaths. However, the ERG noted that SP3[38] was included in the company's NMA, where they used the method of Woods *et al.*[58] to combine the proportions reported in SP3[38] with HR. The results of the meta-analysis suggest that pirfenidone (2,403mg/day) compared with placebo reduces all-cause mortality (HR: 0.52, 95% CI: 0.31 to 0.88, *p*-value not reported) at 52 weeks (see Figure 5). A sensitivity analyses of the 3 trials based on data at 72 weeks for the CAPACITY trials[49] also favours pirfenidone (HR: 0.64, 95% CI: 0.41 to 0.99, *p*-value not reported, see Figure 6), however the reduction in mortality is lower than that observed using the 52 week data. Under the assumption of proportional hazards, we would expect the treatment effect to be constant over time.

Figure 5: Forest plot of hazard ratios for all-cause mortality (CAPACITY data at week 52) (reproduced from CS,[4] Appendix 9)



TE, Treatment effect (log hazard ratio); SE, Standard error

Figure 6: Forest plot of hazard ratios for all-cause mortality (CAPACITY data at week 72) (reproduced from CS,[4] Appendix 9)



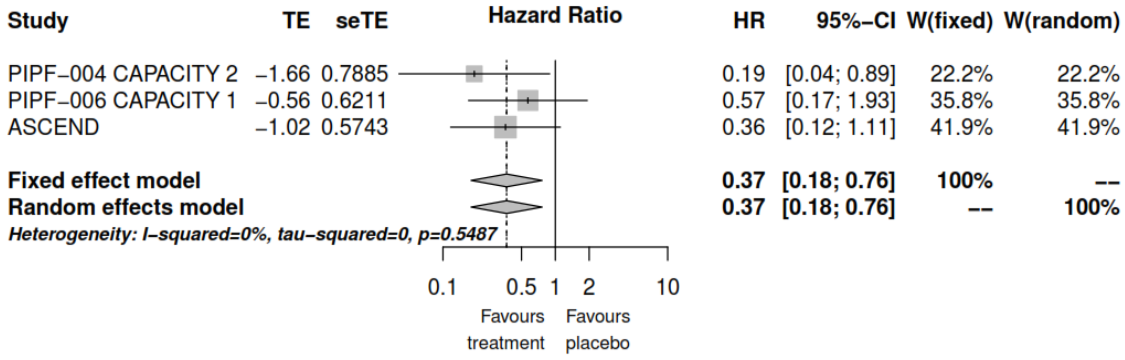
TE, Treatment effect (log hazard ratio); SE, Standard error

Meta-analysis of IPF-related mortality was also conducted using data from CAPACITY 1 & 2[49] and ASCEND.[34] All three trials reported data for 'IPF-related mortality' and 'IPF-related treatment emergent deaths', where treatment-emergent was defined as "the period from baseline to 28 days after

the last dose of the study drug." 'IPF-related mortality' is used in this analysis in line with an ITT approach for analysis. Meta-analysis of the 3 trials[34, 49] at 52 weeks suggests that pirfenidone compared with placebo reduces IPF-related mortality (HR: 0.37, 95%CI: 0.18 to 0.76, *p*-value not reported, see Figure 7).

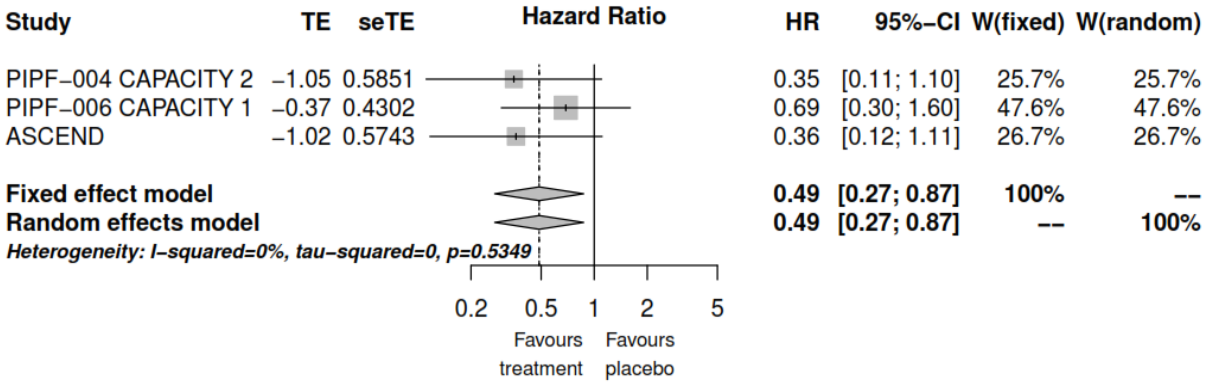
A sensitivity analyses of the three trials[34, 49] based on data at 72 weeks for the CAPACITY trials[49] also favours pirfenidone (HR: 0.49, 95% CI: 0.27 to 0.87, *p*-value not reported, see Figure 8), however, as with the all-cause mortality outcome, the reduction in mortality is lower than that observed using the 52 week data. Under the assumption of proportional hazards, we would expect the treatment effect to be constant over time.

Figure 7: Forest plot of hazard ratios for IPF-related mortality (CAPACITY data at week 52) (reproduced from CS,[4] Appendix 9)



TE, Treatment effect (log hazard ratio); SE, Standard error

Figure 8: Forest plot of hazard ratios for IPF-related mortality (CAPACITY data at week 72) (reproduced from CS,[4] Appendix 9)



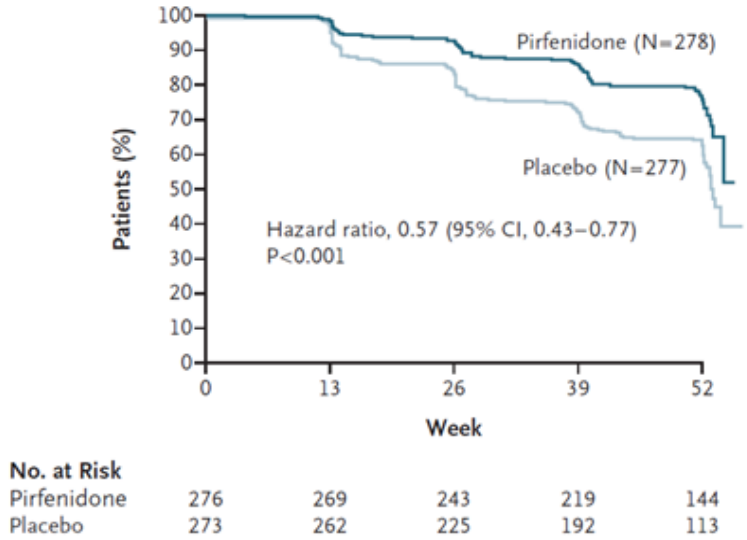
TE, Treatment effect (log hazard ratio); SE, Standard error

4.2.2.3 Progression-free survival (PFS)

The CS makes a case for the inclusion of this outcome based on similarities between IPF and "*the fundamental hallmarks of cancer biology*" (CS,[4] page 99). Four trials reported data for PFS: ASCEND,[34] CAPACITY 1 & 2[49] and SP3.[38] The definitions of PFS varied across the trials. ASCEND[34] defined PFS as the time to the first occurrence of any of the following: a confirmed ≥10% decline from baseline in percent predicted FVC, confirmed ≥50 m decline from baseline in 6MWD, or death.[34] The CAPACITY 1 & 2[49] defined PFS as the time to the first occurrence of any of the following: a confirmed ≥10% decline in percent predicted FVC, ≥15% decline in % predicted DLco or death.[49] In a *post hoc* analysis, the ASCEND[34] definition of PFS was applied to the CAPACITY trials[49] at week 52 and week 72 (see Figure 13). The SP3 trial[38] defined PFS as VC decline of ≥10% or death.

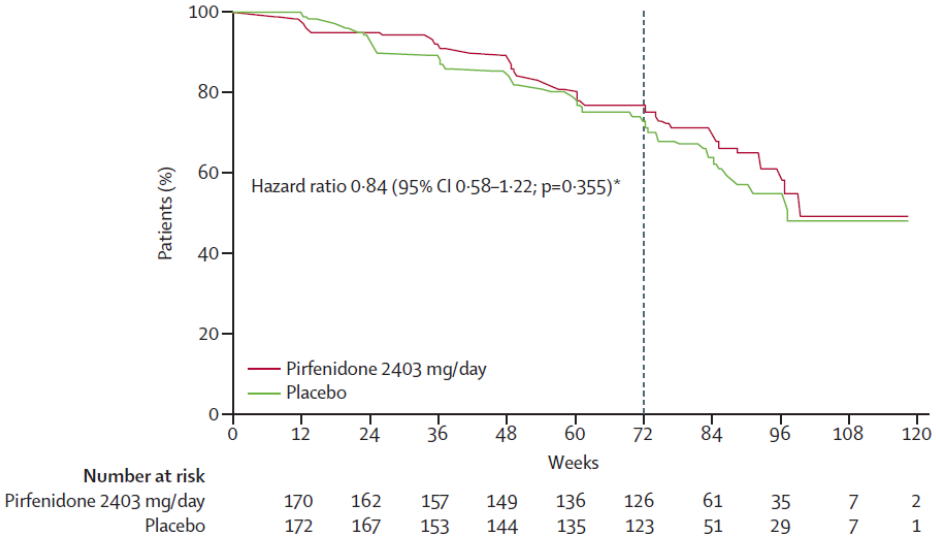
In ASCEND,[34] at 52 weeks, across all randomised patients, there was a statistically significant reduction in the risk of disease progression or death for patients receiving pirfenidone compared with those receiving placebo (HR 0.57; 95% CI, 0.43–0.77, *p*=0.0001, log-rank test, see Figure 9:).[34] That is, for each component of the composite endpoint, fewer patients in the pirfenidone group than in the placebo group had a qualifying event: death (3.6% versus 5.1%); a confirmed absolute decrease of ≥10% in percent predicted FVC (6.5% versus 17.7%); or a confirmed decrease of 50 m or more in the 6MWD (16.5% versus 19.5%).[34]

Figure 9: Kaplan–Meier estimates for PFS in all randomised patients from ASCEND (reproduced from CS,[4] Figure 10 and King 2014[34])



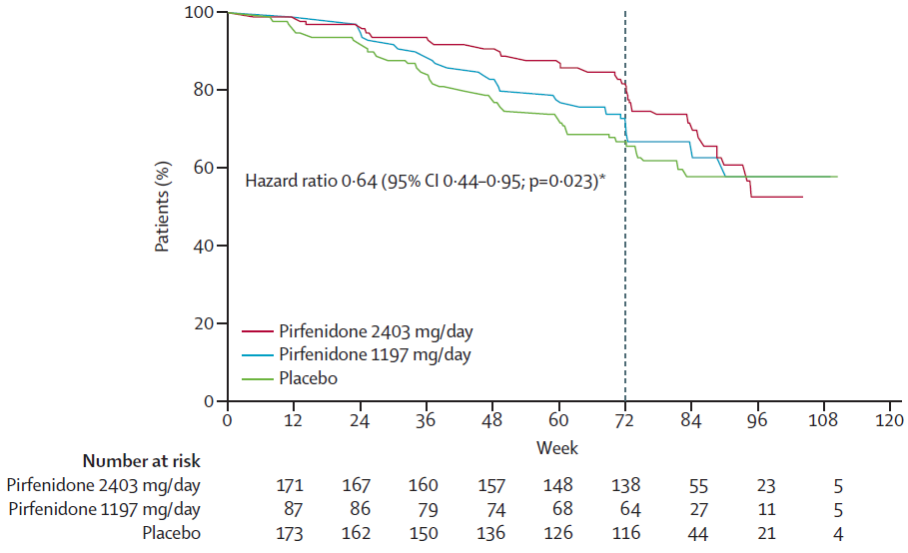
In CAPACITY 1, at 72 weeks, across all randomised patients, there was no statistically significant reduction in the risk of disease progression or death for pirfenidone compared with placebo (HR: 0.84; 95% CI, 0.58 to 1.22, *p*=0.355, see Figure 10).[49]

Figure 10: Kaplan-Meier estimates for PFS in CAPACITY 1 (reproduced from CS,[4] Figure 12)



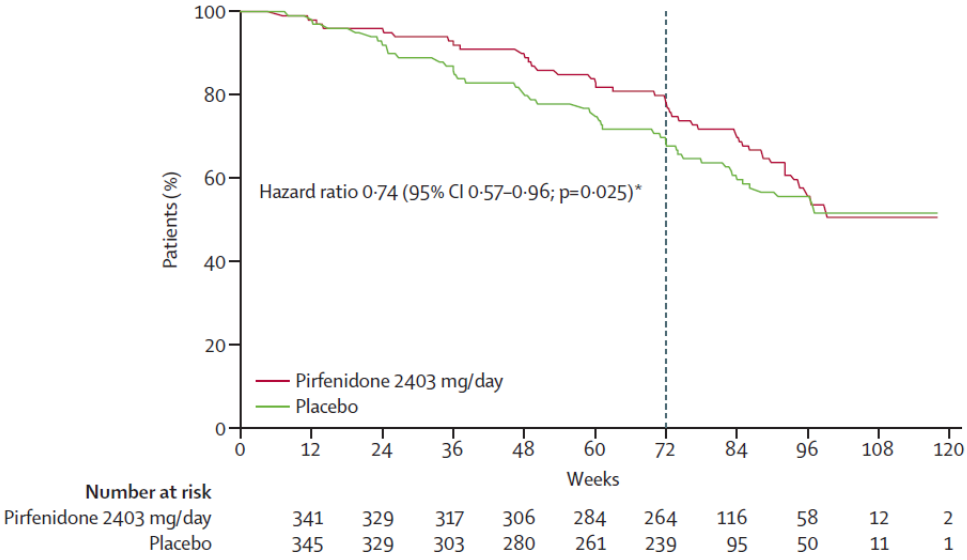
In CAPACITY 2,[49] at 72 weeks, across all randomised patients, there was a statistically significant reduction in the risk of disease progression or death for pirfenidone compared with placebo (HR 0.64; 95% CI, 0.44 to 0.95, *p*=0.023, log-rank test, see Figure 11).

Figure 11: Kaplan-Meier estimates for PFS in CAPACITY 2 (reproduced from CS,[4] Figure 11)



In the pooled population from CAPACITY 1 & 2,[49] at 72 weeks, there was a statistically significant reduction in the risk of disease progression or death for pirfenidone compared with placebo (HR=0.74; 95% CI: 0.57 to 0.96; *p*=0.025, see Figure 12).

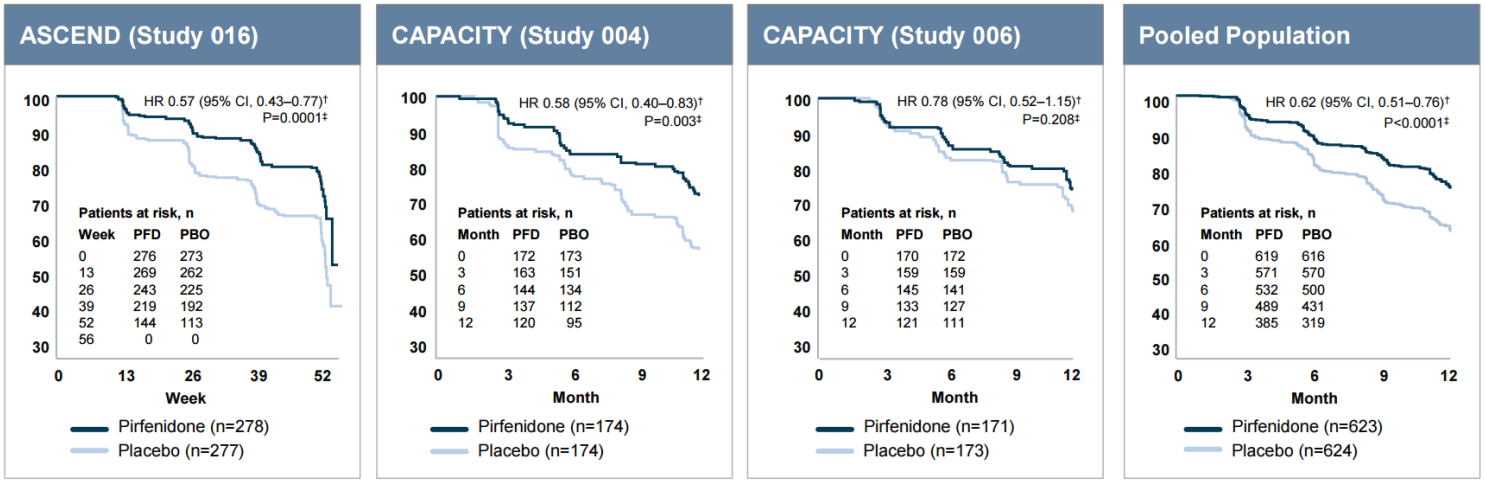
Figure 12: Kaplan-Meier estimates for PFS in the CAPACITY 1 & 2 pooled population (reproduced from CS,[4] Figure 13)



As noted above, an exploratory *post hoc* analysis of PFS was conducted on data from the 52-week CAPACITY 1 & 2[49] populations using the ASCEND[34] definition for disease progression (time to the first occurrence of death, confirmed ≥10% decline in percent predicted FVC, or confirmed ≥50 m decrement in 6MWD). The company justified replacing the DLco criteria with the 6MWD criteria with reference to the relationship between 6MWD and survival.[59] The use of 52-week data and the application of this definition of PFS, which included criteria relating to 6MWD rather than DLco, resulted in reduced HRs and *p*-values in the CAPACITY trials.[49] For CAPACITY 1[49] from HR 0.84 (95% CI, 0.58 to 1.22, *p*=0.355) (original definition using DLco criteria) to HR 0.78 (95% CI, 0.52 to 1.15, *p*=0.208) (using 6MWD criteria), and for CAPACITY 2 from HR 0.64 (95% CI, 0.44 to 0.95, *p*=0.023) to HR 0.58 (95% CI, 0.40 to 0.83, *p*=0.003). See the CS[4] (and Figure 13).

A *post hoc* pooled analysis of these data on PFS from ASCEND,[34] CAPACITY 1 and CAPACITY 2[49] at week 52 was also undertaken: there was a statistically significant reduction in the risk of disease progression or death for pirfenidone compared with placebo (HR: 0.62; 95% CI: 0.51 to 0.76; *p*<0.0001, see Figure 13).

Figure 13: Post hoc analysis of progression-free survival at week 52 in ASCEND, CAPACITY trials, and in the pooled population (reproduced from CS,[4] Figure 14)

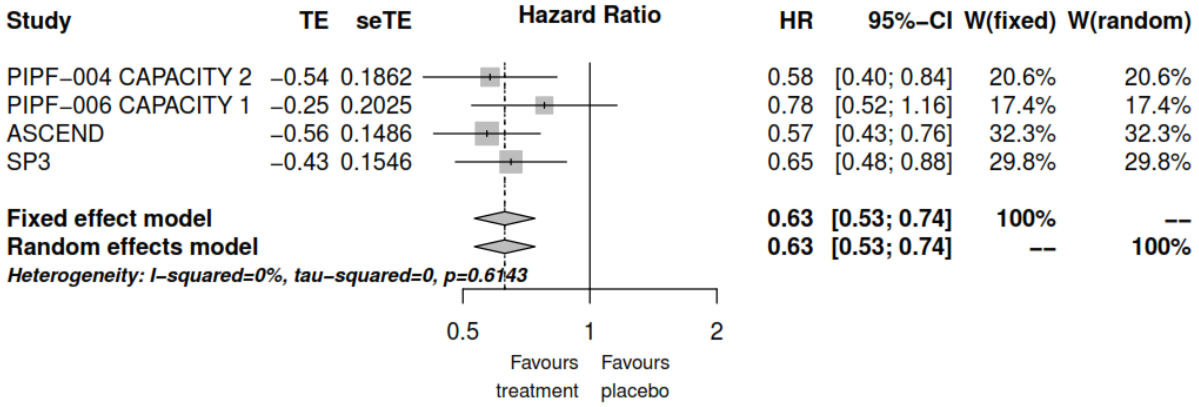


† Cox proportional hazards model
‡ Log-rank test

In SP3,[38] pirfenidone 1,800 mg per day significantly reduced the risk of disease progression or death (defined as VC decline of ≥10% or death) by 55% compared with placebo (HR 0.45; 95% CI 0.11 to 0.79; *p*=0.028, log-rank test).

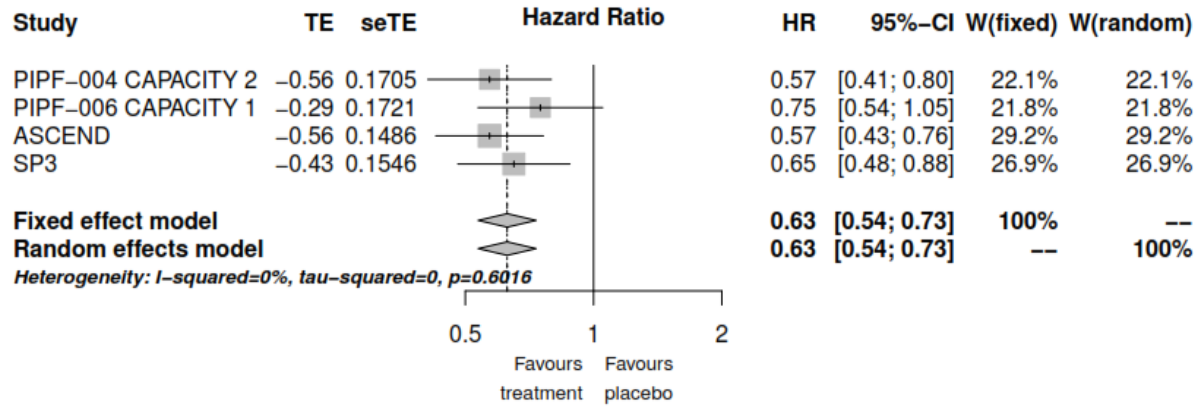
A meta-analysis based on data at 52 weeks was conducted using all four trials (ASCEND,[34] CAPACITY 1&2[49] and SP3[38]). The results of the meta-analysis suggest that pirfenidone compared with placebo reduces the risk of disease progression or death (HR: 0.63, 95% CI: 0.53 to 0.74, *p*-value not reported, see Figure 14). These results are in line with the *post hoc* pooled analysis of ASCEND,[34] CAPACITY 1 and CAPACITY 2[49] at week 52. A sensitivity analysis based on 72 week results for the CAPACITY trials[49] and 52 week results for ASCEND[34] with the assumption that the proportional hazards assumption holds up to 72 weeks gave the same results (see Figure 15). However, as noted above, the definition of PFS varied across the trials and the CS applied the ASCEND[34] definition of PFS to the CAPACITY trials[49] at both week 52 and week 72. The SP3 trial[38] defined PFS as VC decline of ≥10% or death. Hence, the ERG believes caution should be applied when interpreting these results.

Figure 14: Forest plot of hazard ratios for progression-free survival (CAPACITY data at week 52) (reproduced from CS,[4] Appendix 9)



TE, Treatment effect (log hazard ratio); SE, Standard Error

Figure 15: Forest plot of hazard ratios for progression-free survival (CAPACITY data at week 72) (reproduced from CS,[4] Appendix 9)



TE, Treatment effect (log hazard ratio); SE, Standard Error

4.2.2.4 Acute exacerbations

All five trials provided data on acute exacerbations, although the criteria for this outcome varied across the trials. The definitions are provided in secondary outcomes Table 7. For ASCEND,[34] acute exacerbations were identified *"via a post hoc analysis of adverse events based on the MedDRA lower level term 'acute exacerbation of IPF'"* (CS,[4] page 104). The publications for ASCEND[34] and the CAPACITY studies[49] did not report the incidence of acute exacerbations, and the latter trials recorded this outcome only as part of the protocols' composite outcome "Worsening of IPF" (see Table 7). These data were therefore extracted from the CSRs and presented in the CS[4] (Table 27, page 106), for use in the pairwise meta-analysis and NMA.

The rates of acute exacerbation were much higher in the ASCEND trial than in the CAPACITY trials,[49] with a higher incidence in the placebo than the pirfenidone arms in the ASCEND[34] and CAPACITY 2 trials[49]: no *p*-values were reported (see Table 15).

Table 15: CSR data for acute exacerbations for ASCEND, CAPACITY 1 & 2 (reproduced from CS,[4] Table 27)

Trial	Intervention	Time point	n
ASCEND (Data on file)	PFN n=278	52 weeks	24
	PBO n=277		40
CAPACITY 1 (Data on file)	PFN n=171	52 weeks	1
	PBO n=173		0
CAPACITY 2 (Data on file)	PFN n=174	52 weeks	0
	PBO n=174		3

- For ASCEND, acute exacerbations were not reported in the primary manuscript King 2014[34]. Acute exacerbations at 52 weeks were available as data on file.
- For CAPACITY 1 & 2, acute exacerbations were not reported in the primary manuscript Noble 2011[49]. Data at 52 weeks were available as data on file and were handled as separate studies.

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

In SP3,[38] according to the CS[4] and Taniguchi *et al*[38], the incidence of acute exacerbation during the study or within 28 days after the termination of the study was 5.6% (n=6), 5.5% (n=3) and 4.8% (n=5) in the pirfenidone 1,800mg/day, pirfenidone 1,200mg/day and placebo groups, respectively. No statistically significant differences were seen between the three groups. According to a published abstract,[60]stepwise multivariate analysis revealed that decline in VC ≥10% within 6 months was a significant risk factor for acute exacerbations (HR, 3.951, *p*=0.012).

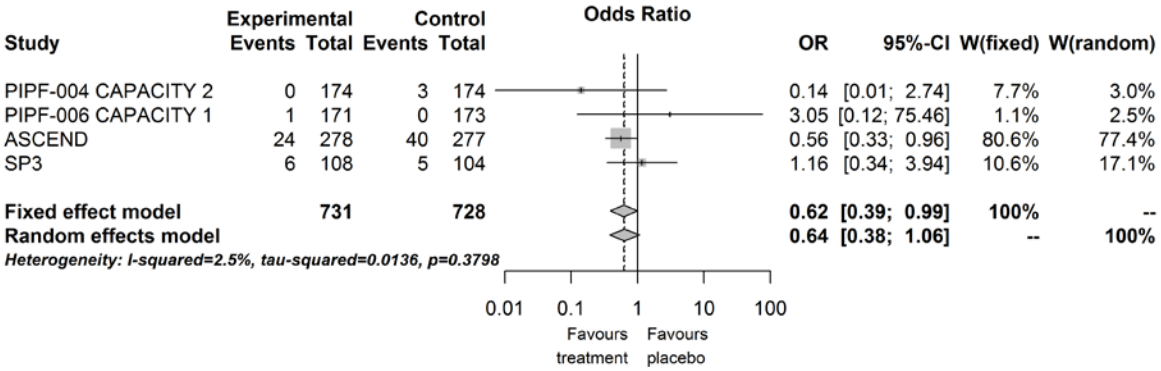
In SP2,[39] according to the CS[4] and Azuma *et al*[39], the incidence of acute exacerbation of IPF was 14% (n=5) in the placebo group and was zero in the pirfenidone group during the 9 months (*p*=0.0031).

There was no consistency in the frequency of acute exacerbation reported across trials. This might be explained by the different definitions used and the difficulty in diagnosis (see clarification response,[10] question A15).

A meta-analysis was conducted based at 52 weeks, using data from ASCEND,[34] CAPACITY 1 & 2[49] and SP3[38] (see Table 15). However, according to the CS,[4] Appendix 9, page 73), data used for the CAPACITY trials[49] were from 48 weeks. The results show that pirfenidone is associated with a reduced risk of acute exacerbations of IPF compared with placebo, with a OR of 0.64 (95% CI: 0.38 to 1.06, *p*-value not reported, see Figure 16). Analyses using a fixed effects model suggest a statistically significant

treatment effect in favour of pirfenidone (OR: 0.62: 95% CI: 0.39 to 0.99, *p*-value not reported). Caution should be applied when interpreting these results as the definition of 'acute exacerbation' varied across trials and there were very few events in the CAPACITY 1 and 2[49] trials whilst in ASCEND[34] the event rate was high.

Figure 16: Forest plot of odds ratios for acute exacerbations up to week 52 (reproduced from CS,[4] Appendix 9)



4.2.2.5 Hospitalisations

This outcome was only reported for CAPACITY 1 & 2[49] and SP2.[39] The protocols for CAPACITY 1 & 2[49] included respiratory hospitalisations as part of the "Worsening of IPF" outcome and SP2[39] reported respiratory hospitalisations, but the CS[4] also reported non-respiratory hospitalisations for the CAPACITY trials[49] (see Table 7).

The CS[4] reported *post hoc* analyses for this outcome (pages 106 and 107), including number of patients hospitalised; number of hospitalisations; mean length of stay in hospital and total number of days in hospital. The data for respiratory and non-respiratory hospitalisations are reported in Table 16. In the pooled CAPACITY 1 & 2[49] population, the number of patients with at least one hospitalisation for respiratory causes (14.8% for pirfenidone versus 15% for placebo) or non-respiratory causes (20.9% versus 16.1% respectively) was similar across treatment arms. However, the duration of hospital stay was consistently numerically greater in the placebo arms.

Table 16: 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
Non-respiratory hospitalisations (NRH)						
Average number of NRH days per patient	1.5	3.7	1.5	2.8	1.5	3.2
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 17: Categorical outcomes for UCSD SOBQ in ASCEND at week 52† (reproduced from CS,[4] Table 34)

Outcomes, n (%)	PFN (n=278)	PBO (n=277)	*p*-value*
Worsening score ≥20 points or death	81 (29.1)	100 (36.1)	0.1577
Worsening score <20 to 0 points	124 (44.6)	115 (41.5)	
No worsening (score change <0 points)	73 (26.3)	62 (22.4)	

†Missing data due to reasons other than death were imputed using the sum of squared differences (SSD) method and included in the ≥20 points category
*p-value by rank ANCOVA

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

In CAPACITY 1 & 2,[49] there were no significant differences between the pirfenidone and placebo groups for the change from baseline to week 72 (see Table 18). There was therefore no evidence of a treatment effect in any of the three key trials.

Table 18: Mean change in UCSD SOBQ score from baseline for the relevant RCTs (ITT population, reproduced from CS,[4] Table 35)

Study	Time point	Treatment group	Mean change in dyspnoea score	*p*-value*
CAPACITY 1[49]	72 weeks	PFN (n=171)	11.9	p=0.604
		PBO (n=173)	13.9	
CAPACITY 2[49]	72 weeks	PFN (n=174)	12.1	p=0.509
		PBO (n=174)	15.2	
*Rank ANCOVA (PFN vs placebo)				

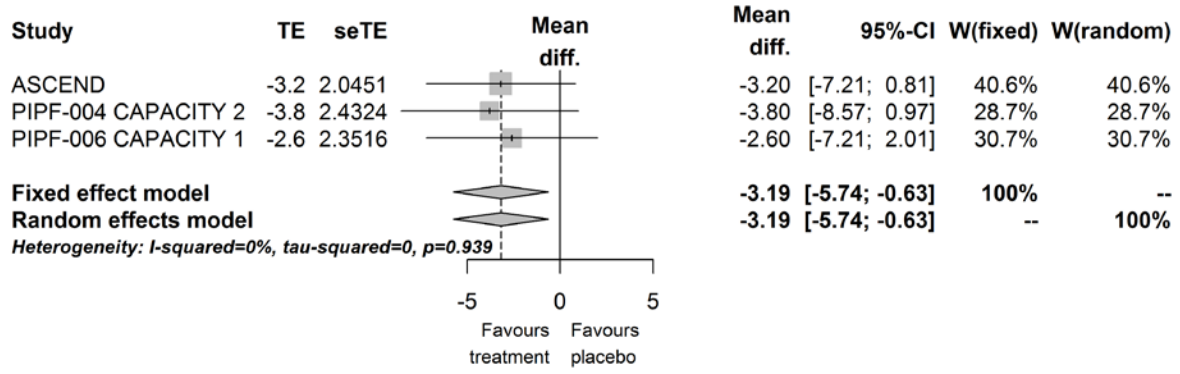
PFN: pirfenidone 2,403mg/d; PBO: placebo

The CS,[4] (page 112) reported that pooled data from the three studies showed pirfenidone treatment reduced the proportion of patients who experienced a ≥20 point increase or death (*p*=0.0471).[37]

The meta-analysis included data from the ASCEND trial[34] at 52 weeks and the CAPACITY trials[49] at 48 weeks. The results suggest that, at this time point, pirfenidone reduces the decline in USCD SOBQ compared with placebo (Mean difference: -3.19 (95% CI: -5.74, to -0.63, *p*-value not reported, see Figure 17), although the mean difference in the individual studies was not statistically significant. The ERG notes that both CAPACITY 1 & 2[49] report smaller treatment effects at week 72 (MD: 2.0% % in CAPACITY 1 and MD: 3.1% in CAPACITY 2) than at week 48 and so the observed statistically

significant difference does not necessarily hold for time points beyond 48/52 weeks, The results of the meta-analysis are consistent with the pooled analysis of CAPACITY 1 & 2[49] and ASCEND.[34]

Figure 17: Forest plot of the mean difference in change from baseline in UCSD SOBQ up to week 52 (reproduced from CS,[4] Appendix 9)



TE, Treatment effect; SE, Standard error

St. George's Respiratory Questionnaire

Only CAPACITY 1 & 2[49] reported data for this outcome. However, it was not listed in any protocols and was not reported in the original publication. It therefore appears to be a *post hoc* analysis. The CS[4] (page 111) states that this measure has demonstrated a strong correlation between physical impairment and disease severity, clinical symptoms, and functional disability in patients with IPF. At week 72, the difference in change in SGRQ between pirfenidone and placebo was not statistically significant in either trial (see Table 19).

Table 19: SGRQ measure of change in health status from baseline to week 72 in CAPACITY 1 & 2 (reproduced from CS,[4] Table 32, page 111)

	Change from baseline to week 72 (mean ± SD)		*p*-value*
	PFN	PBO	
CAPACITY 1[49]	**(n=166)**	**(n=169)**	
SGRQ	7.2 ± 16.85	7.3 ± 20.37	0.766
CAPACITY 2[49]	**(n=163)**	**(n=165)**	
SGRQ	7.6 ± 18.89	9.0 ± 18.86	0.495
*Rank ANCOVA stratified by geographic region (USA and rest of world). Missing data due to a patient's death were ranked as worse than any non-death and according to time until death			

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

As only the CAPACITY trials[49] reported data for this outcome, the company did not conduct a meta-analysis.

4.2.2.7 6-Minute Walking Distance (6MWD) or 6-Minute Walking Test (6MWT)

Three trials reported data on this outcome: ASCEND[34] and CAPACITY 1 & 2.[49] Data were analysed according to categories of decrements of ≥50 metres or <50 metres, and mean change from baseline.

Categorical analysis of change from baseline in 6MWD

The CS[4] states that a decrement of ≥50 metres in 6MWD is considered an appropriate and clinically relevant threshold for a categorical assessment of response to therapy because it has been associated with an increased risk of mortality.[25] Categorical analysis of 6MWD data was carried out *post hoc* in the CAPACITY 1 & 2 studies,[49] but was pre-specified as a secondary endpoint in ASCEND in the protocol accompanying the publication,[55] but not in the clinical trials register protocol.[33] However, the CS[4] (Table 29, page 108) reported findings for these trials for a *post hoc* composite outcome of mean decline ≥50 m from baseline in 6MWD or death.

At week 52, the absolute difference in the proportion of patients with a mean decline ≥50 m from baseline, or death, for pirfenidone 2,403mg per day compared with placebo in ASCEND[34] was statistically significant (absolute difference: 9.8%; relative reduction: 27.5%; *p*=0.04, see Table 20).

Table 20: Proportion of patients with a mean decline of ≥50 m in 6MWD from baseline or death in ASCEND and CAPACITY 1 &2 (ITT population) (reproduced from CS,[4] Table 29)

Trial	Time point	Treatment group	Mean decline of ≥50 m in 6MWD or death, n (%)	Difference, *p*-value
ASCEND[34]	52 weeks	PFN (n=278)	72 (25.9)	Absolute difference: 9.8% Relative reduction: 27.5% *p*=0.04*
		PBO (n=277)	99 (35.7)	
CAPACITY 1[49]	72 weeks	PFN (n=169)	56 (33.1)	*p*=0.10**
		PBO (n=168)	79 (47.0)	
CAPACITY 2[49]	72 weeks	PFN (n=170)	62 (36.5)	*p*=0.049**
		PBO (n=170)	80 (47.1)	
Pooled CAPACITY 1 & 2[49]	72 weeks	PFN (n=339)	118 (34.8)	Absolute difference: 12.2% Relative risk: 26% *p*=0.001**
		PBO (n=338)	159 (47.0)	

PFN: pirfenidone; PBO: placebo
**Rank ANCOVA (pirfenidone 2,403mg/day vs placebo)*
***Cochran-Mantel-Haenszel test*

At week 72, the difference in the proportion of patients with a mean decline ≥50 m from baseline, or death, for pirfenidone 2,403mg per day compared with placebo in CAPACITY 1[49] was not statistically

significant (p=0.10, see Table 20). At week 72, the difference in the proportion of patients with a mean decline ≥50 m from baseline, or death, for pirfenidone 2,403mg per day compared with placebo in CAPACITY 2[49] was statistically significance (p=0.049, see Table 20).

At week 72, the absolute difference in the proportion of patients with a mean decline ≥50 m from baseline, or death, for pirfenidone 2,403mg per day compared with placebo across CAPACITY 1 & 2[49] was statistically significant (absolute difference: 12.2%; relative reduction: 26%; p=0.001, see Table 20).

The CS[4] (page 108) reported that a *post hoc* pooled analysis of ASCEND[34] and CAPACITY 1 & 2 (data from weeks 52 and 48 respectively)[49] reported a statistically significant improvement in 6MWD for patients receiving pirfenidone 2,403mg per day compared with placebo (p=0.0004). The CS[4] cites Nathan 2014[19] as the supporting study, but the reference provided does not contain this analysis; the source of this analysis and its data is therefore unclear.

Mean change in 6MWD from baseline

The CS[4] states that the reliability and validity of 6MWD as a responsive measure of disease status and a valid endpoint for clinical trials has been demonstrated in a recent study, where the minimally clinical important difference (MCID) was estimated at 24-45 meters.[25]

At week 52, in ASCEND,[34] the absolute difference in the mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo was statistically significant (absolute difference: 26.7m; relative reduction: 44.2%; p=0.036) and satisfied the lower end of the MCID (see Table 21).

At week 72, in CAPACITY 1,[49] the absolute difference in the mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo was statistically significant (absolute difference: 31.8m; relative difference: not reported; p<0.001) and satisfied the MCID (see Table 21). Therefore, the CAPACITY 1[49] results for the categorical analysis of 6MWD (not statistically significant) and the mean change in 6MWD (statistically significant) were different in terms of statistical significance.

Table 21: Mean change from baseline in 6MWD in ASCEND and CAPACITY 1 & 2 (reproduced from CS,[4] Table 30)

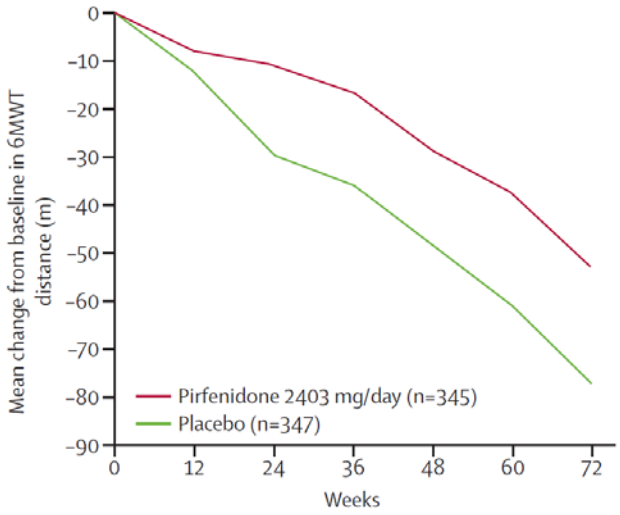
Study	Time point	Treatment group	Mean decline, metres	Difference, *p*-value†
ASCEND[34]	52 weeks	PFN (n=278)	33.5 m	Absolute difference: 26.7 m Relative reduction: 44.2% *p*=0.036
		PBO (n=277)	60.2 m	
CAPACITY 1[49]	72 weeks	PFN (n=174)	45.1 m	Absolute difference: 31.8 Relative difference: NR *p*<0.001
		PBO (n=174)	76.9 m	
CAPACITY 2[49]	72 weeks	PFN (n=171)	60.4 m	Absolute difference: 16.4 m Relative difference: NR *p*=0.171
		PBO (n=173)	76.8 m	
Pooled CAPACITY 1 & 2	72 weeks	PFN (n=345)	52.8 m	Absolute difference: 24 m Relative difference: 31.2% *p*=0.0009
		PBO (n=347)	76.8 m	

PFN: pirfenidone; PBO: placebo; m: metres
†Rank ANCOVA (pirfenidone 2,403 mg/day vs placebo)

However, at week 72, in CAPACITY 2,[49] the absolute difference in the mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo was not statistically significant (absolute difference: 16.4m; relative difference: not reported; *p*=0.171) and did not satisfy the lower end of the MCID (see Table 21). Therefore, the CAPACITY 2[49] results for the categorical analysis of 6MWD (statistically significant) and the mean change in 6MWD (not statistically significant) were different in terms of statistical significance.

At week 72, in the pooled analysis of CAPACITY 1 & 2,[49] the absolute difference in the mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo was statistically significant (absolute difference: 24m; relative difference: 31.2%; *p*=0.0009) and satisfied only the lowest threshold of the MCID (see Table 21 and Figure 18).

Figure 18: Mean change from baseline in 6MWD in CAPACITY 1 & 2 pooled population (reproduced from CS,[4] Figure 15)

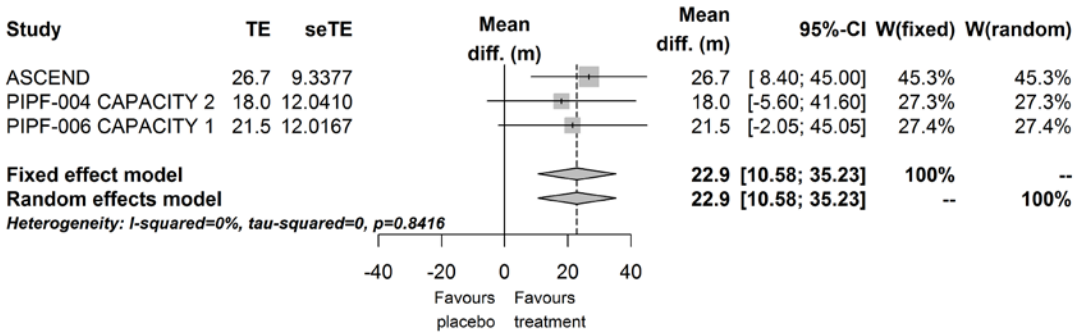


Absolute difference* (m)	3·9	18·6	18·7	19·8	23·3	24·0
Relative difference*	32·2%	62·8%	52·5%	40·6%	38·2%	31·2%
p value†	0·760	0·042	0·053	0·004	0·002	0·0009

**Pirfenidone 2,403 mg/day vs placebo*
†Rank ANCOVA (pirfenidone 2,403 mg/day vs placebo)

Three trials (ASCEND,[34] CAPACITY 1 & 2[49]) were included in the meta-analysis to assess change in distance walked from baseline in the 6MWT. Data at week 48 from the CAPACITY trials[49] were combined with data from week 52 in the ASCEND trial.[34] The meta-analysis suggests that, on average, patients receiving pirfenidone declined by 22.9m less than patients receiving placebo with a 95% CI of (10.58m to 35.23m, *p*-value not reported, see Figure 19).

Figure 19: Forest plot of the mean difference in change from baseline in 6MWD up to week 52 (reproduced from CS,[4] Appendix 9)



TE, Treatment effect; SE, Standard error

4.2.2.8 Measurement of the carbon monoxide diffusing capacity of the lungs (DLco)

Four trials (CAPACITY 1 & 2,[49] SP3,[38] SP2[39]) reported data on the change from baseline in DLco. The CAPACITY trials reported the change in % predicted DLco, while SP2[39] and SP3[38] reported the mean decline (mL/min/mmHG). None of the trials showed a statistically significant treatment effect compared to placebo for this outcome measure.

CAPACITY 1[49] reported a mean change of -9.8% for pirfenidone and -9.2% for placebo, respectively (p=0.996); and CAPACITY 2[49] reported a mean change of -7.9% for pirfenidone and -9.9% for placebo (p=0.145). A published, pooled analysis also indicated that there was no evidence of a statistically significant treatment effect for this outcome (p=0.301).[49] In both the SP2[39] and SP3 trials,[38] there was no statistically significant difference in mean decline of DLco between pirfenidone 1,800mg/day and placebo. The company did not conduct a meta-analysis for DLco as the measurements were not considered comparable.

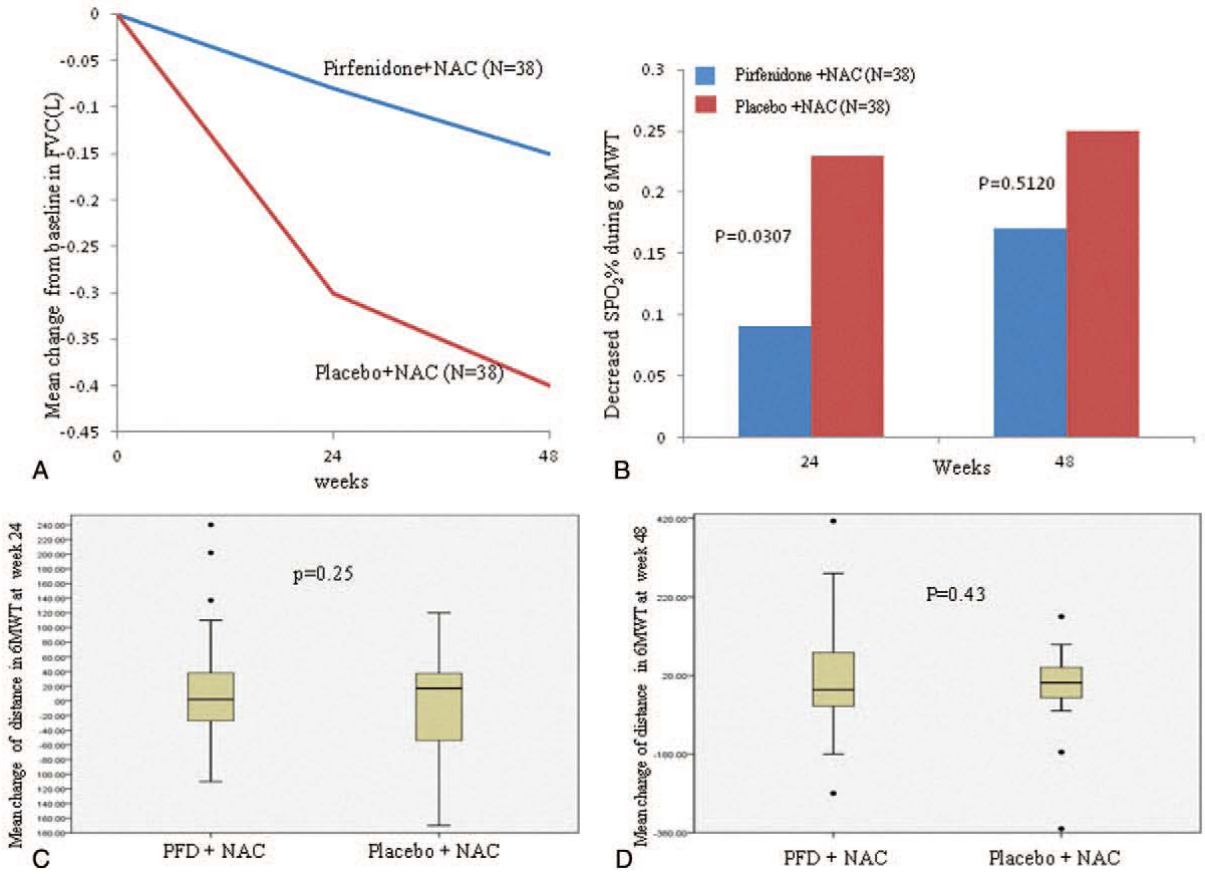
4.2.2.9 Supporting evidence from the Huang et al. trial [48] of pirfenidone plus NAC versus placebo plus NAC

For the purposes of this appraisal, as supporting evidence, only details of the Huang *et al.* efficacy results for FVC, 6-Minute Walking Test (6MWT) and PFS, are presented here[48] (see Figure 20).

The Huang *et al.* trial[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). The authors performed *post hoc* analyses to explore possible reasons behind the change from week 24 to week 48. In doing so, they identified four patients (three in the pirfenidone group and one in the placebo group) who experienced a substantial decline in pulmonary function test parameters (including FVC and DLco) due to AEs after 24 weeks but before 48 weeks. When these patients, were excluded from the analyses, the authors reported that they found a significant treatment effect at both 24 weeks (p=0.018) and 48 weeks (p=0.048).

This trial[48] also reported that there was no statistically significant mean change in 6MWT from baseline for pirfenidone plus NAC compared with placebo plus NAC at either 24 (p=0.25) or 48 weeks (p=0.43, see Figure 20).

Figure 20: Mean change from baseline in FVC at 24 and 48 weeks and in 6MWT at 24 and 48 weeks (reproduced from Huang 2015,[48] Figure 2A-D)



FVC: Forced vital capacity; 6MWT: 6-minute walking test

PFS was also evaluated, defined as the time until the first occurrence of any one of the following: a confirmed ≥10% decline in the percentage predicted FVC, a confirmed ≥15% decline in the percentage predicted DLco (corrected based on the patient's actual haemoglobin levels), a confirmed progression of fibrosis defined by the HRCT fibrosis score, AE-IPF, or death. For PFS, pirfenidone plus NAC had a significant treatment benefit compared with placebo plus NAC (HR=1.88, 95% CI: 1.092–3.242, *p*=0.02). No significant differences were observed in the percent change in the secondary outcomes of arterial blood gas (ABG) ($PaCO_2$, PaO_2, and SaO_2) levels, the dyspnoea score, the HRCT findings, the SGRQ score, or the number of IPF-related adverse events between the pirfenidone and placebo groups.

4.3 Subgroup analyses

4.3.1 Pre-specified analyses

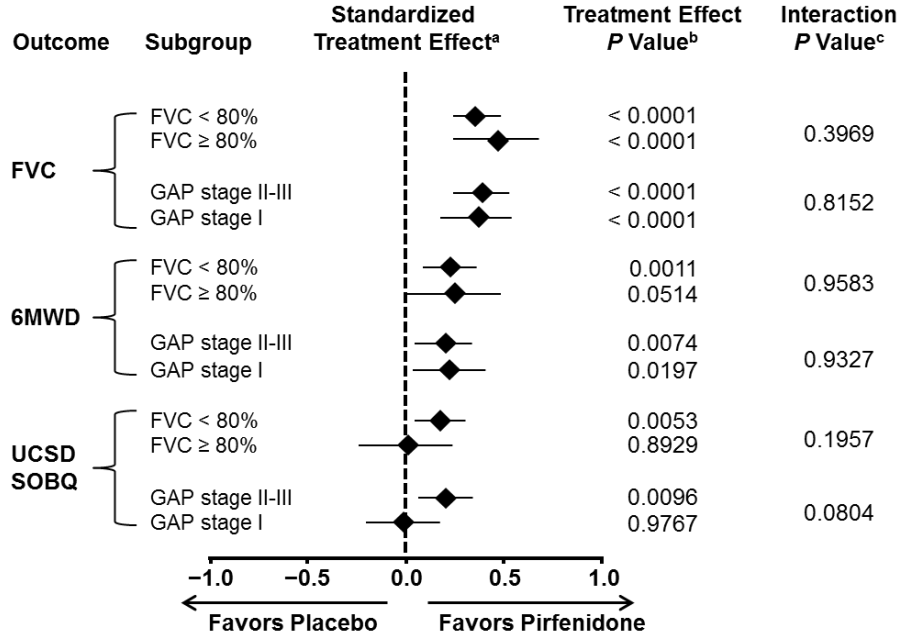
No subgroup analyses were pre-specified for the ASCEND,[34] SP3[38] or SP2[39] trials. Subgroup analyses based on pooled CAPACITY 1 & 2 data were reported in the CS,[4] Figure 16 (page 114) for the primary efficacy outcome variable (difference between pirfenidone and placebo in mean change from baseline to week 72 in percent predicted FVC). There was no evidence for differential treatment effects

according to: sex (p=0.263), age (p=0.864), race (p=0.807), geographic region (p=0.359), and baseline IPF severity (p=0.352). However there was evidence of an interaction between treatment and time from IPF diagnosis (p=0.021), with patients diagnosed >1 year before randomisation experiencing greater effect). In response to a clarification request by the ERG concerning some of the subgroups in this analysis, the company stated: *"As results reported in Figure 16 deviated from [the] more robust approach for the primary outcome, we believe they should not be further used for assessment of robustness and consistency of results in subpopulations"* (see clarification response,[10] question A29).

4.3.2 Post hoc analyses

A *post hoc* analysis of pooled data from ASCEND[34] and CAPACITY 1 & 2[49] was conducted to examine the effects of pirfenidone on patients stratified by earlier versus more advanced disease severity, i.e. "earlier" being "mild" IPF: baseline FVC ≥80% (pirfenidone, n=146; placebo, n=170); and "more advanced" being "moderate" IPF: baseline FVC <80% (pirfenidone, n=477; placebo n=454). According to the CS,[4] (page 115), baseline characteristics and demographics were similar across groups. Efficacy outcomes of interest included absolute ≥10% FVC decline, ≥50m 6MWD decline, and ≥20-point worsening of dyspnoea as measured by UCSD SOBQ. Treatment-by-subgroup interactions were tested based on rank ANCOVA models. Missing values were imputed by using the sum of squared differences method. Factors in the model include study, geographic region, treatment group, subgroups, and treatment-by-subgroup interaction. A proportional hazards model estimated the HR between subgroups. The analysis indicated that there was no significant difference (treatment-by-subgroup interaction) between those patients with baseline FVC ≥ 80% predicted and those with FVC < 80% predicted (see Figure 21).

Figure 21: Treatment effect of pirfenidone by baseline disease severity from pooled data of ASCEND, CAPACITY 1 & 2 (reproduced from CS,[4] Figure 17)



A separate *post hoc* analysis (unpublished) was conducted to evaluate the outcomes for patients who experienced a ≥10% decline in percent predicted FVC during the first 6 months of treatment across the three ASCEND and CAPACITY 1 & 2 trials.[61, 62] Eight-four out of 1,247 patients experienced a ≥10% decline in % FVC during the first 6 months of treatment across these trials: 24 had received pirfenidone (it is unclear if any of these had received the 1,197mg per day dose) and 60 had received placebo. Of these, one (4.2%) had experienced >10% decline in FVC in the pirfenidone group, and 15 (25%) in the placebo group (p=0.032) (see Table 22). The CS[4] states that these findings suggest a potential benefit to continued treatment with pirfenidone despite an initial decline in FVC; this is not consistent with the stopping rule currently recommended in NICE TA282.[2]

Table 22: Outcomes following previous ≥10% decline in FVC at 6 months in ASCEND and CAPACITY 1 & 2 (reproduced from CS,[4] Table 37)

Outcome, n (%)	PFN (n=24)		PBO (n=60)		Relative Difference*	*p*-value
≥10% decline in FVC or death	1	(4.2)	15	(25.0)	−83.3%	0.032
Death	0	(0)	10	(16.7)	−100%	0.056
>0% to <10% decline in FVC	9	(37.5)	23	(38.3)	-2.2%	ND
No further decline in FVC	14	(58.3)	22	(36.7)	59.1	0.089
**Relative difference calculated using the following formula: 100 × [pirfenidone − placebo]/[placebo]*						

These results were supported by an additional *post hoc* analysis,[37] which evaluated the effect of pirfenidone on subgroups based on age, smoker status, and baseline disease status; this analysis found no evidence for differential effects between subgroups.

Exploratory subgroup analyses were conducted in SP3[38] and SP2[39] also. Both analyses found that, in terms of percent predicted VC, IPF patients with baseline percent predicted VC ≥ 70% had better outcomes in terms of VC and PFS at week 52 than those patients with a baseline percent predicted VC <70% although for SP2 the actual data were not reported.[39]

In response to a clarification request from the ERG (see clarification response,[10] question A31), the company also provided results on OS (see Table 23) and PFS (see Table 24) from the ASCEND[34] and CAPACITY trials[49] on groups with a baseline percent predicted FVC of ≤80% (moderate IPF) and >80% (mild IPF). However, numbers within each trial and trial arm were not reported.

The findings did not suggest differential treatment effects according to disease severity for either outcome (as judged by the reported HR and 95% CI), however a treatment-by-subgroup interaction test was not reported so it is unclear if the difference between these subgroups was statistically significant.

Table 23: OS results by baseline FVC percent predicted subgroup at 52 and 72 weeks (reproduced from Clarification response,[10] question A31, Table 12)

Study & time point	Baseline FVC ≤80% predicted			Baseline FVC >80% predicted		
	Adjusted HR	95% CI	*p*-value	Adjusted HR	95% CI	*p*-value
CAPACITY 1[49]						
52 weeks	0.6	0.17-2.04	0.4051	0.77*	0.11-5.59	0.7932
72 weeks	0.89	0.40-1.99	0.7763	0.77	0.11-5.59	0.7932
CAPACITY 2[49]						
52 weeks	0.25	0.08-0.76	0.0080	NE**	**	**
72 weeks	0.29	0.10-0.79	0.0102	4.04***	0.42-38.87***	0.1900***
ASCEND[34]						
52 weeks	0.63	0.29-1.34	0.2215	<0.01	0.00-NE	0.1231
72 weeks	N/A	N/A	N/A	N/A	N/A	N/A
Pooled trials						
52 weeks	0.48	0.27-0.83	0.0071	0.59	0.14-2.51	0.4682
72 weeks	0.58	0.36-0.94	0.0240	0.90	0.27-2.99	0.8610

NE: not evaluable
** Only two deaths occurred in CAPACITY 1 before 52 weeks*
*** There were no additional deaths observed in either arm of CAPACITY 2 between 52 and 72 weeks in patients with FVC >80% predicted*
**** Low number of events*

Table 24: PFS results by baseline FVC percent predicted subgroup at 52 and 72 weeks (reproduced from clarification response,[10] question A31, Table 11)

Study & time point	Baseline FVC ≤80% predicted			Baseline FVC >80% predicted		
	Adjusted HR	95% CI	*p*-value	Adjusted HR	95% CI	*p*-value
CAPACITY 1[49]						
52 weeks	0.84	0.53-1.32	0.4438	0.63	0.29-1.41	0.2571
72 weeks	0.85	0.58-1.26	0.4128	0.56	0.28-1.11	0.0919
CAPACITY 2[49]						
52 weeks	0.60	0.40-0.92	0.0159	0.40	0.18-0.89	0.0193
72 weeks	0.58	0.39-0.86	0.0590	0.48	0.25-0.92	0.0233
ASCEND[34]						
52 weeks	0.56	0.41-0.76	0.0002	0.64	0.30-1.40	0.2584
72 weeks	N/A	N/A	N/A	N/A	N/A	N/A
Pooled trials						
52 weeks	0.62	0.52-0.78	<0.0001	0.54	0.35-0.75	0.0069
72 weeks	0.64	0.52-0.79	<0.0001	0.53	0.35-0.79	0.0017

4.4 Non-randomised and non-controlled evidence

The CS[4] reported findings from RECAP (PIPF-012),[40] a non-randomised, non-controlled, OLE of the ASCEND[34] and CAPACITY trials.[49] The study was designed to assess the long-term safety of pirfenidone 2,403mg per day in patients with IPF. To be included in the extension study, patients must have "completed treatment", that is, they must have received ≥80% of scheduled doses (of either active

treatment or placebo) and completed the week 72 final study visit in CAPACITY 1 or 2[49] (CS,[4] page 159, see Table 25).

Table 25: Summary of RECAP study design

	RECAP (PIFP-012) (Costabel, 2014[40]; Kreuter 2014[63])
Study design	Open-label, uncontrolled, Phase III extension study in which eligible patients receive treatment with pirfenidone 2,403mg/day
Intervention	Eligible patients received pirfenidone 2,403mg/day Concomitant therapy with corticosteroids, azathioprine, cyclophosphamide, and/or NAC were permitted if judged appropriate by investigator
Population	IPF patients that completed the ASCEND[34] or CAPACITY 1 & 2
Objectives	**Primary objective:** To examine the long-term safety and tolerability of pirfenidone in patients with IPF who were previously randomised to the placebo group in either CAPACITY 1 or 2 studies (later adjusted to allow enrolment from the ASCEND trial, Kreuter 2014[63]) **Secondary objective:** To obtain additional efficacy data for pirfenidone 2,403mg/day in patients with IPF
Inclusion/Exclusion criteria	**Inclusion criteria:** • Completes the ASCEND or CAPACITY studies final visit • In the opinion of the principal investigator has been generally compliant (received ≥80% of scheduled doses) with study requirements during the qualifying study, or must be considered eligible to enrol in RECAP by the InterMune medical monitor • Is able to provide informed consent and comply with the requirements of the study **Exclusion criteria:** • Pregnant or lactating women • In the opinion of the PI, is not a suitable candidate for study participation • Known hypersensitivity to any of the components of the study drug • Participates in another interventional clinical trial between the end of participation in ASCEND or either CAPACITY studies and time of enrolment in RECAP • Receives concomitant medications defined in the protocol • Permanently discontinues study drug during the ASCEND or CAPACITY studies for any reason

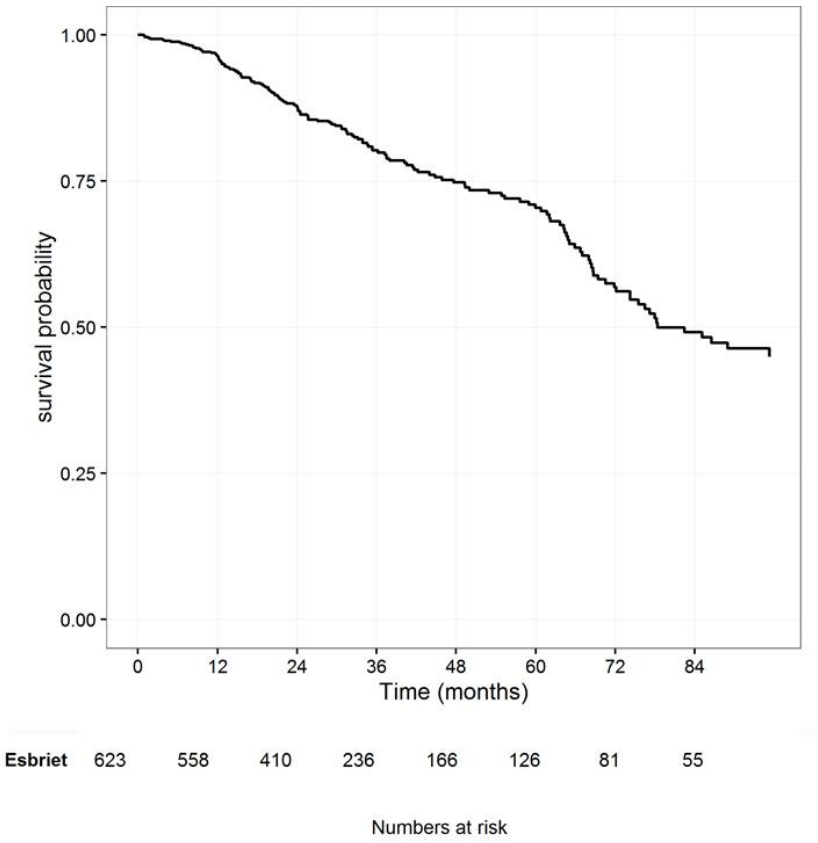
To facilitate comparison with outcomes from the 72-week CAPACITY trials,[49] subgroup analyses were conducted for those who had received placebo in the original trials and who either had baseline FVC and DLco values that met ASCEND[34] or CAPACITY[49] entry criteria (n=178) or did not (n=96)[40,63], although no results were reported in the CS.[4] The publication by Kreuter et al[63] found that discontinuation rates were highest in those patients who had originally received placebo and especially those who did not meet the ASCEND[34] or CAPACITY[49] entry criteria.

In total 603 patients were enrolled in RECAP from the CAPACITY trials.[63] Participants from the ASCEND trial[34] have also been eligible since 2014. The CS[4] (page 158) states that no published data analysis including ASCEND is available to date, but the text refers to CAPACITY/ASCEND data. No results were reported for this specific population in the CS.[4]

The RECAP study[40, 56] is ongoing. The most recent data-cut was performed in June 2015 and the next data-cut is planned in June 2016; some analyses based on summary data from this data-cut were provided by the company as an unpublished conference presentation.[64] As noted in Table 23, the primary objective was to evaluate the safety of pirfenidone 2,403mg per day: data on AEs are included in the integrated analyses set reported under Section 4.5, Table 29.

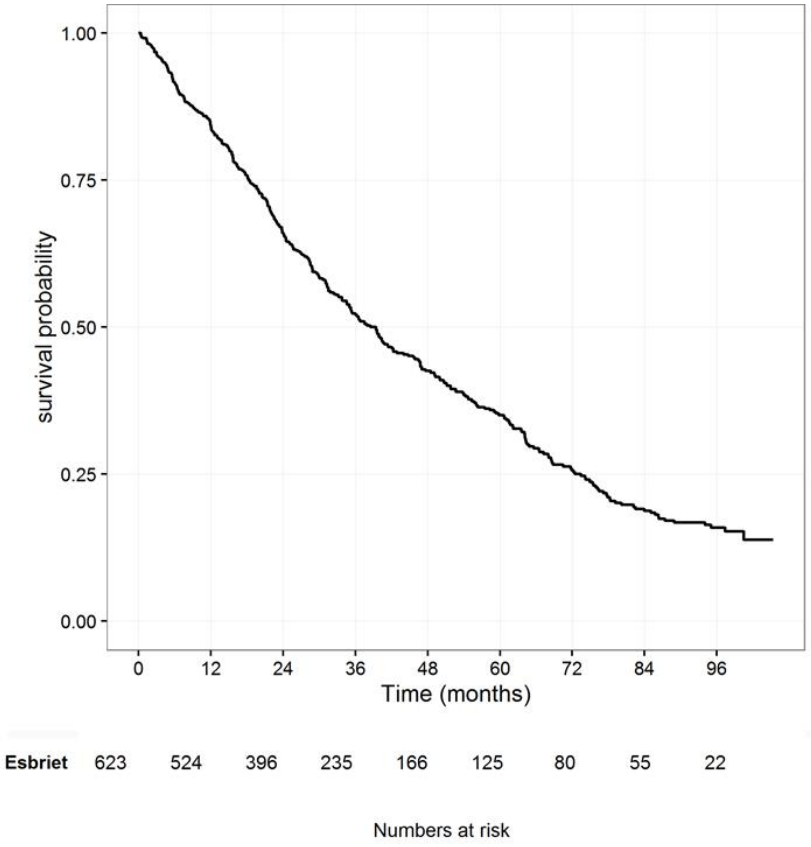
Survival data and time-on-treatment data were reported in the CS[4] (pages 159-161) and are presented here for patients who received pirfenidone 2,403mg per day from baseline onwards in CAPACITY and ASCEND, and through the RECAP extension period,[40, 56] for whom data are available through to 8.8 years (see Figure 22).

Figure 22: RECAP Kaplan Meier estimates for Overall Survival: patients continuing on pirfenidone 2,403mg per day (data cut: June 2015, reproduced from CS,[4] Figure 35)



Time on treatment data for these patients from the latest data-cut of RECAP are presented in Figure 23.

Figure 23: RECAP Kaplan Meier estimates for time on treatment: patients continuing on pirfenidone 2,403mg per day (data cut: June 2015) (reproduced from CS,[4] Figure 36)



The CS,[4] (pages 116-117)[4] also reported the following data [redacted]

No details were provided in the CS[4] about how this study was identified or whether any other potentially relevant studies were excluded (for example, a second non-randomised open-label study, PIPF-002,[65] is only mentioned in the safety section of the CS[4] (Section 4.12); it was not reported how data were extracted or analysed, and no critical appraisal was conducted by the company or reported in the CS,[4] so the risk of bias affecting the RECAP study[40] is unknown.

However, it has been stated by Kreuter et al[63] that, *"the RECAP data must be interpreted with caution due to possible selection bias with regard to both pirfenidone (patients selected for tolerability and treatment response) and placebo (selection for mild progression because death or significant worsening led to informed drop out)."*

Registry data

The CS[4] then used IPD from selected registries with the aim of providing potential long-term comparative data for RECAP[40] based on "best supportive care." The CS[4] stated that the company contacted the holders of various registries reporting outcomes for patients with IPF in real-world practice were contacted, resulting in the availability of patient-level information from three registries: the Edinburgh registry, INOVA registry and the EuroIPF registry (see Table 26).

Table 26: Summary of available registries for best supportive care, registries with patient level demographic data (reproduced from CS,[4] Table 58, pages 164-165)

	Edinburgh	**INOVA**	**EuroIPF**
Geographic Region	UK	USA	Europe
Dates of registry information	1 January 2001 – 30 May 2014	November 1996 - June 2015	2008 - 2011
Patient population	• Incident IPF cases with a definite or possible UIP pattern on HRCT based on the 2011 ATS/ERS diagnostic guidelines for IPF • Event time available • Patients diagnosed up to 48 months prior to data collection date	Confirmed as incident IPF cases based on the 2011 ATS/ERS/JRS/ALAT diagnostic guidelines for IPF.	Verified diagnosis of IPF
n	323	815	409
Follow-up	Patients were followed from index date (date of IPF diagnosis) to date of death or May 30, 2014. Vital status was ascertained on May 30, 2014. Patients were censored on May 30, 2014, if their death could not be confirmed. None of the patients seen at this center underwent lung transplantation during the follow-up period, so this was not included as a censoring criterion for this cohort.	Patients were followed from index date (date of IPF diagnosis) to date of death or date of last visit. Date of last vital status is provided in the dataset. Patients were censored on their date of last visit, if their death could not be confirmed. If patients had a transplant, it was indicated in the dataset, but no dates were provided for treatment or transplant.	Patients were followed from index date (date of inclusion in registry) to date of death or date of last visit. Date of last visit and vital status check was provided. Patients were censored on date of last visit, if their death could not be confirmed
Treatments received during follow-up	BSC only	BSC only	BSC only
Inclusion/exclusion criteria applied to match ASCEND/CAPACITY	• FVC/VC<90% and/or DLco<90% • FVC/VC>50% • DLco >30% • FEV1/FVC>0.7 • Age 40 - 80	• FVC/VC<90% and/or DLco <90% • FVC/VC>50% • DLco >30% • FEV1/FVC>0.7 • Age 40 - 80	• FVC/VC<90% and/or DLco <90% • FVC/VC>50% • DLco >30% • FEV1/FVC>0.7 • Age 40 - 80

	• Gender known	• Gender known • Event time available	• Gender known • Event time available
Number of patients following application of ASCEND/ CAPACITY inclusion/ exclusion criteria	182	286	115
Parameters included in the propensity score model	• Age • Sex • Baseline %predicted FVC • Baseline %predicted DLco • First order interaction terms	• Age • Sex • Baseline %predicted FVC • Baseline %predicted DLco • Baseline FEV/FVC • First order interaction terms	• Age • Sex • Baseline %predicted FVC • Baseline %predicted DLco • Baseline FEV/FVC • Baseline smoking status • First order interaction terms
Number of patients remaining after trimming	125	254	89
Age, mean years ± SD	69.4 ± 7.6	66.2 ± 7.9	66.3 ± 8.4
Male (%)	72%	80%	85%
FVC ± SD	81.2 ± 12.4	70.9 ± 12.8	75.4 ± 14.3
DLco **± SD**	51.6 ± 11.8	46.5 ± 11.1	46.0 ± 10.6
FEV1/FVC ± SD	0.83 ± 0.07	0.83 ± 0.06	0.83 ± 0.07
Propensity score model	logOdds(Trial=1) = Age + Sex + DLco + FVC + Age* DLco + Age*FVC	logOdds(Trial=1) = Age + Sex + DLco + FVC +FEV/FVC + Age* DLco + Sex*FEV/FVC	logOdds(Trial=1) = Age + Sex + DLco + FVC +FEV/FVC + Smoke + Age*FVC + Age*Sex + Age*FEV/FVC + Sex*FVC + Sex*Smoke + DLco *Smoke
Key: *DLco, carbon monoxide diffusing capacity of the lungs; FEV, forced expiratory volume; FVC, forced vital capacity; HRCT, high-resolution computed tomography; IPF, idiopathic pulmonary fibrosis; SD, standard deviation; UIP, usual interstitial pneumonia.*			

To improve the comparability between the data from the ASCEND[34] and CAPACITY trials[49] and these registry data, a two-stage process was conducted. The company selected from these registries: (1) those patients that were considered most likely to satisfy the eligibility for the RECAP trial by applying the ASCEND and CAPACITY trials' inclusion and exclusion criteria, and; (2) applied a propensity score model that calculated the probability of being included in a clinical trial based on baseline characteristics, and excluding patients with unusual profiles based upon propensity-score based trimming (see CS,[4] page 162).

The CS[4] (page 163) argued that, based on the kernel density distributions for each of the logistic models post trimming for each of three registries, the INOVA and EuroIPF registries provided the most comparable patient sample to the patients in the pirfenidone Phase III RCTs. The comparative effectiveness estimated across the three registries was comparable with or better than the comparative effectiveness observed in the pooled ASCEND/CAPACITY data.[34, 49] Results were similar comparing the pooled hazard ratio versus BSC from ASCEND/CAPACITY[34, 49] and INOVA which represented the study with the largest sample size and most similar patient characteristics post trimming (HR 0.52 versus [redacted] see Table 27).[4] The CS[4] accepts that there are limitations in comparing data from a Phase III trial with real-world evidence.

Table 27: Overall survival comparison: pirfenidone versus BSC (from registry data) (reproduced from CS,[4] Table 57, page 163)

Outcome	Edinburgh registry	INOVA registry	EuroIPF registry	Pooled CAPACITY and ASCEND data
Hazard ratio for pirfenidone vs BSC (post trimming unadjusted data)	[redacted]	[redacted]	[redacted]	0.64 (0.41;0.99) at 72 weeks
Hazard ratio for pirfenidone vs BSC (post trimming data using propensity score model to adjust for remaining imbalances)	[redacted]	[redacted]	[redacted]	0.52 (0.31; 0.88) at 52 weeks
Key: BSC, best supportive care.				

In addition to the registries where IPD were available, three additional sources of supportive information were provided on probability of survival:

1. CPRD data (n=4,527) were obtained from 2000 to 2012 (inclusive), before pirfenidone was available in the UK .[66] Patients were selected based on the following criteria:

- A clinical or referral event record for IPF as defined by Read (general practices coding system in the UK) as specified in Navaratnam 2011.[67]
- No clinical or referral codes for connective tissue disease, extrinsic allergic alveolitis, sarcoidosis, pneumoconiosis, or asbestosis at any time in the patient record
- IPF events whilst alive and registered at an up-to-standard general practice
- At least 1 year of registration prior to the index date (date of IPF record)

To improve the similarity between the CAPACITY[49] and CPRD cohorts, the following restrictions were applied to the CPRD data:

- Survival times were adjusted using random-sampling of diagnosis to randomisation collected in the CAPACITY studies (n=2,888)[49]
- Patients with an FVC<50% were excluded, this was determined based on data within 1 month of the patient's index date (n=193)

Full propensity scoring was not possible as only FVC data were available for patients within the CPRD dataset. Standard care patients were followed up to 9.53 years; a median survival of 3.41 years was observed (95% CI: 2.67, 4.93).

2. Strand *et al.*[41] report overall survival for patients prospectively enrolled from the National Jewish Health Institutional Review Board-approved ILD database for patients between January 1, 1985 and January 1, 2011 diagnosed with IPF according to consensus guidelines. Median survival was 4.4 years (95% CI: 4.1-5.2) for IPF.

3. Kondoh *et al.*[50] retrospectively studied patients diagnosed with IPF based on ATS/ERS criteria.[50] Median survival was 3.7 years. A stepwise multivariate Cox regression model demonstrated the prognostic significance of FVC progression (10% decline in FVC at 6 months), acute exacerbations, BMI and disease severity measured via the modified MRC scale.

A summary of the characteristics of the patients contained within the three additional registries, and the patients in the CAPACITY/ASCEND trials,[34, 49] is provided in Table 28. The CS[4] states that the patients within the Strand registry[41] appear to be most similar to those in the CAPACITY[49] and ASCEND[34] trials.

Table 28: Summary of available registries for best supportive care, registries without patient level demographic data (reproduced from CS,[4] Table 59, page 168)

	CPRD	Strand[41]	Kondoh[50]	CAPACITY / ASCEND[34, 49]
Geographic region	UK	USA	Japan	Global
Data collection dates	2000 - 2012	Jan 1985 – Jan 2011	Jan 2000 - Dec 2005	
Patient population	ICD10 codes: H563.00 H563.11 H563.12 H563100 H563z00	Subgroup diagnosed with IPF according to consensus guidelines including ATS/ERS	Patients diagnosed with IPF based on ATS/ERS criteria	Diagnosis of IPF in accordance with the ATS international consensus statement
n	193 in FVC reported and ≥50 subgroup	321	74	623 on high dose PFN arms
Age, mean years ± SD	73.5 ± 9.2	66.1 ± 9.1	64.1 ± 7.4	67.2 ± 7.6
Male (%)	68%	75%	82%	74%
FVC ± SD	79.3 ± 15.7	71.4 ± 17.4	77.0 ± 19.2	67.8 ± 11.2
DLco ± SD	NR	52.3 ± 18.7	59.3 ± 18.7	47.1 ± 9.7

Key: DLco, carbon monoxide diffusing capacity of the lungs; FEV, forced expiratory volume; FVC, forced vital capacity; HRCT, high-resolution computed tomography; IPF, idiopathic pulmonary fibrosis; SD, standard deviation; UIP, usual interstitial pneumonia; PFN, pirfenidone; NR, not reported

4.5 Safety evidence

Serious adverse events

In the ASCEND trial[34] there were 55 patients (19.8%) and 69 patients (24.9%) in the pirfenidone and placebo groups, respectively, who experienced a serious AE (see Table 29). The most common serious AE was "worsening of IPF", which was reported in 7 patients (2.5%) in the pirfenidone group and 27 patients (9.7%) in the placebo group. According to trial protocols, "worsening of IPF" is defined as, "acute IPF exacerbation, IPF-related death, lung transplant or respiratory hospitalization, whichever comes first."[36] "Worsening of IPF" was not specifically categorised as either an efficacy outcome (see CS, [4] pages 96-99 and 104-107) or a safety outcome (unless it could be designated as certainly due to the drug), but its presence was simply reported by the investigator (see clarification response,[10] question A18). The other most frequently-reported serious AEs in the pirfenidone arm were pneumonia, prostate cancer, angina pectoris, nausea, congestive cardiac failure and rib fracture (see Table 29). Other than the more frequent occurrence of "worsening of IPF" in the placebo arm, none of the differences in serious AEs between arms was statistically significant.

Table 29: Serious treatment-emergent adverse events reported by ≥2 patients in ASCEND at 52 weeks (reproduced from clarification response,[10] question A25)

Adverse event	Number of patients, n (%)		Rate ratio (95% CI)	Pr>chi2
	PNF 2,403mg/d (n=278)	Placebo (n=277)		
Worsening of Idiopathic Pulmonary Fibrosis	7 (2.5)	27 (9.7)	0.26 (0.11, 0.58)	<0.001
Pneumonia	11 (4.0)	14 (5.1)	0.78 (0.36, 1.69)	0.533
Prostate Cancer (*M)	2 (0.7)	4 (1.4)	0.50 (0.09, 2.70)	0.409
Angina Pectoris	3 (1.1)	0 (0.0)		0.083
Nausea	3 (1.1)	0 (0.0)		0.083
Atrial Fibrillation	1 (0.4)	2 (0.7)	0.50 (0.05, 5.46)	0.561
Bronchitis	1 (0.4)	2 (0.7)	0.50 (0.05, 5.46)	0.561
Dyspnoea	1 (0.4)	2 (0.7)	0.50 (0.05, 5.46)	0.561
Pulmonary Embolism	1 (0.4)	2 (0.7)	0.50 (0.05, 5.46)	0.561
Septic Shock	1 (0.4)	2 (0.7)	0.50 (0.05, 5.46)	0.561
Cardiac Failure Congestive	2 (0.7)	0 (0.0)		0.157
Rib Fracture	2 (0.7)	0 (0.0)		0.157
Aortic Aneurysm	0 (0.0)	2 (0.7)	0.00	0.156
Gastroenteritis Viral	0 (0.0)	2 (0.7)	0.00	0.156

*Each patient is counted only once for each preferred term. For terms followed by (*M), percentages are based on the number of males within each treatment group. Preferred terms are listed in order of decreasing frequency in the total study population.*
TE SAE = treatment-emergent serious adverse events, defined as occurring after the first dose and within 28 days after the last dose of study treatment.

The CS[4] did not report any serious AEs for CAPACITY 1 & 2, but these were reported in the publication[49] and the clarification response,[10] (question A25). The principal serious AEs for pirfendione, excepting IPF, occurring in >2 patients in any pirfenidone group are reported in Table 30.

Table 30: 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

* Male patients only

None of the differences in serious AEs between arms, including IPF, were statistically significant within the CAPACITY trials[49] (see clarification response,[10] question A25).

The publications for SP3[38] did not report AE data, but these were provided by the company in response to a request for clarification from the ERG (see clarification response addendum,[30] question A28)[10]: serious AEs occurring in ≥1% of participants in the pirfenidone arm were pneumonia (5.5% for pirfenidone versus 2.8% for placebo), bronchitis (1.8% versus 1.9%), worsening of IPF (5.5% versus 4.7%) and pneumothorax (1.8% versus 2.8%). The SP2[39] publications did not report any serious AEs and there was no additional information available for this trial (see clarification response,[10] question A25).

The company conducted a meta-analyses for treatment-emergent serious AEs using data from ASCEND,[34] CAPACITY 1 & 2[49] and SP3[38] at week 52 (see Figure 24) and at 72 weeks (see Figure 25) using data from CAPACITY 1 & 2[49] only (see CS,[4] Appendix 9, page 76). Both analyses showed no difference between the pirfenidone and placebo groups (OR: 0.90, 95% CI: 0.70 to 1.15, *p*-value not reported) and (OR: 1.06, 95% CI: 0.77 to 1.46, *p*-value not reported).

Figure 24: Forest plot of odds ratios for treatment emergent serious adverse events at week 52 (reproduced from CS,[4] Appendix 9)

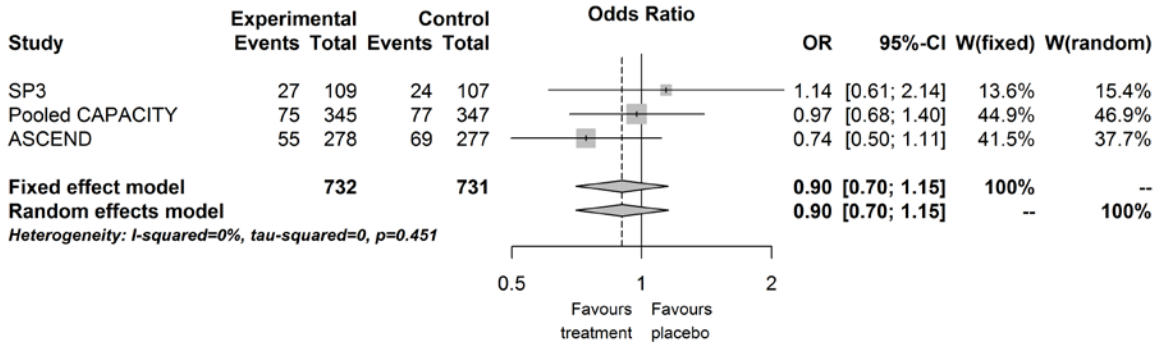
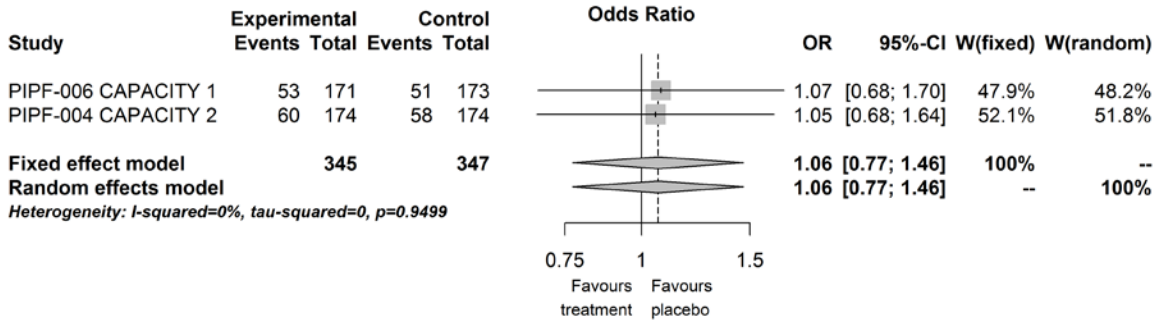


Figure 25: Forest plot of odds ratios for treatment emergent serious adverse events at week 72 (reproduced from CS,[4] Appendix 9)



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 occurred in 3 patients (1.1%) in each trial arm.

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.

SP3[38] reported that 15 patients in the high dose group (1,800mg/d) and 9 patients in the low dose group (1,200mg/d), compared with 7 patients in the placebo group, discontinued the study due to AEs. The CS did not report details of these adverse events, but the publication did so:[38] the principal adverse events affecting discontinuation from pirfenidone treatment were: photosensitivity (n=5); lung carcinoma (n=3); fever (n=2); respiratory failure (n=2); rash (n=2) and an increase in aspartate aminotransferase (AST) and/or alanine amino-transferase (ALT) (n=2).

All adverse events

The most common "treatment-emergent" AEs with higher incidence in the pirfenidone group were primarily gastrointestinal and skin-related events. The CS[4] reported data for any AE with a frequency of at least 15% in any arm (in ASCEND)[34] or a frequency of at least 10% and 1.5 times in the pirfenidone arm compared with the placebo arm (in the CAPACITY trials).[49] Nausea was the most frequent AE: 36% in the pirfenidone arm compared with 13.4% in the placebo arm in ASCEND, and 36% in the pirfenidone arm compared with 17% in the placebo arm CAPACITY trials (*p*-values not reported, see Table 31).[49] The second most frequent event was rash: 28.1% in the pirfenidone arm compared with 8.7% in the placebo arm in ASCEND,[34] and 32% in the pirfenidone arm compared with 12% in the placebo arm in the CAPACITY trials.[49] Dyspepsia was also much more frequent in the pirfenidone arms than the placebo arm: 17.6% in the pirfenidone arm compared with 6.1% in the placebo arm in ASCEND,[34] and 19% in the pirfenidone arm compared with 7% in the placebo arm in the CAPACITY trials;[49] as was anorexia: 15.8% in the pirfenidone arm compared with 6.5% in the placebo arm in ASCEND,[34] and 11% in the pirfenidone arm compared with 4% in the placebo arm in the CAPACITY trials;[49] and dizziness: 17.6% in the pirfenidone arm compared with 13% in the placebo arm in ASCEND, and 18% in the pirfenidone arm compared with 10% in the placebo arm in the CAPACITY trials.[49]

Headache, cough, diarrhoea, fatigue and upper respiratory tract infection were all frequent (between 20% and 26%), but were similar across pirfenidone and placebo arms (see Table 31) According to the CS[4] (page 171), no instances of Stevens-Johnson syndrome or toxic epidermal necrosis were reported in the CAPACITY trials.[49]

Table 31 Adverse events in ASCEND at 52 weeks and CAPACITY 1 & 2 at 72 weeks (adapted from CS,[4] Tables 60 and 61)

	ASCEND*[34]		CAPACITY 1 & 2†[49]	
Adverse event, n (%)	**PFN (n=278)**	**PBO (n=277)**	**PFN (n=345)**	**Placebo (n= 347)**
Nausea	100 (36)	37 (13.4)	125 (36)	60 (17)
Rash	78 (28.1)	24 (8.7)	111 (32)	40 (12)
Headache	72 (25.9)	64 (23.1)		
Cough	70 (25.2)	82 (29.6)		
Diarrhoea	62 (22.3)	60 (21.7)		
Upper respiratory tract infection	61 (21.9)	56 (20.2)		
Fatigue	58 (20.9)	48 (17.3)		
Dizziness	49 (17.6)	36 (13)	63 (18)	35 (10)
Dyspepsia	49 (17.6)	17 (6.1)	66 (19)	26 (7)
Anorexia	44 (15.8)	18 (6.5)	37 (11)	13 (4)
Dyspnoea	41 (14.7)	49 (17.7)		
Vomiting			47 (14)	15 (4)
Photosensitivity reaction			42 (12)	6 (2)
Anorexia			37 (11)	13 (4)
Arthralgia			36 (10)	24 (7)
Insomnia			34 (10)	23 (7)
Abdominal distension			33 (10)	20 (6)

** Occurring in ≥15% of patients in either treatment group; † Occurring in ≥10% of patients on pirfenidone and with an incidence of 1.5 x greater than that in patients receiving placebo; PFN: pirfenidone 2,403mg/day; PBO: placebo*

The SP3[38] and SP2[39] trials also reported a relatively high incidence of the following AEs for pirfenidone compared with placebo: photosensitivity; anorexia; dizziness; nausea; heartburn; fatigue and elevated gamma-GTP (see Table 32). *P*-values were reported for the SP2 trial[39] and the incidence of many of the AEs was significantly higher in the pirfenidone group than the placebo group (see Table 32). Respiratory infections were reported to be more common in patients treated with placebo.

The CS[4] (page 173) stated that most of the AEs reported for SP2[39] disappeared with decrease of the dose or temporarily holding the medication.

It is unclear why there is some inconsistency between trials in the frequency of some AEs, such as photosensitivity, nausea and anorexia.

Findings on AEs from Huang *et al*[48] were consistent with the other published trials, including, for example, the significantly higher incidence of rash in patients receiving pirfenidone.

Table 32: Adverse events reported from the SP3 at 52 weeks and SP2 at 26 weeks (adapted from CS,[4] Tables 62 and 63)

	SP3*[38]			SP2†[39]	
Adverse event n (%)	**PFN 1,800mg/d (n=109)**	**PFN 1,200mg/d (n=55)**	**Placebo (n=107)**	**PFN 1,800mg/d (n=72)**	**PBO (n=32)**
Any adverse event§	109 (100.0)	54 (98.2)	106 (99.1)	72 (98.6)	32 (88.9)
Photosensitivity§	56 (51.4)	29 (52.7)	24 (22.4)	32 (43.8)	0 (0)
Anorexia§	18 (16.5)	6 (10.9)	3 (2.8)	23 (31.5)	2 (5.6)
Abdominal discomfort	3 (2.8)	4 (3.7)	0 (0.0)	22 (30.1)	3 (8.3)
Nausea§				16 (21.9)	2 (5.6)
Heartburn				12 (16.4)	1 (2.8)
Fatigue§				16 (21.9)	1 (2.8)
Dizziness	8 (7.3)	0 (0.0)	1 (0.9)		
Nasopharyngitis	54 (49.5)	30 (54.5)	70 (65.4)		
Upper respiratory tract infection	1 (0.9)	3 (5.5)	9 (8.4)	12 (16.4)	3 (8.3)
γ-GTP elevation§	25 (22.9)	12 (21.8)	10 (9.3)	20 (27.4)	3 (8.3)

***** With an incidence of ≥5%; † With an incidence of ≥10% at six months; § Difference between pirfenidone 1800mg per day and placebo is significant at level of $p<0.05$ or better in trial SP2; PFN: pirfenidone 2,403mg/day; PBO: placebo

Integrated analysis of safety data from ASCEND, CAPACITY 1 & 2, and two ongoing open-label studies

Data from the three principal Phase III trials (ASCEND,[34] CAPACITY 1 & 2[49]) were analysed together with data from the non-randomised, non-controlled, OLE study that included a set of patients who completed either ASCEND, CAPACITY 1 or 2 (RECAP)[40] (see Section 4.4) and PIPF-002, an ongoing open-label compassionate-use study in US patients with either IPF or secondary pulmonary fibrosis.[68] No critical appraisal was reported for either the RECAP[40] or the PIPF-002 study.[65] Safety outcomes were assessed from baseline until 28 days after study drug discontinuation.

The latest interim analyses of the integrated population were conducted using a data cut-off date of 17 January 2014.[69] A total of 1,299 patients were included in the integrated population and the reported data only concern AEs occurring in at least 15% of patients in the cumulative clinical database. The cumulative total exposure to pirfenidone was 3,160 person exposure years (PEY). The median duration of exposure was 1.7 years (range, 1 week–9.9 years); 545 (42%) patients received pirfenidone for ≥2 years and 325 (25%) patients received pirfenidone for ≥4 years. The majority of patients (n=964, 74.2%) received a mean daily dose between 1,800mg and 2,600mg. Cumulative safety outcomes in the pooled pirfenidone 2,403mg/day and placebo treatment groups in the Phase III studies were presented for comparison (see Table 33).

Table 33: AEs in the integrated population compared with the pooled pirfenidone 2,403mg/day and placebo groups from the ASCEND and CAPACITY 1 & 2 trials* (reproduced from CS,[4] Table 64)

	Integrated population† (n=1,299)	Pooled ASCEND, CAPACITY 1 & 2 population	
		PFN (n=623)	PBO (n=624)
Median duration of exposure, years (range)	1.7 (>0, 9.9)	1.0 (>0, 2.3)	1.0 (>0, 2.3)
Treatment-emergent adverse event, %			
Nausea	37.6	36.1	15.5
Cough	35.1	27.8	29.2
Dyspnoea	30.9	16.9	20.2
Upper respiratory tract infection	30.6	26.8	25.3
Idiopathic pulmonary fibrosis	29.3	13.0	19.9
Fatigue	28.2	26.0	19.1
Diarrhoea	28.1	25.8	20.4
Rash	25.0	30.3	10.3
Bronchitis	23.8	14.1	15.4
Headache	21.6	22.0	19.2
Nasopharyngitis	21.3	16.7	17.9
Dizziness	21.2	18.0	11.4
Dyspepsia	18.4	18.5	6.9
Vomiting	15.9	13.3	6.3
Weight decreased	15.6	10.1	5.4
Back pain	15.4	10.4	10.4
Anorexia	15.2	13.0	5.0

**Occurring in ≥15% of patients in the cumulative clinical database*
†Includes two patients from PIPF-002 with a diagnosis of "pulmonary fibrosis"
PFN: pirfenidone 2,403mg/d; PBO: placebo

The findings for the integrated population are consistent with the findings of the ASCEND[34] and CAPACITY trials[49] (though not always with the SP3[38] and SP2[39] trials), i.e. gastrointestinal and skin-related events were among the most common AEs. The CS[4] (page 174) states that these were mainly mild to moderate in severity, reversible, and rarely led to treatment discontinuation. Elevations in liver enzymes (ALT or AST >3 x Upper Limit of Normal [ULN]) occurred in 40/1,299 (3.1%) patients in the integrated population, compared with 23/623 (3.7%) and 5/624 (0.8%) in the pooled pirfenidone and placebo groups in the Phase III trials. All elevations were reversible without clinical sequelae. Respiratory AEs were more common in the integrated population than the placebo and pirfenidone-treated patients from the pooled Phase III trials. The CS[4] (page 175) states that this finding is expected from a chronic progressive respiratory disease followed over a long period of observation.

The CS[4] stated that the safety and AE profile of pirfenidone is different from that of nintedanib, for which most frequently reported adverse reactions are diarrhoea, nausea and vomiting, abdominal pain, decreased appetite, weight loss and elevation of hepatic enzymes.[1] However, the pirfenidone trials did

report nausea, diarrhoea, vomiting, weight loss and anorexia as frequent events (see Table 30 and Table 31).

4.5.1 Ongoing studies

As noted above, there are two ongoing studies to evaluate safety: the non-randomised, non-controlled, OLE study that included a set of patients who completed either ASCEND, CAPACITY 1 or CAPACITY 2 (RECAP) (see Section 4.5, the final data collection date is listed as December 2015 https://clinicaltrials.gov/ct2/show/record/NCT00662038) and PIPF-002, an ongoing open-label compassionate-use study in US patients with either IPF or secondary pulmonary fibrosis (https://clinicaltrials.gov/ct2/show/NCT00080223), which has a listed completion date of April 2015.

4.6 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

In the absence of any direct head-to-head RCTs comparing pirfenidone and nintedanib, for the treatment of IPF, the company conducted an NMA. This is an extension of the conventional pairwise meta-analysis that can be used to combine direct and indirect evidence about treatment effects across RCTs that share at least one treatment in common with at least one other study.

The company conducted a systematic review to collate the published RCTs which assess the efficacy and safety of therapies prescribed for the treatment of IPF. The inclusion criteria for the NMA systematic review were as follows (see CS,[4] pages 122-123): the population of interest was adults (aged 18 or older) with suspected or diagnosed IPF; the interventions of interest were pirfenidone, double therapy (prednisone and azathioprine), N-acetylcysteine (NAC), nintedanib, and triple therapy (prednisone and azathioprine and NAC); the relevant study designs were Phase II or Phase III RCTs and the outcomes of interest included lung capacity, gas transfer, physical functioning (6MWD), PFS, adverse effects of treatment, HRQoL measured using SGRQ, SOBQ, dyspnoea score or EQ-5D, hospitalisations, acute exacerbations, mortality (all cause or IPF-related), categorical declines in FVC (0%, 5% and 10%), discontinuation and compliance of study treatments.

The systematic review methods undertaken for the NMA (e.g. literature searching, study selection, data extraction and quality assessment) were the same as those undertaken for the pirfenidone systematic review. As noted in Section 4.1.1, adequate systematic searches were undertaken to identify all relevant RCT studies assessing the efficacy and safety of NAC, nintedanib and triple therapy for the treatment of IPF. NAC, triple therapy and double therapy were not included in the NICE scope as comparators of interest, however, the company's literature search was developed to support submissions of pirfenidone to all national agencies and as such some comparators of interest included in the searches were beyond the scope of this appraisal.

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[54] (pirfenidone), CAPACITY 1[49] (pirfenidone), CAPACITY 2[49] (pirfenidone), SP3[38] (pirfenidone), IMPULSIS 1[72] (nintedanib), IMPULSIS 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 34: Summary of trials included in the company's NMA: (adapted from CS,[4] Table 12, page 66-67 and Appendix 10)

Study	Design, Location	Population	Treatment, dose and sample size (used in NMA)	Study durations (week)	Key outcomes measured in NMA
Pirfenidone					
ASCEND [34]	Phase III, randomised, double-blind, placebo-controlled trial. **Location** 127 sites (no sites in UK)	• Patients aged 40–80 years with confident clinical and radiographic diagnosis of IPF, in accordance with the International consensus statement [ATS, 2000] of >6 months but <48 months before randomisation, confirmed by central review. • FVC (% predicted value) 50-90% • DLco 30-90% • 6MWD ≥150 m • No improvement of IPF in preceding year.	Pirfenidone 2,403mg/day (n=278) Placebo (n=277)	52 weeks	**Primary outcomes** Change in percent predicted FVC or death at week 52. **Secondary outcomes** Change from baseline to Week 52 in 6MWD and PFS, change in dyspnoea (UCSD SOBQ); rate of death from any cause and the rate of death from IPF.
CAPACITY 2[49]	Phase III, randomised, double-blinded, placebo-controlled trial **Location** 110 centres (including 3 sites in the UK)	• Patients aged 40–80 years with confident clinical and radiographic diagnosis of IPF, in accordance with the International consensus statement [ATS, 2000] in the previous 48 months. • FVC (% predicted value) ≥50% at Screening and Day 1 (before randomisation) • DLco ≥35% • FVC or DLco ≤90% • No improvement of IPF in preceding year	Pirfenidone 2,403mg/day, (n=174) Placebo (n=174)	72 weeks	**Primary outcomes:** Change in percent predicted FVC from baseline to week 72. **Secondary outcomes:** Categorical FVC, PFS, worsening IPF, dyspnoea (SOBQ), 6MWD, worst SpO2 during the 6MWT, % predicted DLco, and fibrosis by use of HRCT.

CAPACITY 1[49]	Phase III, randomised, double-blinded, placebo-controlled trial **Location** 110 centres(no UK sites)	• Patients aged 40–80 years with confident clinical and radiographic diagnosis of IPF, in accordance with the International consensus statement [ATS, 2000] in the previous 48 months. • FVC (% predicted value) ≥50% at Screening and Day 1 (before randomisation) • DLco ≥35% • FVC or DLco ≤90% • No improvement of IPF in preceding year	Pirfenidone 2,403mg/day (n=171) Placebo (n=173)	52 weeks	**Primary outcomes:** Change in percent predicted FVC from baseline to week 72. **Secondary outcomes:** Categorical FVC, PFS, worsening IPF, dyspnoea (SOBQ), 6MWD, worst SpO2 during the 6MWT, % predicted DLco, and fibrosis by use of HRCT.
SP3[38]	Phase III, randomised, double-blind, placebo-controlled trial. **Location** 73 centres in Japan.	• Patients aged 20 -75 years, with confident clinical and radiographic diagnosis of IPF in accordance with the International consensus statement [ATS/ERS, 2000]. • O2 desaturation of 5% between resting SpO2 and min SpO2 during 6MET • SpO2 >85% during 6MET (air). • No decrease in symptoms during the preceding 6 months	Pirfenidone 1,800mg/day (n=108) Placebo (n=104)	52 weeks	**Primary outcomes:** Change in VC from baseline to week 52 (originally was the change in lowest SpO2 during the 6MWT). **Secondary outcomes:** PFS time, change in the lowest SpO2 during the 6MWT
Nintedanib					
TOMORROW[73]	Phase II, randomised, double-blind, placebo-controlled trial **Location**	• Patients >40 years of age with diagnosis of IPF in accordance with ATS and ERS criteria and who had received the diagnosis of IPF <5 years before screening	Nintedanib 300mg/day (n=86) Placebo (n=85)	52 weeks	**Primary outcome** The annual rate of decline in FVC. **Secondary outcome** Changes from baseline in percent predicted FVC and DLco; changes in SpO_2 and TLC (as measured by body plethysmography); 6MWD; SGRQ; a

	92 sites in 25 countries including UK	• Patients had to have undergone HRCT <1 year before randomisation • FVC (% predicted value) ≥50% • DLco (% predicted value) 30 to 79% • PaO2 when breathing ambient air that was 55 mm Hg or greater at altitudes up to 1500m or a PaO2 of 50mm Hg or greater at altitudes above 1500 m.			decrease from baseline in FVC of > 10% or > 200 ml; SpO_2 decrease of more than 4%; incidence of acute exacerbations; survival at 52 weeks; and death from a respiratory cause
INPULSIS 1[72]	Phase III, randomised, double-blind, placebo-controlled trial **Location** 98 study sites including UK	• Age > 40 years; • IPF diagnosed, according to most recent ATS, ERS, JRS, ALAT IPF guideline for diagnosis and management, within 5 years; • Combination of HRCT pattern, and if available surgical lung biopsy pattern, as assessed by central reviewers, are consistent with diagnosis of IPF • Dlco (corrected for Hb): 30%-79% predicted of normal; • FVC> 50% predicted of normal	Nintedanib 150mg/bid (n=309) Placebo (n=204)	52 weeks	**Primary outcome** Annual rate of decline in FVC (mL) from baseline to week 52. **Secondary outcome** Time to the first acute exacerbation, change from baseline in SGRQ total score, acute exacerbations, absolute change from baseline in FVC (mL) and as a % predicted value over the 52-week treatment period, proportion of patients with an FVC response, risk of acute exacerbation, change from baseline in SGRQ domain scores over the 52-week treatment period, death from any cause, death from a respiratory cause, and death that occurred between randomisation and 28 days after the last dose of the study drug.
INPULSIS 2[72]	Phase III, randomised, double-blind, placebo-	• Age > 40 years; • IPF diagnosed, according to most recent ATS, ERS, JRS, ALAT IPF guideline for diagnosis and management, within 5 years;	Nintedanib 150mg/bid (n=329) Placebo	52 weeks	**Primary outcome** Annual rate of decline in FVC (mL) from baseline to week 52. **Secondary outcome**

	controlled trial **Location** 108 study sites, no sites in UK	• Combination of HRCT pattern, and if available surgical lung biopsy pattern, as assessed by central reviewers, are consistent with diagnosis of IPF • Dlco (corrected for Hb): 30%-79% predicted of normal; • FVC> 50% predicted of normal	(n=219)		Time to the first acute exacerbation, change from baseline in SGRQ total score, acute exacerbations, absolute change from baseline in FVC (mL) and as a % predicted value over the 52 week treatment period, proportion of patients with an FVC response, risk of acute exacerbation, change from baseline in SGRQ domain scores over the 52 week treatment period, death from any cause, death from a respiratory cause, and death that occurred between randomisation and 28 days after the last dose of the study drug.

Note: only trials relevant to the decision problem are reported
ALAT, Latin American Thoracic Association ATS; American Thoracic Society; bid, twice a day; DLco, Diffusing capacity of the lungs for carbon monoxide; EQ-5D, The EuroQoL Group 5-Dimension Self-Report Questionnaire; ERS, European Respiratory Society; 6MWD, 6-Minute walking distance; 6MWT, 6-Minute walking distance; FVC, Forced vital capacity; HRCT, High-resolution computed tomography; ICECAP, Investigating Choice Experiments for the Preferences of Older People Capability Instrument; IPF, Idiopathic pulmonary fibrosis; JRS, Japanese Respiratory Society; MedDRA, Medical Dictionary for Regulatory Activities; 6MET, 6 min exercise test; mL, millilitres; PaO2, Partial pressure arterial oxygen; PFS, progression-free survival;SF-36; Medical Outcomes Study 36-Item Short-Form Health Survey; SGRQ, St. George's Respiratory Questionnaire; SpO2, Peripheral oxygen saturation; tid, three times a day; TLC, total lung capacity; UCSD SOBQ, University of California San Diego Shortness of Breath Questionnaire and VC, Vital capacity

The main differences noted between the studies relate to definition of the outcomes, patient characteristics, methods used for handling missing data, and the time period of outcome assessment. The CS states that "*due to the limited number of studies contributing to each network, a pragmatic approach was adopted, whereby trials were included regardless of minor differences in outcome definitions, timing of assessment and analysis methods. It was assumed that the differences in definitions and methods did not influence the relative treatment effects*" (CS,[4] page 153). The main sources of heterogeneity are discussed in turn below.

Handling of missing data

In ASCEND[34] and CAPACITY 1 & 2[49], missing values as a result of death were assigned the worst rank in the ANCOVA analyses, and worst possible outcome in mean change analyses (e.g., FVC=0) and categorical analyses. Other missing data were imputed with the average value from three patients with the smallest sum of squared differences at each visit with data that were not missing. For the SP3[38] study and the analysis of secondary endpoints in the TOMORROW trial,[73] LOCF imputation was used when data for the entire 52 week period were not available. In the INPULSIS trials,[72] the statistical model used for the primary analysis allowed for missing data, assuming that they were missing at random; missing data were not imputed for the primary analysis. The company[4] acknowledged that the inclusion of all the trials in the NMA regardless of how missing data were handled may produce bias in the results but strict exclusion criteria on the handling of missing data, could lead to the exclusion of most trials from the network.

Study duration

The time of outcome assessment for data included in the NMA varied (see Table 35 and Table 39). The primary endpoint in the CAPACITY trials[49] was evaluated at 72 weeks with assessments for certain endpoints conducted every 12 weeks. The company considered that data at 48 weeks was the most appropriate data cut-off to use in the NMA so that it could be compared with the 52 week data from the other trials. The CS[4] (page 125) assumes that the treatment effect will be similar across these time points. The ERG asked the company to provide additional analyses to explore the sensitivity of the results to this assumption (discussed in Section 4.7). For a highly progressive disease such as IPF, if trials enrol participants at the same point in their disease course then those with a shorter follow-up might be expected to observe fewer negative outcomes (e.g. exacerbations, decline in lung function, deaths) whilst trials with a longer follow-up would be expected to observe worse outcomes.

Outcome definition

The definitions of the outcomes included also varied. In the SP3[38] study, lung function was reported as VC whilst the remaining trials used FVC. The CS[4] (page 93) stated that *"given that there is little difference between VC and FVC in subjects without obstructive pathology,[76] and IPF patients have a*

restrictive pathology, it is appropriate that VC and FVC are treated as comparable endpoints." The ERG noted that the exclusion criteria for SP3[38] were not as explicit regarding the exclusion of patients with obstructive airway disease as the exclusion criteria for the ASCEND[34] and CAPACITY trials.[49] Therefore the ERG considers that the combination of VC data from SP3[38] with FVC data from the ASCEND[34] and CAPACITY trials[49] is questionable.

Definition of PFS and mortality also differed across the studies. PFS was assessed as composite endpoint and in response to clarification question A14,[10] the company provided the definition of PFS used in each of the trials (see Table 35) and stated that *"To maintain similarity as far as possible, for CAPACITY 1 and 2, the PFS estimate based on the definition used in the ASCEND trial was included in the analysis. For the definitions of SP3, PANTHER and INPULSIS, it is assumed that they will lead to similar hazard ratios and odds ratios between a given pair of treatments, and thus that it is appropriate to combine them in an NMA. We believe this to be a reasonable assumption because in a comparison between the CAPACITY and ASCEND trials, the replacement of DLco by 6MWD led to an increase in qualifying events without changing the HR estimate*."[10] In response to clarification question A33,[10] the company demonstrated that the HRs using the ASCEND[34] definition of PFS provide more conservative estimates of treatment effect (as compared to placebo) than those using the definition utilised in SP3[38] and PANTHER.[74, 75]

In response to clarification question A17,[10] the company confirmed that the definition of OS was the same across all the trials in the NMA; this was defined as patients who died due to any cause (all-cause mortality) in the ITT populations.

Table 35: 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]	IMPULSIS 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

Outcome	ASCEND[34]	*CAPACITY 1 & 2[49]	SP3[38]	IMPULSIS 1&2[72]	TOMORROW[73]	PANTHER[74]
	decline from baseline in percent predicted FVC, confirmed ≥50 m decline from baseline in 6MWD, or death	FVC, ≥15% decline in % predicted DLco or death. In a *post hoc* analysis, the ASCEND definition of PFS was applied to the CAPACITY trials at 52 weeks and at 72 weeks, and used within the NMA			proportion of patients who progressed, rather than the proportion of patients who either progressed or died. It was unclear how many patients progressed before they died and therefore PFS cannot be calculated	≥10% in FVC or death.
Acute Exacerbations	Identified via a *post hoc* analysis of AEs based on the MedDRA lower level term "acute exacerbation of IPF" (CS,[4] page 104)	Defined 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	Defined 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 (PaO_2) is lower by more than 10 Torr than previous	Yes	Yes	Yes (NAC only)

Outcome	ASCEND[34]	*CAPACITY 1 & 2[49]	SP3[38]	IMPULSIS 1&2[72]	TOMORROW[73]	PANTHER[74]
		All other cardiac, thromboembolic, aspiration, infectious processes ruled out	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			
Physical function						
6MWD	Defined as the change from Baseline to week 52 in distance walked during the 6MWD test as measured in metres (m).	Defined as the change from baseline to week 48 in distance walked during the 6WMD test as measured in meters (m).	No	No	Yes	Yes
Health Related Quality of Life						
SGRQ	No	Yes	No	Yes	Yes	No
UCSD SOBQ	The SOBQ is used to assess shortness of breath with various activities of daily living (for example, brushing ones teeth or mowing the lawn). Patients rated the severity of their shortness of breath experienced on an average day during the past week on a 6 point scale (0 to		No	No	No	No

Outcome	ASCEND[34]	*CAPACITY 1 & 2[49]	SP3[38]	IMPULSIS 1&2[72]	TOMORROW[73]	PANTHER[74]
	5),with 0= not at all breathless, 4= severely breathless and 5= Maximally or unable to do because of breathlessness					
All cause discontinuation of treatment	Defined as the count of patients who "did not complete the planned observation time"					

**In CAPACITY 1 & 2[49] assessments were conducted every 12 weeks therefore data at 48 weeks was considered most appropriate to use for comparing with 52 week data from other trials*

*** Note that duration of follow up varies by outcomes.*

CRP, C reactive Protein; DLco, Diffusing capacity of the lungs for carbon monoxide; 6MWD, 6-Minute walking distance; FVC, Forced vital capacity; HRCT High-resolution computed tomography; IPF, Idiopathic pulmonary fibrosis; KL-6, Krebs von den Lungen 6; LDH, Lactate dehydrogenase; MedDRA, Medical Dictionary for Regulatory Activities; mmHg, millimetres of mercury; PaO2, Partial pressure arterial oxygen; PFS, progression-free survival; Sp-A; Surfactant protein A; Sp-D, Surfactant protein D and VC, Vital capacity;

Baseline characteristics

The CS,[4] page 128 notes that there were some differences between the baseline populations in the included trials (see Table 36), but there were no major concerns regarding the inclusion of any of these trials in the network. The CS,[4] page 127 notes that that the populations included in the trials are in line with the licensed indications[1] and the scope[3] and all patients had mild to moderate impairment in pulmonary function at baseline. However, the ERG notes that the SP3[38] study was conducted in a Japanese population and used a lower dose of pirfenidone (1,800mg/day) than that licensed in the UK (2,403mg/day). The CS,[4] page 127 notes that the difference in dosage reflects the difference in mean weights in the North American and European population compared to the Japanese population, hence the trials are comparable. However, the ERG notes that the INPULSIS trials,[72] which compared nintedanib with placebo, also had a high Japanese contingent compared with the other trials assessed, but no reported dose adjustments were made in these studies. The ERG is unsure how this would impact on the evaluation of effectiveness and safety of the therapy.

Despite stating that patients had mild to moderate impairment in pulmonary function at baseline, the measure of function was reported inconsistently across trials at baseline (see Table 36). The ASCEND[34] and the CAPACITY trials[49] used percentage predicted FVC and percentage predicted DLco; SP3 trial[38] used percentage predicted total lung capacity and vital capacity; TOMORROW[73] and the INPULSIS[72] trials used percentage predicted FVC and DLco **(**ml/min/mm Hg). As highlighted in CS,[4] page 82 patients recruited in the ASCEND trial[34] were at higher risk of disease progression with a reported percentage predicted FVC approximately 7-8% lower than the CAPACITY trials.[49]

The time since patients were diagnosed with IPF varied between the trials. Approximately half of the patients in the CAPACITY trials had a diagnosis for less than 1 year,[49] whilst the majority of patients in the remaining trials had been diagnosed for just over 1 year and 38% of patients in SP3[38] had disease duration of greater than 3 years. The ERG notes that due to the progressive and unpredicted clinical course of IPF, difference in disease duration will have an impact on outcomes as reported in the company's subgroup analysis: *"There was evidence of an interaction between treatment and time from IPF diagnosis to randomisation, with those patients diagnosed more than a year before randomisation experiencing a significantly greater treatment effect"* (CS,[4] page 113).

Table 36: Summary of baseline characteristic of trials included in the company's NMA: (CS,[4] Table 16, page 83 and Appendix 10)

	CAPACITY 2 [49]		ASCEND [34]		CAPACITY 1 [49]		SP3[38]		PANTHER (NAC) [74]		TOMORROW [73]		INPLUSIS 1[72]		INPLUSIS 2[72]	
	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 1800 mg/day	Placebo	NAC 600mg/tid	Placebo	Nintedanib 300mg	Placebo	Nintedanib 150mg/bid	Placebo	Nintedanib 150mg/bid	Placebo
N	174	174	278	277	171	173	108	104	133	131	86	85	309	204	329	219
Mean Age, years (SD)	65.7 (8.2)	66.3 (7.5)	68.4 (6.7)	67.8 (7.3)	66.8 (7.9)	67.0 (7.8)	65.4 (6.2)	64.7 (7.3)	68.3 (8.4)	67.2 (8.2)	65.4 (7.8)	64.8 (8.6)	66.9 (8.2)	66.9 (8.4)	67.1 (7.5)	66.4 (7.9)
Males (%)	118 (68)	128 (74)	222 (79.9)	213 (76.9)	123 (72)	124 (72)	85 (78.7)	81 (77.9)	107 (80.5)	98 (74.8)	65 (76.5)	63 (74.1)	163 (79.9)	251 (81.2)	171 (78.1)	256 (77.8)
White (%)									94.7	95.5	71.8	76.5	66.2	64.1	51.6	49.2
Previously smoked (%)	63.0	66.0	66.2	61.0	66.0	58.0	75.0	67.3	70.5	71.0			70.6	70.2	63.5	66.3
Never smoked (%)									27.3	25.2			25	23.0	32.4	31.3
Currently smokes (%)									2.3	3.8			4.4	6.8	4.1	2.4
Definite IPF (HRCT)	159 (91)	164 (94)	266 (95.7)	262 (94.6)	149 (87)	158 (91)			103 (77.4)	99 (75.6)	33 (38.8)	24 (28.2)				
Mean time since IPF diagnosis, years (SD)	*1.3 (0.96)	*1.4 (1.12)	*1.7 (1.1)	*1.7 (1.1)	*1.2 (1.09)	*1.1 (1.04)	38 (35.2) $\leq$ 1y 29 (26.9) 1-3y 41 (38.0) >3y	20 (28.0) $\leq$ 1y 17 (24.0) 1-3y 35 (49.0) >3y	1.0 (1.0)	1.1 (1.0)	1.0 (1.2)	1.4 (1.5)	1.6 (1.4)	1.7 (1.4)	1.6 (1.3)	1.6 (1.3)
Desaturation <80% during 6MWT							34 (31.5)	24 (23.1)								
Mean (SD) 6MWD (m)	411.1 (91.8)	410.0 (90.0)	415.0 (98.5)	420.7 (98.1)	378.0 (82.2)	399.1 (89.7)			371.4 (115.5)	375.4 (104.7)						

	CAPACITY 2 [49]		ASCEND [34]		CAPACITY 1 [49]		SP3 [38]		PANTHER (NAC) [74]		TOMORROW [73]		INPLUSIS 1 [72]		INPLUSIS 2 [72]	
	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 1800 mg/day	Placebo	NAC 600mg/tid	Placebo	Nintedanib 300mg	Placebo	Nintedanib 150mg/bid	Placebo	Nintedanib 150mg/bid	Placebo
Mean (SD) SpO_2 % predicted							89.0 (2.3)	89.0 (2.0)	95.75 (2.45)	96.12 (2.3)	95.6 (1.7)	95.3 (2.2)	95.9 (1.9)	95.9 (2.0)	95.7 (2.1)	95.8 (2.6)
Mean (SD) FVC % predicted	74.5 (14.5)	76.2 (15.5)	67.8 (11.2)	68.6 (10.9)	74.9 (13.2)	73.1 (14.2)			72.2 (15.9)	73.4 (14.3)	79.1 (18.5)	81.7 (17.6)	80.5 (17.3)	79.5 (17.0)	78.1 (19.0)	80.0 (18.1)
Mean (SD) FVC (L)									2.9 (0.8)	2.9 (0.8)	2.7 (0.8)	2.8 (0.8)	2.76 (0.74)	2.85 (0.82)	2.67 (0.78)	2.62 (0.79)
Mean (SD) DLco % predicted	46.4 (9.5)	46.1 (10.2)	43.7 (10.5)	44.2 (12.5)	47.8 (9.8)	47.4 (9.2)	52.1(16.8)	55.2 (18.2)	44.7 (10.8)	46.0 (12.2)						
Mean (SD) DLco (ml/min/mm Hg)									13.2 (3.7)	13.5 (3.8)	3.7 (1.0)	3.8 (1.1)	4.0 (1.2)	4.0 (1.1)	2.7 (1.3)	3.8 (1.2)
Mean (SD) PaO_2							79.8 (10.2)	17.4 (9.7)	80.7 (10.5)	81.5 (11.8)	79.6 (13.3)	76.5 (14.1)				
Mean (SD) $P(A\text{-}a)O_2$							18.4 (11.3)	17.4 (9.7)	17.81 (9.95)	17.34 (10.96)						
Mean (SD) VC % predicted							77.3 (16.8)	79.1 (17.4)					79.5 (17)	80.5 (17.3)	80 (18.1)	78.1 (19)
Mean (SD) VC (L)							2.40 (0.64)	2.47 (0.70)								

	CAPACITY 2 [49]		ASCEND [34]		CAPACITY 1 [49]		SP3 [38]		PANTHER (NAC) [74]		TOMORROW [73]		INPLUSIS 1 [72]		INPLUSIS 2 [72]	
	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 2403 mg/day	Placebo	Pirfenidone 1800 mg/day	Placebo	NAC 600mg/tid	Placebo	Nintedanib 300mg	Placebo	Nintedanib 150mg/bid	Placebo	Nintedanib 150mg/bid	Placebo
Surgical Lung Biopsy (%)	86 (49)	85 (49)	86 (30.9)	79 (28.5)	94 (55)	94 (54)	26 (24.1)	28 (26.9)			29 (34.1)	19 (22.4)				

Note: Data only reported from arms included in NMA

** Data reported as time since IPF diagnosis, years ± SD (ERG assumes this is mean time)*

DLco, Diffusing capacity of the lungs for carbon monoxide; 6MWD, 6-Minute walking distance; 6MWT, 6-Minute walking distance; FVC, Forced vital capacity; HRCT, High-resolution computed tomography; IPF, Idiopathic pulmonary fibrosis; m, meter; PaO2, Partial pressure arterial oxygen; SD, Standard deviation; SpO2, Peripheral oxygen saturation; TLC, total lung capacity and VC, Vital capacity

Risk of bias

The methodological quality of the studies included in the NMA was assessed in the CS,[4] (page 128-129 and Appendix 10) using standard criteria adapted from the CRD guidance for undertaking systematic reviews.[44] A summary of the quality assessment results, as reported by the company, is provided in Table 37.

The CS[4] noted that a potential risk of bias arises from the different methods used for handling missing data across the studies and the process undertaken for randomisation was unclear in the TOMORROW[73] and the SP3 trial.[38] In the TOMORROW trial,[73] an interactive voice-response system (IVRS) was used to perform randomisation; however, no information was provided on how randomisation was generated. In the SP3[38] study, patients were allocated to treatment groups using a modified minimisation method, including some random allocation based on biased coin design to balance baseline SpO_2. However, for the purpose of these analyses it was assumed that randomisation process was adequate for all (CS,[4] page 128). The ERG agrees that the majority of the studies were of good quality, with low risk of bias, however, the ERG disagrees with categorising SP3[38] as a study with low risk of bias, principally because of the absence of any published protocols and the inadequacy of the information contained within the published manuscripts. Further details are provided in Section 4.2.

Table 37: Quality assessment summary of RCTs included for NMA (reproduced from CS,[4] page 129)

	Was randomisation carried out appropriately?	Was the concealment of treatment allocation adequate?	Groups similar at baseline in terms of prognostic factors?	Were the care providers, participants, and outcome assessors blind to treatment allocation?	Unexpected imbalances in drop-outs between groups?	Authors measured more outcomes than they reported?	Did the analysis include an ITT basis?	Risk of bias of the study
CAPACITY 1 & CAPACITY 2[49]	Yes	Yes	Yes	Yes	No	No	Yes	Low risk
SP3[38]	Unclear	Yes	Yes	Unclear	No	No	Yes	Low risk
ASCEND[34]	Yes	Yes	Yes	Yes	No	No	Yes	Low risk
PANTHER[74, 75]	Yes	Yes	Yes	Unclear	Yes (at interim analysis)	No	No	Some risk of bias
TOMORROW[73]	Unclear	Yes	Yes	Yes	No	No	No	Low risk
INPLUSIS 1 & INPULSIS 2[72]	Yes	Yes	Yes	Yes	No	No	No	Low risk

Scenarios considered

For the statistical analysis (see CS,[4] pages 126-129), the company used a base-case network which included all Phase II and III trials. Sensitivity analyses were performed using a restricted network which was limited to Phase III trials and excluded the triple therapy arm of the PANTHER trial.[77]

The ERG also asked the company to perform a sensitivity analysis without the SP3[38] and PANTHER studies[74, 75] (see clarification response,[10] question A38).[10] The company did not agree on the relevance of excluding SP3[38] from the network, stating in their response that *"SP3 has been recognised as providing valuable evidence in several reviews: the initial NICE technology appraisal of pirfenidone[2]; the nintedanib appraisal[78], and; as part of the EMA's review of the marketing authorisation application for pirfenidone."* The company did, however, provide results excluding the PANTHER study[74, 75] for one outcome (all-cause mortality up to 52 weeks) as proof of concept that excluding PANTHER[74, 75] does not change the comparative efficacy of pirfenidone, nintedanib and placebo. The ERG considers that the stated concerns relating to population difference, statistical methods for handling missing data, and risk of bias provide reason to consider excluding SP3[38] from the analyses and have consequently not included SP3[38] in the ERG base-case network. Table 38 summarises the studies included in the company's base-case network, and how this differs to the company's restricted network and ERG base-case network. Note that the inclusion of studies in the NMA analyses varies by outcome.

Table 38: Summary of the trials used in the network meta-analysis (reproduced from CS, page 124)

Trial (reference) included in CS base-case	CS restricted network?	ERG base-case network?	Treatments		
			Placebo	PFN	NTB
ASCEND (King 2014[34])	Yes	Yes	Yes	Yes	
CAPACITY 1 (Noble 2011[49])	Yes	Yes	Yes	Yes	
CAPACITY 2 (Noble 2011[49])	Yes	Yes	Yes	Yes	
SP3 (Taniguchi 2010[38])	Yes		Yes	Yes	
INPULSIS-1 (Richeldi 2014[72])	Yes	Yes	Yes		Yes
INPULSIS-2 (Richeldi 2014[72])	Yes	Yes	Yes		Yes
TOMORROW (Richeldi 2011[73])		Yes	Yes		Yes
PANTHER NAC (Martinez 2014[74])			Yes		
PANTHER Triple therapy (Raghu 2012[77])	Yes		Yes		

4.7 Critique of the NMA

4.7.1 Efficacy

Summary of analyses undertaken

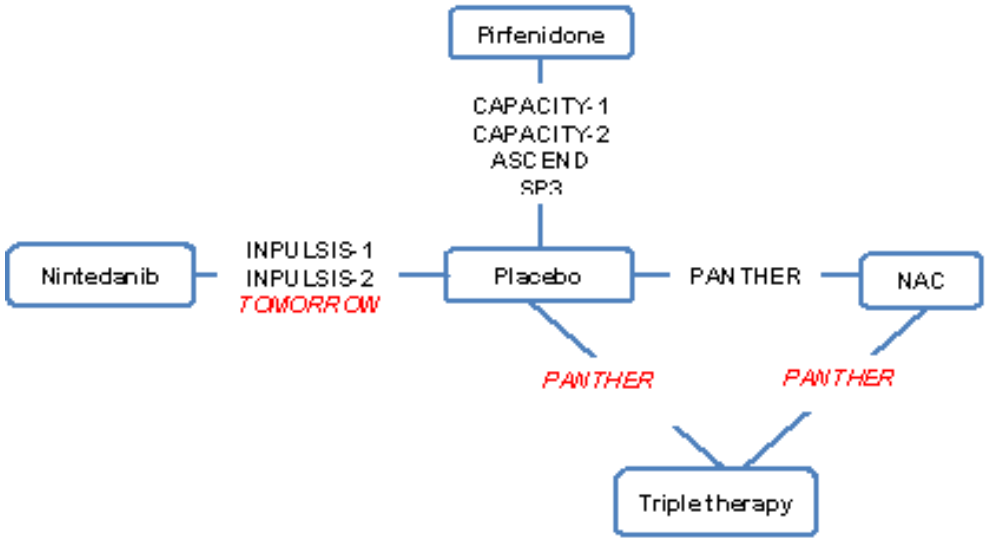
NMA were performed by the company to compare the treatment effects of pirfenidone, nintedanib, NAC, triple therapy and placebo for 11 outcomes relevant to the decision problem, as listed in Table 35. The results of four of these outcomes (OS, PFS, time to treatment discontinuation, acute exacerbations) are used to inform the economic model. Separate NMAs were undertaken for each outcome.

The base-case NMAs included all Phase II and III trials (eight trials in total). The network diagram for these studies is presented in Figure 26, however not all trials reported data that could contribute to all NMA outcomes. Table 39 summarises data available in each trial, for each outcome. A full summary of the NMA results and the number of studies included by scenario is provided in Table 41.

The company also performed sensitivity analyses using a restricted network which was limited to Phase III trials and excluded the triple therapy arm of the PANTHER trial. A full summary of the NMA results for the restricted network is provided in CS appendix 14.

Additional analyses performed by the ERG are summarised in Section 4.8.

Figure 26: Network diagram including all trials for NMA (reproduced from CS, Figure 19 page 125)



All trials are included in the base-case network

Red italicised text indicates trials excluded from the restricted network

Table 39: Summary of evidence for the company's base-case NMAs (adapted from clarification response,[10] question A32, Table 13)

	Study duration (weeks)	All-cause mortality	IPF-related mortality	Progression-Free Survival	Exacerbations	10% categorical decline FVC% pred [d]	FVC % pred	FVC Litres	6MWD	SGRQ	UCSD SOBQ	All cause discontinuation
CAPACITY1 and 2	72	52 and 72	52 and 72	52 and 72	52	48	48	48	48	48	48	48
CAPACITY1 and 2	72	52 and 72	52 and 72	52 and 72	52	48	48	48	48	48	48	48
ASCEND	52	52	52	52	52	52	52	52	52	-	52	52
SP3	52	52[a]	-	52	52	[e]	52	52	-	-	-	-
INPULSIS1	52	52	52*	52*	52	52	52	52	-	52	-	52
INPULSIS 2	52	52			52	52	52	52	-	52	-	52
TOMORROW	52	52[a]	52[b]	[f]	52	[g]	52	52	52	52	-	52
PANTHER (NAC)	60	60[b]	60[a]	60[a]	60	60	60	60	60	60	60	60
PANTHER (Triple)	32	32 (mean)	32 (mean)[a]	60[a]	[h]	[h]	-	60[c]	60[c]	60[c]	60[c]	-

[a] *HRs were unavailable: number of events and number of patients were used as an alternative (via the Woods model)*

[b] *HR was calculated from other available data using the methods of Parmar*

[c] *For FVC (L), 6MWD, SGRQ and UCSD SOBQ, publication presented estimated changes over 60 weeks (based on a repeated measures model)*

[d] *The NMA for this outcome assumes that all patients with missing values are non-responders (i.e. have a decline of more than 10%)*

[e] *Taniguchi 2010 reported some results for FVC 10% (Table E2 in the supplementary appendix) however there was insufficient information to calculate FVC 10% in line with the above definition*

[f] *only reported the proportion of patients who progressed, rather than the proportion of patients who either progressed or died. Although the number of deaths was reported, it was unclear how many patients progressed before they died and therefore PFS cannot be calculated.*

[g] *The outcome is not clearly defined in the nintedanib company submission to NICE.[26] Based on the company's response to clarification questions, the submission may be measuring any decline up to 52 weeks, whereas the other studies are measuring declines at exactly 48/52 weeks.*

[h] *Results were reported but the time point was not comparable.*

** pooled HR used for INPULSIS trials. Post hoc analysis only supplied as part of the submission to NICE by Boehringer Ingelheim.[26]*

Methods for the NMA

The CS specified the use of a random effects model for the principal analysis and also performed sensitivity analyses using fixed effects models (results provided in CS Appendix 14). Model fit statistics (total residual deviance and deviance information criterion [DIC]) were not provided in the original CS, but were provided for key outcomes upon clarification. The DIC provides a relative measure of goodness-of-fit that penalises complexity and was used to compare different models for the same likelihood and data[79]. The company reported that no meaningful differences in DIC between random effects and fixed effect models were observed. Random effects models were considered more appropriate due to the stated concerns in heterogeneity between the studies and the ERG considers that this decision was appropriate.

Where there were sufficient sample data, conventional reference prior distributions were used, however for certain endpoints there were too few studies to estimate the between study variance from the sample data alone and weakly informative priors were used. Although prior distributions should not be used without reasonable justification, the company considered *"the assumption of no heterogeneity made in the fixed effect model to be unrealistic."* (CS, [4] page 131). In the absence of further information on which to base the choice of prior, these were based on the recommendations of Turner *et al.*,[80], with details provided in CS,[4] Appendix 12. The ERG considers the company's choice of model and priors to be appropriate.

Statistical heterogeneity was assessed by presenting I^2 statistics from pairwise comparisons. Estimates of between study standard deviation from the conducted NMA were not reported in the original CS however the company provided this information upon response to clarification question A34, for key outcomes informing the economic model: all-cause mortality at 52 weeks; all-cause mortality at 72 weeks; PFS at 52 weeks; PFS at 72 weeks; IPF-related mortality at 52 weeks; IPF-related mortality at 72 weeks, and acute exacerbations.

Despite describing PANTHER as a multi-arm trial, it was treated as two separate placebo controlled trials for the statistical analyses. This was justified by describing PANTHER as an "atypical multi-arm trial", in which the "correlations between the arms will be less than those in a regular multi-arm trial" (see CS, Appendix 12). The ERG does not believe the issue is of importance, given that the interventions considered in PANTHER are not of relevance to the decision problem, but notes that if it is to be included, appropriate methods including correction for multi-arm trials should be used.

Reporting of results for the NMA

For continuous outcomes (FVC, 6MWD, SGRQ, UCSD SOBQ) the mean difference in the change from baseline is reported; for binary outcomes (acute exacerbations, discontinuation, categorical decline in

FVC) ORs are reported, and for survival outcomes (all-cause mortality, PFS, IPF related mortality) HRs are reported.

Results were summarised using posterior medians and 95% credible intervals (CrI). In the presence of heterogeneity, it is recommend that the predictive distribution, rather than the distribution of the mean treatment effect, better represents uncertainty about comparative effectiveness for a future rollout of a particular intervention.[81] The 95% predictive intervals (PrI) for key outcomes were provided by the company following a request for clarification (see clarification response,[10] question A34). The predictive intervals from the ERG analyses reported in Section 4.8 are used to inform the ERG base-case model in Section 5.

Implementation

Analyses were conducted using JAGS version 3.3.0 [82] and R version 3.0.1 or above.[83] The 'R2JAGS'[84] package was used to run JAGS from within R. The company stated that *"an appropriate burn-in period and number of iterations were allowed for."* (CS, Appendix 12).

Main results of NMA

Input data for the company's base-case network is provided for all outcomes in CS Appendix 11. In the original submission, pooled results for the two INPULSIS studies were used to inform the NMA for all survival outcomes (all-cause mortality, PFS, IPF related mortality). In response to clarification question A35 the company provided results using the individual study HR for all-cause mortality. The updated data used for the all-cause mortality NMA are presented in Table 40.

A full summary of the NMA results from the company's base-case network is provided in Table 39. The treatment effects for pirfenidone are broadly similar to those for nintedanib for all outcomes, with the pairwise treatment effects indicating that neither treatment is statistically significantly more effective.

For change from baseline in absolute (litres) and percent predicted FVC/VC, both pirfenidone and nintedanib were associated with beneficial effects relative to placebo. Pirfenidone was also associated with beneficial effects relative to placebo for all three time to event outcomes (all-cause mortality, PFS and IPF-related mortality). For nintedanib, the direction of the treatment effect favoured the active treatment, however the results were not statistically significant relative to placebo. For acute exacerbations, the treatment effects were not statistically significant for either treatment. For all-cause discontinuation of treatment, nintedanib was associated with increased odds of all-cause discontinuation relative to placebo, however the treatment effect was not statistically significant for pirfenidone.

The heterogeneity in treatment effects between studies is summarised for key outcomes in Table 41. The estimate of between-study standard deviation is mild-moderate for all outcomes, but with considerable uncertainty for IPF-related mortality and acute exacerbations. The network for IPF-related mortality contains fewer studies than that for all-cause mortality with no outcome data provided by SP3, and only pooled results were available for the INPULSIS trials. The NMA for acute exacerbations utilised a weakly informative prior for the between-study heterogeneity, as described above.

Sensitivity analyses conducted using fixed effects models were consistent with those reported from the random effects models. Results were also consistent across the company's base-case and restricted network.

For all-cause mortality, PFS and IPF-related mortality, the company's principal analyses use data from CAPACITY 1 and 2 evaluated at 52 weeks, rather than the full trial duration of 72 weeks. The main rationale behind this choice was to provide a comparison of data across similar timeframes for all studies. Other factors discussed by the company to justify this decision are that full follow up data were available for the majority of patients at this time point, that clinical data from ASCEND and CAPACITY was pre-specified to be pooled at 52 weeks, and that are no data available to support an assumption of proportional hazards in the longer term for nintedanib versus placebo (see clarification response,[10] question A37). The ERG considers that the use of the 52 week data would be appropriate if the purpose of analysis was to estimate the treatment effects at the specified time point, and there was reason to believe that treatment effects may not be consistent over the extended follow up period (and therefore bias results). However, the purpose of the analysis is to estimate the population mean survival time, and for the cost effectiveness modelling it was considered appropriate by the company to extrapolate the treatment effects over the full lifetime. The ERG therefore considers that the full evidence base with 72 week follow up should be used. Consequently, the 72 week data have been used in the additional ERG analyses presented in Section 4.8 and in health economic model .The ERG notes that the use of a constant HR in the economic model is appropriate only if the assumption of proportional hazards can be justified over both the observed and unobserved time period. The company's observation that there are no data available to support an assumption of proportional hazards in the longer term for nintedanib versus placebo therefore raises concerns over the reliability of the results based on extrapolated HR.

In response to clarification question A36 the company performed additional NMAs to justify the assumption that treatment effects are constant over time by including a covariate for trial duration through meta-regression, as described in the NICE TSD [81]. Analyses were conducted for the three time-to-event outcomes (all-cause mortality, IPF related mortality, PFS) only. Results of the company's meta-regressions (see clarification response addendum,[30] question A36, pages 17 - 21) showed that

including a covariate for study duration did not improve model fit, as judged using the DIC, and resulted in higher estimates for the between trial standard deviation. However these analyses were limited by the small number of studies and effect estimates at different trial durations were available only for the pirfenidone studies. The results should therefore be interpreted with caution and not viewed as robust evidence for a lack of treatment by time interaction.

Table 40: Input data for all-cause mortality NMA, company's base-case network (adapted from clarification response,[10] question A35 Table 29)

Study	Treatment	Comparator	HR	logHR	SE	N	n
CAPACITY 2	Pirfenidone	Placebo	0.37	-0.9942523	0.5304795	NA	NA
CAPACITY 1	Pirfenidone	Placebo	0.66	-0.4155154	0.5237481	NA	NA
ASCEND	Pirfenidone	Placebo	0.55	-0.597837	0.3793018	NA	NA
INPULSIS 1	Nintedanib	Placebo	0.63	-0.4620355	0.3942242	NA	NA
INPULSIS 2	Nintedanib	Placebo	0.74	-0.3011051	0.3103049	NA	NA
PANTHER	Triple therapy	Placebo	9.26	2.225704	1.0604775	NA	NA
PANTHER	NAC	Placebo	1.995622	0.6909556	0.6666667	NA	NA
SP3	Pirfenidone	Placebo	NA	NA	NA	110	3
SP3	Placebo	Placebo	NA	NA	NA	109	6
TOMORROW	Nintedanib	Placebo	NA	NA	NA	85	7
TOMORROW	Placebo	Placebo	NA	NA	NA	85	9

Table 41: Summary of results from company's base-case NMA, random effects model (adapted from CS Section 4.10 and clarification response,[10] question A34)

Outcome	Base-case network, RE model					
	Number of trials*		Treatment effect			Between study heterogeneity
	PFN	NTB	PFN vs placebo	NTB vs placebo	PFN vs NTB	
Lung Capacity		-				
Change from baseline in Percent Predicted FVC/VC (%)	4	3	3.39 (1.94,4.84)	3.33(2.34,4.5)	0.05 (-0.81,1.80)	NR
Change from baseline in FVC/VC (L)	4	3	0.12 (0.04,0.20)	0.12 (0.04,0.21)	0.00 (-0.11,0.12)	NR
FVC decline ≥10% Percent Predicted (OR)	3	2	0.58 (0.40,0.88)	0.65(0.42,1.02)	1.12(0.60,2.01)	NR
Physical Functioning and HRQoL		-				
Change in 6MWD	3	1	22.70 (8.82,36.31)	6.00 (-28.25,40.66)	16.63 (-20.83,53.81)	NR
SGRQ	2	3	-1.24(-4.94,2.39)	-2.11 (-5.48,0.37)	0.88 (-3.45,5.94)	NR
UCSD SOBQ	3	0	-3.19 (-6.24, -0.17)	NA	NA	NR
Time to event outcomes		-				
All-Cause Mortality up to 52 wks (HR)	4	3	0.52 (0.30, 0.88)	0.71 (0.43,1.16)	0.73 (0.35,1.50)	0.11 (0.03,0.54)
All-Cause Mortality up to 72 wks (HR)			0.62 (0.38, 0.99)	0.71 (0.43, 1.16)	0.87 (0.44, 1.72)	0.11(0.03,0.53)
PFS HR up to 52 wks (HR)	4	2**	0.63 (0.50, 0.80)	0.74(0.51,1.08)	0.85 (0.55,1.34)	0.09 (0.02,0.45)
PFS HR up to 72 wks (HR)			0.63 (0.50, 0.78)	0.74 (0.51,1.07)	0.85(0.55,1.31)	0.09 (0.02, 0.43)
IPF-Related Mortality up to 52 wks (HR)	3	3**	0.36(0.14, 0.90)	0.60 (0.22,1.33)	0.61 (0.18,2.34)	0.19 (0.03,1.44)
IPF-Related Mortality up to 72 wks (HR)			0.48 (0.22, 1.01)	0.60 (0.23, 1.28)	0.80 (0.27,2.63)	0.18 (0.03,1.29)
Other		-				
Acute Exacerbations (OR)	4	3	0.62 (0.29,1.39)	0.55 (0.26,1.09)	1.14 (0.41,3.44)	0.29(0.04,1.07)
All-cause Discontinuation of Treatment (OR)	4	3	1.28 (0.91,1.78)	1.42 (1.01,2.01)	0.90 (0.55,1.44)	NR

PFN, pirfenidone; NTB, nintedanib; HR, hazard ratio; OR, odds ratio

* number of trials are summarised for interventions relevant to the decision problem only. Network also includes NAC and triple therapy trials (PANTHER)

** uses pooled HR for INPULSIS

4.7.2 Safety

In response to clarification question A39, the company performed additional NMA to compare the treatment effects of pirfenidone, nintedanib, NAC, triple therapy and placebo for four key adverse events outcomes; diarrhoea, rash, discontinuation due to adverse event and serious cardiac adverse events. The results of these NMA were used to inform the updated economic model.

The base-case NMAs included all Phase II and III trials (eight trials in total), however not all trials reported data that could contribute to all AE outcomes. Table 42 summarises data available in each trial, for each AE outcome. As with the NMA of efficacy outcomes, the company also performed sensitivity analyses using a restricted network. A random effects model was specified for the principal analysis and sensitivity analyses were performed using a fixed effects model. Weakly informative priors, based on the recommendations of Turner *et al.*,[80] were used for the between study variance.

As with the data for the NMA of efficacy outcomes, there were differences in follow up time between studies. Data for the CAPACITY trials was collected at 72 weeks, rather than using intermediate follow up data (as was done for the NMA of efficacy outcomes) providing a greater range of follow up times. The CS page 126 states *"It is difficult to justify whether treatment effects will be stable over this longer time period"* and acknowledge that the difference in follow up time may lead to bias in the results.

A full summary of the NMA results and the number of studies included by scenario is provided in Table 43 for the company's base-case network, random effects model. Additional analyses are presented in the clarification response appendix D.[10]

Pirfenidone was 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.

Table 42: Summary of evidence for the company's base-case adverse event NMAs (adapted clarification response,[10] appendix D, Table 137)

Trials	Study duration (weeks)	Diarrhoea	Rash	Discontinuation of treatment due to AE	Serious cardiac events
CAPACITY1	72	√	√	√*	√*
CAPACITY 2	72	√	√		
ASCEND	52	√	√	√	√
SP3	52	-	-	√	-
PANTHER (NAC)	60	√	-	√	√
PANTHER (Triple)	32	√	√	-	√
INPULSIS1	52	√	√	√	√*
INPULSIS2	52	√	-	√	
TOMORROW	52	√	√	√	√

** Pooled trials*

Table 43: Summary of results from the company's base-case AE NMAs, random effects model (adapted clarification response,[10] appendix D, Table 137)

Outcome	Number of trials*		Base-case network, RE model Treatment effect; OR (95% CrI)		
	PFN	NTB	PFN vs placebo	NTB vs placebo	PFN vs NTB
Diarrhoea	3	3	1.39 (0.94, 2.11)	7.32 (4.82, 11.13)	0.19 (0.11, 0.35)
Rash	3	2	3.85 (2.38, 6.29)	1.29 (0.49, 3.35)	2.99 (1.03, 8.88)
Discontinuation due to adverse event	4**	3	1.58 (1.04, 2.39)	1.52 (1.01, 2.29)	1.04 (0.58, 1.85)
Serious cardiac events	3**	3**	1.36 (0.54, 3.46)	0.64 (0.17, 1.49)	2.11 (0.65, 11.34)

PFN, pirfenidone; NTB, nintedanib; HR, hazard ratio; OR, odds ratio

** number of trials are summarised for interventions relevant to the decision problem only. network also includes NAC and triple therapy trials (PANTHER)*

*** uses pooled HR*

4.8 Additional work on clinical effectiveness undertaken by the ERG

4.8.1 Network meta-analysis

Additional analyses were conducted by the ERG, using the ERG base-case network described in Table 38. NMAs were conducted using random effects models for the following key outcomes used to inform the company's health economic model: all-cause mortality up to 72 weeks; PFS up to 72 weeks and acute exacerbations.

Analyses were conducted in the freely available software package WinBUGS[85] and R,[83] using the R2Winbugs[86] interface package. For all-cause mortality, there was evidence of poor convergence and so a weakly informative half-normal prior with variance 0.32^2 was used. WinBUGS code using this prior was provided by the company (see CS, Appendix 15). Under this prior, the between-study SD has a mean of 0.26. For all outcomes, a burn-in of 300,000 iterations of the Markov chain was used with a further 100,000 iterations retained to estimate parameters. Samples from the posterior distributions exhibited moderate correlation between successive iterations of the Markov chain and so were thinned by retaining every 10^{th} sample.

All-cause mortality

Six trials were included in the network for all-cause mortality (Figure 27). The treatment effects are summarised in Figure 28.

Figure 27: Network of evidence for all-cause mortality, acute exacerbations and all-cause discontinuation, ERG base-case

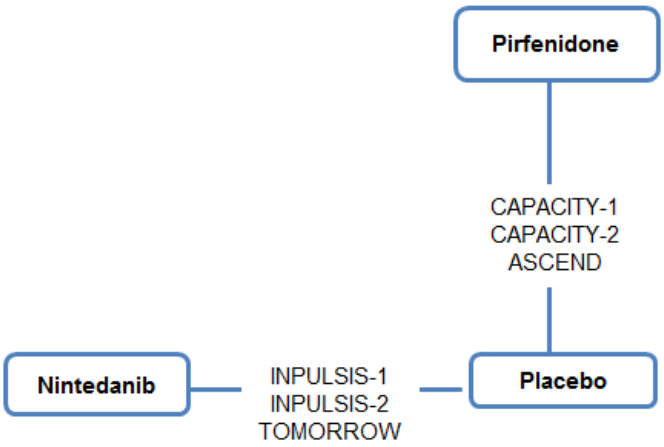
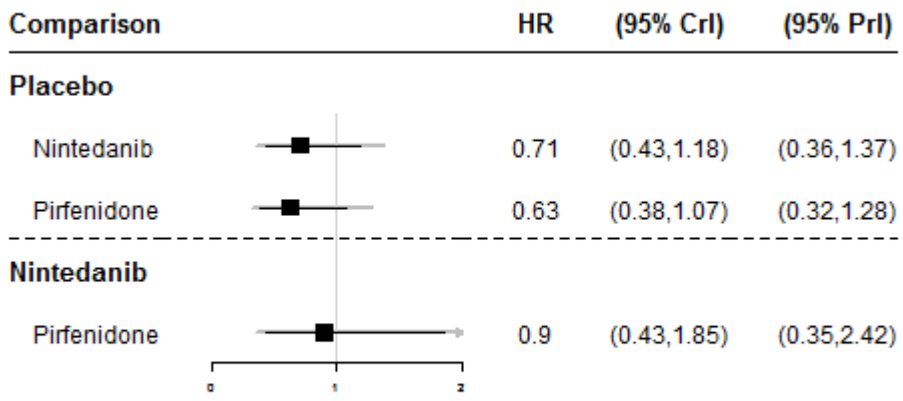


Figure 28: All-cause mortality, ERG base-case network - HR, 95% CrI and 95% PrI



Heterogeneity: between-study variance is 0.14 (95% CrI; 0.01, 0.52)

PFS

Five trials were included in the network for PFS (Figure 29), but a pooled HR was used for the INPULSIS trials since the individual study-level treatment effects were not available. The results of the NMA are summarised in

Figure 30.

Figure 29: Network of evidence for PFS, ERG base-case

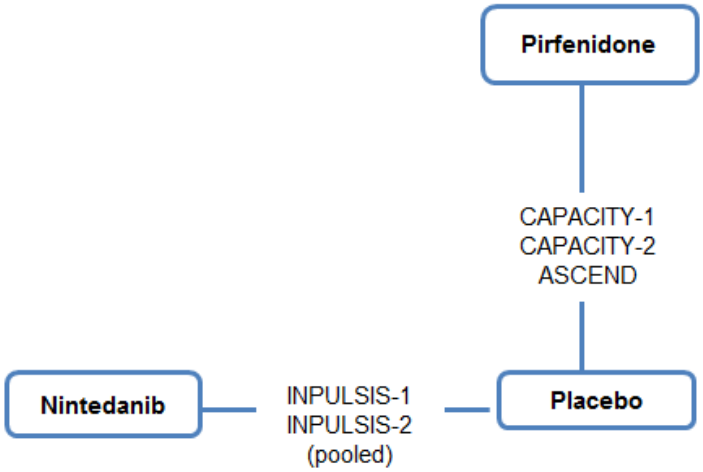
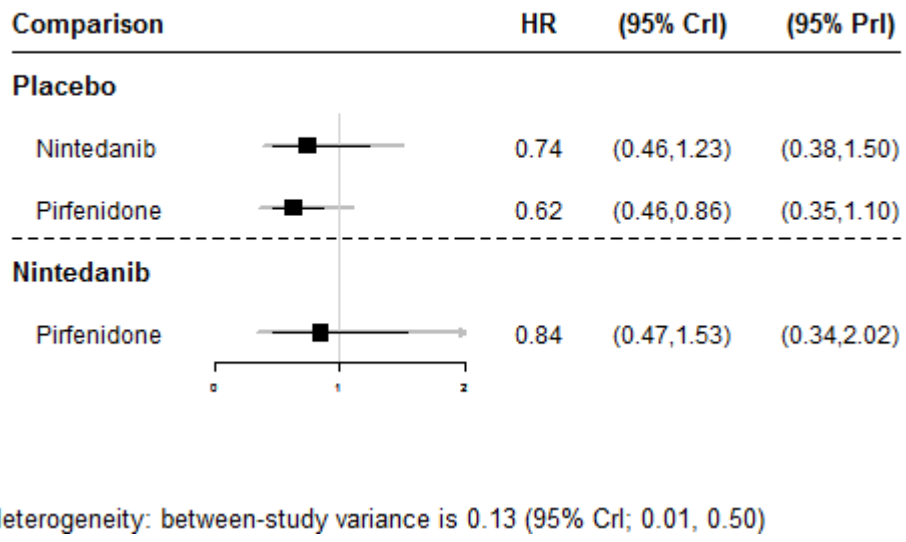


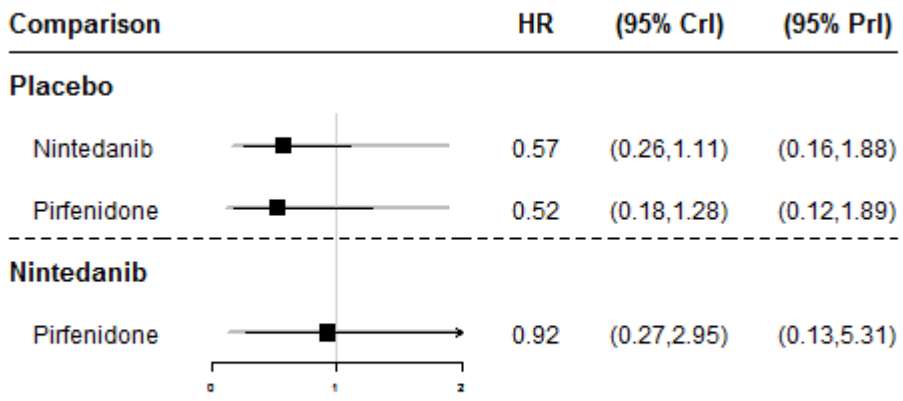
Figure 30: PFS, ERG base-case network - HR, 95% CrI and 95% PrI



Exacerbations

Six trials were included in the network for acute exacerbations (Figure 27). The pooled treatment effects are summarised in Figure 31.

Figure 31: Acute exacerbations, ERG base-case network - HR, 95% CrI and 95% PrI



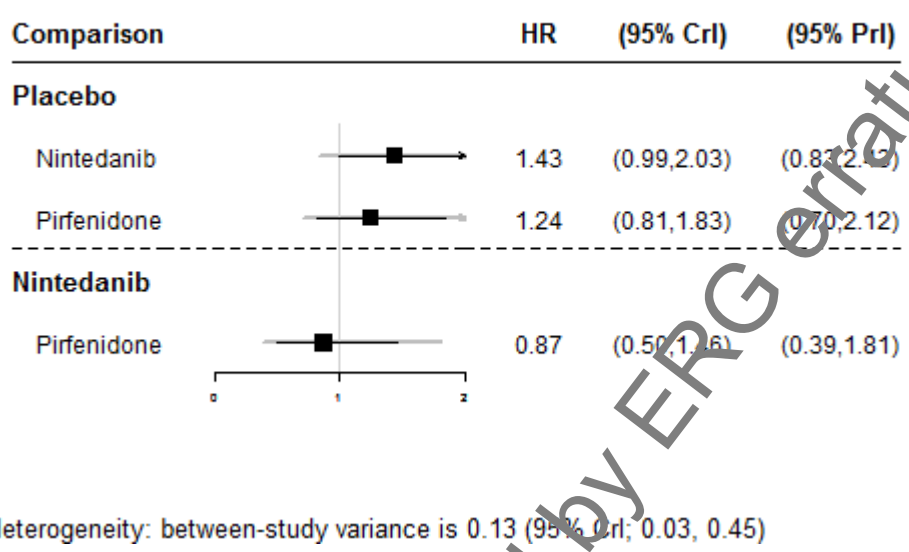
Heterogeneity: between-study variance is 0.29 (95% CrI; 0.05, 1.16)

All cause discontinuation

Six trials were included in the network for acute exacerbations (Figure 27). The pooled treatment effects (odds ratios) are summarised in Figure 32.

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 32: 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, unlicensed doses of pirfenidone, 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 but was not statistically significant (absolute difference: 3.8 [95% CI: -2.7 to 10.2], $p=0.440$), whilst CAPACITY 2[49] did report a statistically significant difference in favour of pirfenidone compared with placebo at week 72 (absolute difference: 14.4 [95% CI: 7.4 to 21.3], $p=0.001$). ASCEND also reported a significantly higher proportion of patients with no decline in percent predicted FVC (22.7% for pirfenidone versus 9.7% for placebo, $p<0.000001$), but CAPACITY 1[49] reported no difference between pirfenidone and placebo on this outcome measure (25.8% versus 22%, *p*-value not reported). CAPACITY 2[49] reported a higher proportion of patients with no decline in percent predicted FVC for

pirfenidone compared with placebo (24.1% versus 13.8%), but did not report a *p*-value. A meta-analysis of the ASCEND trial (52 weeks)[34] and the CAPACITY trials (48 weeks) [49] suggested that, compared with placebo, pirfenidone lowers the proportion of patients experiencing decline in FVC percent predicted of ≥10% (OR: 0.50, 95% CI: 0.31 to 0.82, *p*-value not reported).

There were fewer overall deaths or treatment-emergent IPF-related deaths in the pirfenidone than the placebo arms of the ASCEND[34] and CAPACITY[49] trials. However, these differences were not statistically significant in the ASCEND trial[34] at 52 weeks (for all-cause mortality or treatment-emergent IPF-related deaths, *p*=0.105 and *p*=0.226, respectively). The differences were significant in the pooled analyses for the CAPACITY trials[49] at 52 weeks (for all-cause mortality or treatment-emergent IPF-related deaths, *p*=0.047 and *p*=0.012, respectively) and in the pooled ASCEND[34] and CAPACITY[49] trials at 52 weeks (for all-cause mortality or treatment-emergent IPF-related deaths, *p*=0.011 and *p*=0.006, respectively). However, these differences were no longer significant at 72 weeks in the pooled CAPACITY trials (for all-cause mortality, *p*=0.315, IPF related mortality, *p*=0.117, or treatment-emergent all-cause mortality, *p*=0.141). There was only a significant difference between groups for treatment-emergent IPF-related mortality in the pooled CAPACITY trials[49] at 72 weeks (*p*=0.03). There appears to be a markedly increased rate of mortality in the CAPACITY trials[49] between the data reported for 52 weeks and for 72 weeks, the reasons for which are unclear. SP3, SP2 and Huang *et al.* (2015)[48] all reported all-cause mortality and found no statistically significant differences between pirfenidone and placebo arms. Meta-analysis of CAPACITY 1 & 2[49] and ASCEND[34] for pirfenidone compared with placebo, at 52 weeks, suggests that pirfenidone reduces all-cause mortality (HR: 0.52, 95% CI: 0.31 to 0.88, *p*-value not reported) and IPF-related mortality (HR: 0.37, 95% CI: 0.18 to 0.76, *p*-value not reported). Sensitivity analysis of the three trials at 72 weeks gave similar outcomes in favour of pirfenidone for both all cause-cause mortality and IPF- related mortality (HR: 0.64, 95% CI: 0.41 to 0.99, *p*-value not reported) and (HR: 0.49, CI: 0.27 to 0.87, *p*-value not reported). However, the reduction in mortality was lower at 72 weeks compared with 52 weeks.

Four of the key trials reported data for PFS: ASCEND, [34] CAPACITY 1 & 2[49] and SP3.[38] The definitions of PFS varied across the trials, but with a common element of a confirmed ≥10% decline from baseline in percent predicted FVC or VC. As with the findings for FVC outcomes, ASCEND at 52 weeks (HR 0.57; 95% CI, 0.43 to 0.77, *p*=0.0001) and CAPACITY 2[49] at 72 weeks (HR 0.64; 95% CI, 0.44 to 0.95, *p*=0.023) found statistically significant benefits in terms of PFS for those on pirfenidone compared with those on placebo; whilst for CAPACITY 1[49] the treatment effect was not statistically significant (HR: 0.84; 95% CI, 0.58 to 1.22, *p*=0.355). *Post hoc* pooled analyses of the ASCEND[34] and CAPACITY trials, [49] found statistically significant benefits for those on pirfenidone compared with those on placebo (HR: 0.62; 95% CI: 0.51 to 0.76; *p*<0.0001). Huang *et al.* (2015)[48] also reported a significant treatment benefit for pirfenidone plus NAC compared with placebo plus NAC for

PFS (HR=1.88, 95% CI: 1.09 to 3.24, *p*=0.02). Meta-analysis of the four trials showed pirfenidone improves PFS at 52 weeks compared with placebo (HR 0.63 95% CI, 0.53 to 0.74, *p*-value not reported). A sensitivity analysis based on CAPACITY trials[49] at 72 weeks, and ASCEND at 52 weeks,[34] with the assumption that the proportional hazards assumption holds up to 72 weeks, gave the same results.

All five included trials reported outcome data on acute exacerbations but used different definitions. The rates of acute exacerbation were much higher in the ASCEND trial[34] than in the CAPACITY trials, [49] with higher incidence in the placebo than the pirfenidone arms in the ASCEND[34] and CAPACITY 2[49] trials: no p values were reported. None of these three trials showed any statistically significant treatment effects compared to placebo for this outcome measure. SP2 t[39] did find a statistically significant difference in favour of the 1,800mg per day dose of pirfenidone for this outcome, but there was no consistency in the frequency of acute exacerbation reported across trials. This might be explained by the different definitions used. A meta-analysis of ASCEND,[34] CAPACITY 1 & 2[49] and SP3[38] also showed that pirfenidone is associated with a reduced risk of acute exacerbation of IPF with a HR of 0.64 (95% CI: 0.38 to 1.06, *p*-value not reported) compared with placebo, however the treatment effect was not statistically significant for the random effects model. CAPACITY 1 & 2[49] and SP2[39] also reported similarities in rates of hospitalisation (due to respiratory or non-respiratory causes) between pirfenidone and placebo arms.

Patient-reported outcomes were evaluated using the UCSD SOBQ and the SGRQ in the ASCEND and CAPACITY[49] trials. The treatment effects were not statistically significant for any of the individual trials , however results of the meta-analysis (using data from the CAPACITY trials at 48 weeks) suggest that pirfenidone is associated with a statistically significant reduction in in USCD SOBQ compared with placebo (Mean difference: -3.19 (95% CI: -5.74 to -0.63, *p*-value not reported).

The CS[4] reported the findings from two sets of analyses for 6MWD. The ASCEND[34] and CAPACITY[49] trials all reported findings on the pre-specified outcome of mean change from baseline in 6MWD for pirfenidone 2,403mg per day compared with placebo. ASCEND[34] at 52 weeks (absolute difference: 26.7m; relative reduction: 44.2%; *p*=0.036) and CAPACITY 1[49] at 72 weeks (absolute difference: 31.8m; relative difference: not reported; *p*<0.001) both reported a statistically significant and clinically important difference between pirfenidone and placebo on this outcome, but the treatment effect in CAPACITY 2[49] was not statistically significant (absolute difference: 16.4m; relative difference: not reported; *p*=0.171). A pooled analysis of the CAPACITY trials[49] at 72 weeks (absolute difference: 24m; relative difference: 31.2%; *p*=0.0009) also reported a statistically significant and clinically important difference between pirfenidone and placebo on this outcome. Huang *et al.* (2015)[48] reported no difference between the pirfenidone and placebo arms in the 6MWT. Meta-analysis of CAPACITY 1 &

2 (data from week 48) [49] and ASCEND (data from week 52)[34] suggested that pirfenidone reduces the decline in 6MWD (MD: 22.9, 95% CI 10.58 to 35.23, *p*-value not reported).

A *post hoc* categorical analysis based on a mean decline ≥50 m in 6MWD from baseline, or death, in ASCEND[34] and CAPACITY 1 & 2,[49] also found that there was a statistically significant difference between pirfenidone and placebo in ASCEND trial (52 weeks: absolute difference: 9.8%; relative reduction: 27.5%; p=0.04)[34] The treatment effect was not statistically significant in CAPACITY 1 (p=0.10),[49] but was statistically significant for CAPACITY 2 (p=0.049).[49] A pooled analysis of the CAPACITY trials (72 weeks: absolute difference: 12.2%; relative reduction: 26%; p=0.001)[49] also reported a statistically significant effect for pirfenidone compared with placebo for this categorical outcome.

Four trials (CAPACITY 1 & 2,[49] SP3,[38] SP2[39]) reported data on the change from baseline in DLco. The CAPACITY trials[49] reported the change in percent predicted DLco, while SP2[39] and SP3[38] reported the mean decline (mL/min/mmHG). None of the trials reported statistically significant treatment effect for this outcome measure.

It is unclear why CAPACITY 1[49] reports different findings from ASCEND[34] and CAPACITY 2[49] in terms of FVC,PFS and 6MWD. For CAPACITY 1[49] the treatment effect is not statistically significant for FVC or PFS outcomes, unlike ASCEND[34] and CAPACITY 2,[49] but reports a positive statistically significant effect on one measure of 6MWD, which is not found to be statistically significant in CAPACITY 2[49]. An additional, small RCT of pirfenidone in combination with NAC in adults with mild and moderate IPF was identified by the ERG[48] and also reported no statistically significant effect on FVC, 6MWD or mortality outcomes.

The effect of the, "intrinsic variability in rates of FVC decline"[49], acknowledged as an issue in the CAPACITY trials' publication, and expanded on by the company in response to a request for clarification of this issue by the ERG (see clarification response,[10] question A26), might explain differences in outcomes across trials. Clinical advice received by the ERG suggested that there is currently no accepted single criterion by which to identify speed of progression of IPF. Participants in the trials included in the CS were not stratified by rate of progression, so it is possible, for example, that the placebo arm might have had more participants with more rapidly progressing disease than the intervention arm. As a result, the true treatment effect of the intervention relative to placebo might be uncertain. This could work either for or against the intervention.

In response to a clarification request from the ERG (see clarification response,[10] question A31), the company also provided results on OS and PFS from the ASCEND[34] and CAPACITY[49] trials for groups with a baseline percent predicted FVC of ≤80% (moderate IPF) and >80% (mild IPF), although exact,

numbers within each subgroup in each trial arm were not reported. The findings did not suggest differential treatment effects according to disease severity for either outcome (as judged by the reported HR and 95% CI), however a treatment-by-subgroup interaction test was not reported so it is unclear if the difference between these subgroups was statistically significant.

The CS[4] also reported findings from non-randomised and non-controlled studies. First, the RECAP study (PIPF-012),[40] a non-randomised, non-controlled, open-label extension of the ASCEND and CAPACITY trials, which was principally designed to assess the long-term safety of pirfenidone 2,403mg/day in patients with IPF who received ≥80% of scheduled doses and completed the week 72 final study visit in CAPACITY 1 or CAPACITY 2. The RECAP study is ongoing. The most recent data-cut was performed in June 2015 and the next data-cut is planned in June 2016. The publication by Kreuter *et al*[63] found that discontinuation rates were highest in those enrolled patients who had originally received placebo, and especially in those who did not meet the ASCEND or CAPACITY entry criteria. Survival data and time-on-treatment data were reported in the CS,[4] (pages 159-161) and were presented for patients who received pirfenidone 2,403mg per day from baseline onwards in CAPACITY and ASCEND, and through the RECAP extension period, for whom data are available through to 8.8 years. Information on survival of patients with IPF was also presented from six registries to explore the relative survival rates of trial patients receiving pirfenidone compared with these "matched" real-world patients receiving best supportive care. The CS[4] stated that results were similar to the comparisons reported for the trials.

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. For change from baseline in absolute (litres) and percent predicted FVC/VC, both pirfenidone and nintedanib were associated with beneficial effects compared with placebo. Pirfenidone was also associated with beneficial effects relative to placebo for all three time-to-event outcomes (all-cause mortality, PFS and IPF-related mortality). For nintedanib, the direction of the treatment effect favoured the active treatment, however the results were not statistically significant relative to placebo. For acute exacerbations, the treatment effects were not statistically significant for either treatment. For all-cause discontinuation of treatment, nintedanib was associated with beneficial effects relative to placebo; however the treatment effect was not statistically significant for pirfenidone.

The ERG noted that, overall, some adverse events (AEs) were frequent, especially nausea, rash, dizziness, dyspepsia, anorexia and photosensitivity, but that these were generally mild or moderate in severity. The ERG requested from the company more detailed data on serious adverse events and the adverse events leading to discontinuation. The most frequently-reported serious adverse events in the pirfenidone arms of the ASCEND[34] and CAPACITY[49] trials, other than worsening of IPF, were

pneumonia, prostate cancer, angina pectoris, coronary artery disease, congestive cardiac failure, atrial fibrillation and pneumothorax. The AEs leading to discontinuation of treatment in ≥1% of patients in pirfenidone groups were pneumonia, rash, raised hepatic enzyme levels and decreased weight (in ASCEND), [34] photosensitivity, rash and respiratory failure (in CAPACITY 1)[49] and bladder cancer, nausea and rash (in CAPACITY 2). [49] The majority of safety data were from trials with a follow-up of no more than 72 weeks, but the CS[4] did present analyses that included more than 300 patients who had received pirfenidone for more than four years. However, the results for these patients were not presented separately. The ERG noted that the two ongoing studies to evaluate safety would address some outstanding issues: the non-randomised, non-controlled, OLE study that included a set of patients who completed either ASCEND, CAPACITY 1 or 2 (RECAP) and PIPF-002,[65] an ongoing open-label compassionate-use study in US patients with either IPF or secondary pulmonary fibrosis.

Meta-analyses of treatment-emergent serious adverse events using data from ASCEND,[34] CAPACITY 1&2[49] and SP3[38] 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).

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.

There are two ongoing studies to evaluate safety: the non-randomised, non-controlled, open-label extension study that included a set of patients who completed either ASCEND, CAPACITY 1 or 2 (RECAP)[40] and PIPF-002,[65] an ongoing open-label compassionate-use study in US patients with either IPF or secondary pulmonary fibrosis.

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 unlicensed dose (1,800mg/day) of pirfenidone 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 contributing to each network, a pragmatic approach was adopted, whereby trials were included regardless of these differences.

Despite including all available evidence in the NMAs, there were still a limited number of studies for certain outcomes. For binomial outcomes, there were too few studies to estimate the between-study variance from the sample data alone and weakly informative priors were used.

For the INPULSIS studies,[72] trial-level treatment effects were not available for two outcomes (PFS and IPF related mortality). Pooled HRs were therefore used to inform the NMA for these outcomes.

5 COST EFFECTIVENESS

5.1 ERG comment on company's review of cost-effectiveness evidence

5.1.1 Search strategy

A single search strategy was conducted in November 2015 to identify cost-effectiveness studies, HRQoL and resource use data. The ERG notes that the search was sufficiently comprehensive and sensitive and the ERG obtained a similar result when reproducing the searches. The structure of the search strategy was: Idiopathic pulmonary fibrosis AND (resource use OR cost-effectiveness OR utilities).

The following sources were searched:

- MEDLINE 1946 to 2015 November 16
- MEDLINE In-Process
- EMBASE 1974 to 2015 November 16
- Econlit 1886 to October 2015
- The Cochrane Database of Systematic Reviews
- The Health Technology Assessment (HTA) Database Issue 4 of 4, October 2015 (Cochrane Library)
- NHS Economic Evaluation Database (NHS EED): Issue 2 of 4, April 2015 (Cochrane Library
- Cost-Effectiveness Analysis registry
- PROQOLID
- ScHARRHUD
- EuroQol database

Supplementary searching included searching key HTA websites (NICE; the Pharmaceutical Benefits Scheme [PBS]; the Canadian Agency for Drugs and Technologies in Health [CADTH], and the Scottish Medicines Consortium [SMC]). NICE submissions were hand-searched and Google Scholar and conference posters and abstracts were also searched over the period 2014 to 2016. The CS states that a 'recent systematic review' was also hand-searched but does not report the citation details of the particular review.

The CS does not provide a reference to any published filter used in the search. However, the utilities search filter appears to have been directly derived (with no variation) from Arber et al,[87] whilst the cost-effectiveness filter appears to be a slightly modified version of the NHS EED search filter.[88] The company reported in their clarification response (see clarification response,[10] question B5) that this was amended in order to increase the sensitivity.[44]

No date or language restrictions were applied to the searches; however, the CS states that only studies that were published after 2010 were screened. The date limit was applied because although NHS EED was omitted from the original submission, the ERG report relating to the previous pirfenidone appraisal stated that this database was checked and no additional studies were identified.[89]

The ERG agrees that it was not appropriate to apply filters to the searches run on databases with an economic focus including Econlit and NHS EED, as these databases have a specific economic focus. The ERG notes that the reporting of the searches is very thorough and includes screenshots of the searches conducted on Google Scholar and conference abstracts.

5.1.2 Inclusion / exclusion criteria for the review of published cost-effectiveness studies

The CS (page 189) reports that study selection followed a two-stage process involving: (a) the assessment of titles and abstracts of potentially relevant studies by a single reviewer, checked independently by a second reviewer, followed by: (b) re-assessment of full texts of potentially includable studies against what the company refers to as the "*systematic review eligibility criteria.*"

The inclusion and exclusion criteria adopted are not clearly reported within the CS or accompanying appendices. The CS did not provide an explicit list of inclusion/exclusion criteria for the review of published cost-effectiveness studies. The CS states that the aim of the review was to identify cost-effectiveness studies of pirfenidone for adult patients with mild to moderate IPF in England. It also states that full economic evaluations were included as well as relevant economic data reported in technology assessments. The CS states that obviously irrelevant records (such as animal studies and studies about ineligible populations) were removed. Excluded studies are tabulated in Appendix 18 and the most common reasons for exclusion were either an ineligible population or the reporting of ineligible outcomes, however the appropriateness of these exclusions cannot be assessed without knowing explicitly which populations and outcomes were deemed relevant.

Included studies were assessed using the checklist reported by Drummond and Jefferson[90] by one reviewer and checked by a second reviewer. Studies were not selected or excluded from the review based on quality assessment.

5.1.3 *Studies included in the review of published cost-effectiveness studies*

The company's electronic searches yielded 3,474 potentially relevant unique citations for the single search to identify cost-effectiveness studies, HRQoL studies and resource use data. Of these, 4 studies (reported across 5 references according to the company) were included in the review of cost-effectiveness studies.[42, 91-94] The CS justifies the exclusion of the cost-effectiveness model used in the 2015 nintedanib NICE submission[26] on the basis that the model was for "*all patients with IPF and not*

just those patients with mild to moderate disease". The ERG disagrees with this exclusion because the modelled population in the nintedanib appraisal related to patients with a percent predicted FVC above 50%, even though this was a narrower population than that covered by the nintedanib licensed indication. The ERG considers the exclusion of this study to be inappropriate as it addressed a similar decision problem to that considered within the current pirfenidone appraisal.

The ERG notes that a total of 6 references are presented by the company instead of five (corresponding to 4 studies). This includes the model used in the previous submission to NICE reported in two references,[42, 89] the model developed by Loveman *et al.* (2014) for a health technology assessment of all available treatments for IPF reported in two references,[93, 94] and two separate Common Drug Review (CDR) reports published by the CADTH for nintedanib[91] and pirfenidone.[92] A table of reasons for exclusion of studies is presented in CS Appendix 18. The ERG notes some inconsistencies in that the CDR for pirfenidone published in 2015 included in the company's review is a re-submission and that an initial assessment was conducted in 2013; the original submission is not included in the company's systematic review.

The ERG notes that the included studies vary in terms of modelling approach. The model submitted to NICE by the company during the previous appraisal of pirfenidone (TA282), used a micro-simulation approach whereby surrogate outcomes (FVC and 6MWD) are used to estimate the risk of IPF-related mortality.[42, 89] In contrast, the model developed by Loveman *et al.* (2014) used a cohort state transition approach whereby OS is modelled as a function of PFS.[93, 94] The modelling approach used in the CDRs for nintedanib and pirfenidone are less clear given the lack of details provided in these brief reports.[91, 92] Effectiveness data and sources for utility values also vary between these studies. Data from the ASCEND trial[34] were not available during the previous submission to NICE[42, 89] or HTA by Loveman et al[93, 94] and therefore are only included in the two CDRs.[91, 92] Utility values in the previous model submitted to NICE were taken from the CAPACITY trials[49] based on the SGRQ scores mapped onto EQ-5D utilities based on an algorithm developed in COPD by Starkie *et al.* (2011).[95] The model developed by Loveman et al[93, 94] used utility values from two studies conducted under the auspices of the IPFCRN[75, 96] in the US. The pirfenidone model previously submitted to NICE[42, 89] took discontinuation rates from the trials and did not include a stopping rule. Although unclear, it also appears that no stopping rule was applied in the analyses submitted to the CDR for nintedanib[91] and pirfenidone.[92] In contrast, Loveman *et al.* (2014)[93, 94] assumed that treatments are discontinued following progression. ICERs reported also varied between studies with some ICERs only being available after the application of confidential price discounts. The previous pirfenidone model submitted to NICE[42, 89] reported an ICER for pirfenidone versus BSC for patients with percent predicted FVC ≤80% of £25,969 per QALY gained following a confidential price reduction. The ICER for pirfenidone was CAN$78,024 per QALY gained against BSC in the CDR for pirfenidone

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 7 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 IMPULSIS 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 IMPULSIS 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 IMPULSIS 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

original model and the changes made in the revised model are both summarised here, however the results are presented only for the revised model.

The company's economic evaluation (described in Table 44) assesses the cost-effectiveness of pirfenidone versus BSC from the perspective of the UK NHS and Personal Social Services (PSS) in three populations:

(i) the ITT trial population of the ASCEND/CAPACITY/RECAP trials,[34, 40, 49] comprising of adults with mild to moderate IPF at baseline;

(ii) a subgroup of people with a percent predicted FVC >80% at baseline (considered by the company to be mild IPF);

(iii) a subgroup of people with a percent predicted FVC > 50% and ≤ 80% at baseline (considered by the company to be moderate IPF)

Within the percent predicted FVC of 50 - 80% subgroup, a comparison of pirfenidone against both BSC and nintedanib is evaluated.

The company's model uses a lifetime horizon. All costs and health outcomes are discounted at a rate of 3.5% per annum.

Table 44: Scope of the company's health economic analysis

Population	(i) ITT - trial population – people with Mild to Moderate IPF (ii) People with a percent predicted FVC >80% at baseline (considered by the company to be mild IPF) (iii) People with a percent predicted FVC of 50 - 80% at baseline (considered by the company to be moderate IPF).
Interventions and comparators	For the ITT-trial population, the base-case analysis compares: • pirfenidone versus BSC[a] For people with a percent predicted FVC >80% at baseline (considered to be mild IPF), the base-case analysis compares: • pirfenidone versus BSC[a] For people with a percent predicted FVC of 50 - 80% at baseline (considered to be moderate IPF), the base-case analysis compares: • pirfenidone versus (i) nintedanib or (ii) BSC[a]
Primary health economic outcome	Incremental cost per QALY gained
Synthesis of health effects	The majority of clinical effectiveness and safety estimates included in the model are based on a systematic review of the literature and results are taken from NMAs.
Measuring and valuing health effects	The utility values for the main model health states (progression-free and progressed) were derived by mapping from a disease specific HRQoL instrument (SGRQ) measured in people with IPF to the EQ-5D-3L. The mapping algorithm between the SGRQ and EQ-5D-3L was estimated in a population with IPF from England. The utility decrements for AEs were based on the submission made by the company for nintedanib during TA379.[26]
Perspective	NHS and PSS for costs Direct health impact on patients only for outcomes (i.e. no carer QALYs are included)

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 RECAP extension trial; therefore, no stopping rule is applied 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 capacity for carbon monoxide greater than or equal to 35%.*" Clinical experts to the ERG indicated that it is challenging to assess the severity in IPF but considered the population included in the clinical trials and, by extension, in the model, to be consistent with the definition of mild to moderate IPF used in clinical practice.

In addition to the ITT population (adults with mild to moderate IPF), the company reports results for people with a percent predicted FVC > 80% and 50 - 80% at baseline, and considers these populations to be people with mild and moderate IPF, respectively. Clinical experts to the ERG reiterated that it is challenging to assess the severity in IPF and that percent predicted FVC alone may not be a sufficient surrogate marker and that DLco may be a better indicator of the severity in IPF. The final NICE scope[3] suggests that "*if evidence allows, subgroup analysis by disease severity, defined by FVC (such as above and below or 80% FVC) and/or diffusing capacity for carbon monoxide, will be considered.*" The ERG notes that an analysis by DLco is not presented by the company. The ERG further notes that InterMune (pirfenidone's company at the time of the previous NICE appraisal) in their original submission to NICE considered that "*in clinical practice a FVC of 70% or 80% predicted is often considered to represent mild IPF, whilst a FVC >50% and <70% predicted is considered indicative of moderate IPF (Nathan, 2011) although formal definitions within guidelines have not been made.*"[97] The ERG accepts the challenges in defining the severity in IPF, and considers the subgroups defined by the company to be clinically reasonable and broadly consistent with the final NICE scope.[3] Nevertheless, the ERG would have liked to see an analysis by DLco. The direction of the ICER for any subgroups using DLco as a stratification factor is unclear.

The company's model also reflects the population included in the ASCEND,[34] CAPACITY,[49] and RECAP extension trials.[40] As described in Section 3.1, the ERG observes that the populations recruited in those trials may not be fully reflective of a typical clinical population, notably;

- The majority of individuals recruited in the trials (approximately 75%) had a percent predicted FVC of 50 - 80% at baseline but the proportion with mild IPF may be higher in the UK;
- The majority of trial participants were not recruited in the UK and BSC may vary internationally particularly in countries without universal access to healthcare;
- Patients with comorbidities, particularly emphysema, were excluded from the trials but these patients may be offered treatment in current practice if their FVC is in the range of 50% to 80%.

Furthermore, the ERG notes that people included in the RECAP OLE study were pre-selected in that only people who were compliant to the drug (defined as compliance of ≥ 80% of dose) were included.

Finally, the CS reports results from the ITT population, a combination of people with a percent predicted FVC of 50 - 80% and >80% at baseline; as suggested in the final NICE scope.[3] The comparators specified in the final scope are different within these two populations. Nintedanib is a comparator in people with a percent predicted FVC of 50 - 80% (which composed the majority of people included in the trials) but not >80% at baseline. The correct interpretation of the results for the ITT population is therefore problematic, as the comparison is made only against BSC. The ERG advises that the subgroups of people with a percent predicted FVC of 50 - 80% and >80% at baseline should be interpreted separately for this reason.

5.2.1.2. ERG's comments on the treatment regimen assumed for the intervention

The ERG is largely satisfied with the treatment regimen for the intervention (pirfenidone) assumed in the company's model. The ERG notes that according to the SmPC,[1] the dose should be titrated over a 14-day period when initiating pirfenidone treatment according to the following schedule; one capsule, three times a day (801mg/day) in the first week and two capsules, three times a day (1,602mg/day) in the second week of initiating treatment. Individuals receive three capsules, three times a day (2,403mg/day) from week 2 onwards. The ERG notes that dose titrations have not been explicitly included in the company's model. Instead the average dose over the trial period following titration has been applied in the model.

In the company's base-case, people initiating pirfenidone discontinue at the rate observed in the Phase III trials.[34, 49] The appropriateness of the company's decision to not include a stopping rule is questionable. The ERG notes that the licensing of pirfenidone[1] does not specify a stopping rule. However, NICE issued a stopping rule for the use for pirfenidone[98] (TA379) and nintedanib[78] (TA282) in England and recommends that both treatments should be discontinued if there is evidence of disease progression (defined as a decline in predicted FVC of 10% or more within any 12 month period). The company justifies the exclusion of the stopping rule on the basis of: (i) the high unmet need for people with IPF; (ii) evidence that pirfenidone may benefit people with or without disease progression, and; (iii) references to arguments regarding the difficulty of imposing such a stopping rule from the nintedanib submission, and diverse comments received at the scoping consultation for this appraisal and during the consultation on the Appraisal Consultation Document (ACD) for nintedanib.

The ERG recognises that this issue may be open to debate; nevertheless, the ERG considers that an analysis including the stopping rule for pirfenidone and nintedanib should represent the base-case as this reflects current clinical practice in England. Clinical advice received by the ERG confirmed that the stopping rule defined by NICE has been implemented successfully in practice and that audits are regularly conducted to ensure that clinics comply with these rules. The ERG further notes that the NICE

Appraisal Committee considered the views expressed regarding the difficulty of implementing the stopping rule during the appraisal for nintedanib and concluded in the Final Appraisal Determination (FAD)[78] that: "*The Committee recognised the limitations of FVC but understood that in clinical practice the wider patient characteristics would be taken into account in interpreting percent predicted FVC. Clinical experts noted that they follow the stopping rule in NICE's technology appraisal guidance on pirfenidone for treating idiopathic pulmonary fibrosis, but explained that before withdrawing treatment they retest FVC to confirm that the 10% drop is not temporary, which might happen with an infection. The Committee concluded that, although it has some limitations, percent predicted FVC is the most reliable and widely used measure of lung function in clinical practice.*" The ERG further notes that the approach used by the company is somewhat inconsistent in that an identical stopping rule has been included in the NICE guidance for nintedanib (TA282) and pirfenidone (TA379) but the stopping rule is applied for nintedanib in the base-case but not for pirfenidone.

The ERG notes that whilst a scenario analysis is presented by the company including a stopping rule for both pirfenidone and nintedanib, the implementation of the stopping rule within the model lacks validity. This issue is further described in Section 5.2.2.2.

Finally, in the company's base-case analysis (assuming no stopping rule for pirfenidone), the duration and dosage of treatment is based on the discontinuation rate and dosage observed in the clinical trials.[34, 40, 49] The ERG is unclear whether the dosage received is representative of clinical practice and whether people would be treated for a shorter or longer duration than that assumed within the model. Nevertheless, the ERG considers that using the dose intensity and discontinuation rates from the same trials were used to generate the effectiveness estimates, could be considered reasonable as this ensures consistency in the extrapolated costs and benefits.

5.2.1.3. ERG's comments on the comparators included within the CS and company's model

In people with a percent predicted FVC of 50- 80% at baseline (considered to be moderate IPF), pirfenidone is compared with BSC (defined in the trial as symptom relief, pulmonary rehabilitation, management of comorbidities, and end of life care, including oxygen therapy) and nintedanib. The ERG considers the comparators included in the company's model for this subgroup to be appropriate as this is in line with the recent NICE recommendation regarding the use of nintedanib in adults with a percent predicted FVC of 50 - 80% at baseline[78] and the marketing authorisation for nintedanib.[17]

In people with a percent predicted FVC > 80% at baseline (considered to be mild IPF), pirfenidone is compared with BSC only. No analysis is presented against nintedanib. The ERG considers the comparators included for this subgroup to be appropriate. Whilst nintedanib is licensed in this population,[17] NICE did not issue a positive recommendation for nintedanib in this subgroup.[78]

For the ITT-trial population, a combination of people with mild to moderate IPF, the only comparator considered is BSC. This is justified by the company on the basis that nintedanib has not been recommended by NICE for the treatment of people with a percent predicted FVC > 80% at baseline (see CS on page 207). The ITT-trial population represents a combination of those people with a percent predicted FVC of 50 - 80% or >80% at baseline; a proportion of these people are clearly suitable for treatment with nintedanib, which is not a comparator in the ITT analysis. The ERG further observes that a large majority of people (approx. 75% - see Table 16 in CS in page 198) included in the ASCEND/CAPACITY trials[34, 49] had a percent predicted FVC of 50 – 80% at baseline. The ERG advises that the subgroups of people with a percent predicted FVC of 50% to 80% and >80% at baseline should be interpreted separately.

Finally, within the company's model, the efficacy for BSC reflects the mix of therapies used in the ASCEND/CAPACITY trials[34,49] and includes interventions aiming to relieve symptoms, pulmonary rehabilitation, management of comorbidities, and end of life care, including oxygen therapy. The ERG notes that people in the ASCEND[34] and CAPACITY[49] trials were recruited from a large number of centres worldwide (127 sites in Australia, Brazil, Croatia, Israel, Mexico, New Zealand, Peru, Singapore, and the US for ASCEND and 110 centres in Australia, Europe, and North America for CAPACITY), with potentially varying clinical practice. The generalisability to the UK of treatments received as part of BSC within the ASCEND[34] and CAPACITY[49] trial populations is unclear, particularly for patients in those countries without universal access to healthcare.

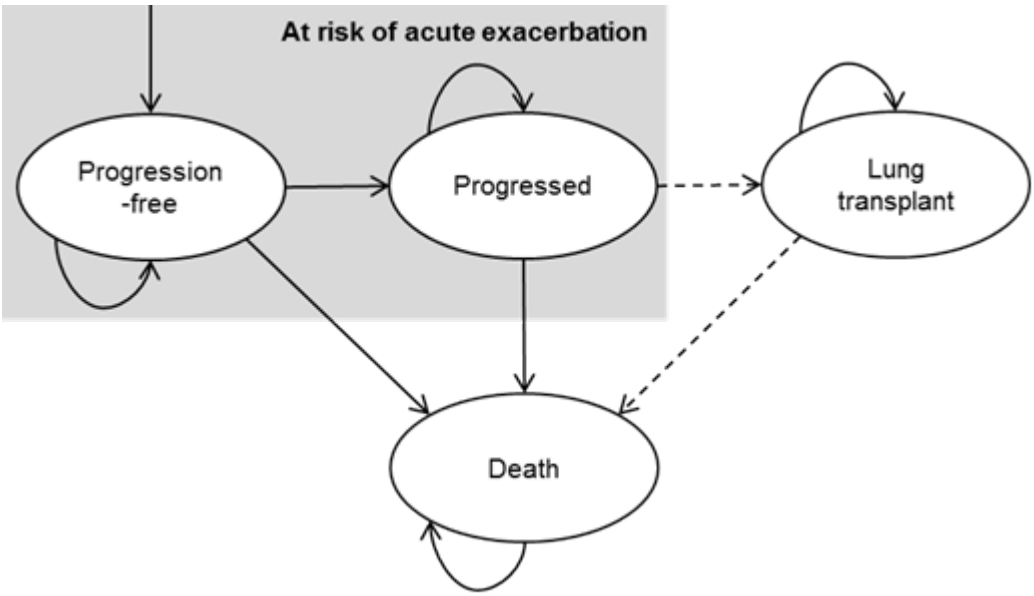
5.2.1.4. ERG's comments on the perspective, discounting and time horizon used in company's base-case

The company's base-case assesses costs and benefits over a lifetime horizon and adopts a UK NHS and PSS perspective. All costs and health outcomes are half-cycle corrected and discounted at a rate of 3.5% per annum. The ERG considers these to be appropriate and in line with the NICE Reference Case.[43]

5.2.2. Description and critique of the company's health economic model structure and logic

The description of the model's logic is based on information contained within the CS, and the ERG's assessment of the economic model. A simplified representation of the company's model structure is shown graphically in Figure 33. In summary, the model structure presented in the CS is based on three main health states; progression-free, progressed disease and death. Health states for progression-free and progressed disease are further sub-divided into 'on-treatment' and 'off-treatment' periods (not shown in Figure 33). The model uses a 3-monthly cycle length.

Figure 33: Model structure (reproduced from CS,[4] Figure 42, page 205)



The company's model adopts a cohort-based partitioned survival approach whereby the OS, PFS and discontinuation curves from the Phase III trials[34,40,49] for pirfenidone are extrapolated over a lifetime horizon using parametric functions. These parametric functions are used to calculate the proportion of individuals in each health state over time. The time in the progressed disease health state is derived as the difference between the extrapolated OS and PFS curves. Consequently, movement between health states is not modelled using transitions probabilities, so this is not a traditional transition-state (Markov) model.

Treatment effects (HRs/RRs) estimated from the NMAs for BSC and nintedanib versus pirfenidone (with pirfenidone representing the baseline) are subsequently applied to the baseline hazards to estimate the hazards in people initiating nintedanib and BSC (see Section 5.2.4). The HRs/RRs are applied over the entire time horizon in the company's base-case, thereby assuming constant proportional hazards. Scenario analyses were conducted by the company whereby the treatment effects was assumed to stop after 7, 10 and 14 years. People initiating pirfenidone and nintedanib are assumed to receive BSC following treatment cessation.

In addition to the three main health states (progression-free, progressed disease and death), lung transplantation is included as a separate health state which is not used in the base-case. The model also includes the impact of acute exacerbations on HRQoL and resource use; these are not modelled 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 three main health states (progression-free, progressed disease and death). 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,
contrary to the company's argument, the ERG considers that this simplification has the potential to overestimate the lifetime QALYs gained as the impact of subsequent progression on HRQoL is not captured. This overestimation could be favourable to pirfenidone as any survival gain for pirfenidone

will translate into a larger QALY gain if subsequent declines in HRQoL after progression are ignored. Whilst the company's model structure made it difficult for the ERG to directly estimate the impact of this simplification on the incremental QALYs and ICER, an exploratory analysis conducted by the ERG (see Section 6) adjusting utility by age (and therefore assuming some form of progression – although with limitations) led to an increase in the ICERs of pirfenidone versus BSC.

Furthermore, within the company's model, all disease progression is assumed to be equally detrimental. Clinical advisors to the ERG considered that a 10% drop in percent predicted FVC would impact on HRQoL differently according to the baseline percent predicted FVC and therefore the clinical impact of disease progression, as defined in the model, would be different across individuals. The ERG notes that the model used in the nintedanib appraisal provides a better representation of the natural history in IPF, whereby individuals transit through multiple health states with different levels of percent predicted FVC (rather than just two), as their disease progresses. This structure allows for different HRQoL and cost estimates to be attached according to the individual's percent predicted FVC level. The model structure used in the nintedanib company submission was also considered by the clinical advisors to the ERG to be more representative of the progressive nature of IPF than the pre/post progression model presented by the company for pirfenidone.

In addition to the three main health states (progression-free, progressed disease and death), the company attempts to includes two key features of IPF; the impact on costs and health outcomes of acute exacerbations in the base-case and lung transplantations in a scenario analysis. The ERG considers the approach taken by the company to include lung transplantations as a scenario analysis to be appropriate given the uncertainty in the data available and the potential difficulty in incorporating lung transplantation within a cohort model. The ERG notes that the company's inclusion of lung transplantations relies on a series of assumptions and adjustments but this scenario analysis has a minimal impact on the ICER (an increase from [redacted] per QALY gained in the ITT population for the comparison of pirfenidone versus BSC).

The ERG considers the inclusion of acute exacerbations in the base-case to be appropriate given that exacerbations are considered to be an important clinical event in IPF.[78] Within the company's model, the impact of acute exacerbations is applied as a cost and HRQoL decrement during each model cycle and individuals could remain in the progression-free health state following an exacerbation. Clinical advisors to the ERG noted that the diagnosis of acute exacerbations is challenging and that it is often difficult to distinguish between an exacerbation and progression. Clinical advisors to the ERG suggested that people who have experienced an exacerbation would usually be considered to have progressed. The ERG further notes from discussions held during the nintedanib appraisal that exacerbations are associated with high morbidity and mortality and therefore delaying/preventing

exacerbations is an important aspect of maintaining quality of life.[78] Nevertheless, the ERG notes that the inclusion of acute exacerbations (as implemented by the company) has a minimal impact on the ICER (an increase from [redacted] per QALY gained in the ITT population for the comparison of pirfenidone versus BSC excluding acute exacerbations). The ERG considers the lack of impact associated with the inclusion of exacerbations in the model to be an artefact of the company's chosen model structure rather than a reflection on the relevance of exacerbations in IPF. This is because acute exacerbations are disconnected from the outcomes of progression and survival and are instead included as a simple cost and utility decrement during each model cycle.

The ERG further notes that within the company's model, the impact of exacerbations on costs and outcomes is modelled inconsistently and relies on a series of strong assumptions which are often not adequately supported by the evidence (especially over the long-term). The ERG notes that the impact of exacerbations on health outcomes is modelled by estimating the risk of exacerbations whilst on a particular treatment and applying utility decrements to those individuals having an exacerbation. In contrast, the impact of exacerbations on costs is included separately as a cost of hospitalisation specific to the treatment received (independent of the rate of exacerbations). It should be noted that in response to a request for clarification from the ERG (see clarification response,[10] question B15), the company confirmed that hospitalisation costs included in the model are not specific to acute exacerbations.

5.2.2.2. ERG's comments on the general modelling approach

The company's model adopts a partitioned survival approach and the CS (page 203) refers to the model published by Loveman *et al.* (2015).[94] The ERG notes that whilst both the Loveman *et al.* model and the company's model are based on PFS (although different definitions are used), each uses a different analytical approach (partitioned survival or state transition).

In the company's model, the OS, PFS and discontinuation curves from the trials are extrapolated using parametric functions and modelled independently from each other; these are used to determine the health state occupancy within the model. Within the company's model, individuals could also remain on treatment following progression. In contrast, in Loveman *et al.* (2015), a state transition approach is used and OS is estimated indirectly by assuming a relationship between OS and PFS. In the model described by Loveman *et al.*, treatment is assumed to be discontinued following progression. The ERG notes that both state-transition and partitioned survival approaches are used in the evaluation of cancer treatments and that both approaches have advantages and limitations. The choice between approaches is often not straightforward and needs to be considered with respect to the quality and quantity of data available and whether the resulting model structure has face validity given the characteristics of the disease being modelled.

The ERG considers that whilst the partitioned-survival modelling approach is commonly used, the implementation of this approach in the company's model means that the outcomes of OS, PFS and discontinuation are modelled independently of each other. In simple terms, in the company's model, a change in either PFS or time to discontinuation has no impact on OS. To illustrate this, the ERG compared outcomes estimated when assuming no stopping rule (scenario 1 –company's base-case) with those estimated when assuming the stopping rule (scenario 2; as programmed by the company – company's scenario analysis). As can be seen from Table 45, different assumptions relating to the time on treatment have no impact on the mean life years, but impact treatment costs, and therefore the ICERs for pirfenidone.

Table 45: Impact of the stopping rule on health outcomes & treatment costs for pirfenidone and ICER against BSC for the ITT-trial population (results are discounted and half-cycle corrected)

	Scenario 1 - no stopping rule (company's base-case)	**Scenario 2 - Stopping rule (company's scenario analysis)**
Mean time on treatment (in years)	3.29	2.08
Mean time in PFS (in years)	2.05	2.05
Mean time in progressed disease (in years)	6.62	6.62
Mean life years	8.67	8.67
Treatment costs	[redacted]	[redacted]
ICER (vs. BSC)	[redacted]	[redacted]

During clarification, the ERG asked the company to provide evidence to support the assumption that time on treatment is independent from PFS and OS (see clarification response,[10] question B6). In response, the company stated that "*The ERG is correct that the model was constructed utilising the simplifying assumption that time on treatment, OS and PFS are independent of each other. This is a common practice in NICE submissions using time to event data (such as oncology submissions where disease is similar in severity and impact to IPF). To accurately quantify the relationship between time on treatment, OS and PFS, additional data would be required which are not publically available for nintedanib. Recent studies comparing the state-transition method (i.e. modelling time on treatment, PFS and OS separately) and area-under-the-curve (AUC) partitioned survival models show that the two methods produce similar results, and that either approach may be considered appropriate to a given*

decision problem, depending on the available data and scope of the evaluation [Briggs 2015]. We consider our approach the most appropriate given the data available".

The ERG considers the response from the company to be misleading. The company makes reference to a single case study conducted in advanced melanoma showing that the two methods provide similar results and could be appropriate in this particular case. However, the ERG is aware that different analytical approaches could lead to different estimates in other conditions, as shown in TA 257.[100] When deciding between modelling approaches it is important to consider the face validity of the model structure and any assumptions inherent within the structure as well as the amount and quality of the data available to parameterise the model.

Importantly, the ERG considers the modelling approach used by the company to be reasonable when the stopping rule is excluded; but inadequate when implementing a stopping rule given that treatment duration and treatment outcomes are disconnected from each other. The company's implementation of the stopping rule using tunnel states was also cumbersome and was not well described in the original submission but additional details were provided following the clarification request by NICE (see clarification response,[10] question B8. The company identified errors in the implementation which were corrected following the clarification request (see clarification response,[10] questions B8 – B10 and B23).

The ERG acknowledges that the implementation of a stopping rule, which was not implemented in the clinical trials, will usually be reliant on some assumptions to estimate treatment outcomes in those that discontinue due to the stopping rule, irrespective of the modelling approach chosen. Clinical advisors to the ERG commented that it hard to understand the relationship between treatment discontinuation and clinical outcomes such as disease progression and all-cause mortality because IPF is a heterogeneous condition with natural variability in the rates of decline in percent predicted FVC and the mechanism of action of pirfenidone has not been fully established. However, the ERG does not believe that the company's assumption that there is no relationship between treatment duration and treatment outcomes, such as PFS and OS, to be plausible. [redacted]

The ERG considers that the ICERs presented by the company using the stopping rule could represent a lower bound of the true ICER when the stopping rule is implemented in clinical practice, as the life-time costs of treatment are reduced when the stopping rule is applied in the model, but the incremental QALYs are not reduced by the shorter duration of treatment.

The ERG further notes that the CS includes a long description of the relationship between percent predicted FVC and OS to justify the definition of progression used in the model, but given that PFS and OS are modelled separately, no relationship is modelled between outcomes and therefore the definition of progression used in the model has no impact on OS.

5.2.2.3. ERG comments on the use HR for the comparators

The company estimates the baseline hazards of death, progression and discontinuation in people initiating pirfenidone from individual IPD from the CAPACITY,[49] ASCEND,[34] and RECAP trials[40] for all three populations evaluated in the model; i.e. the ITT population, and the subgroups of people with a percent predicted FVC of 50% - 80% and >80% at baseline. HRs taken from the NMAs are then applied to the hazards from the pirfenidone arms to estimate the hazards in people initiating nintedanib and BSC. Alternatives for OS for people initiating BSC are explored in scenario analyses such as using the Kaplan Meier (KM) curve up to the end of the observed period followed by extrapolation using HRs.

The ERG considers the use of HRs to capture the treatment effect to be reasonable and pragmatic with respect to the data available and the limited duration of follow-up in the evidence base for both nintedanib and BSC. Nevertheless, the ERG has a number of concerns with the values used and the duration over which the treatment effect is assumed to be constant in the company's base-case analysis. These issues are described in Section 5.2.2.5 and 5.2.4.1 respectively.

5.2.2.4. ERG's comments regarding the representation of the treatment pathway

The company's model assumes that people initiating pirfenidone and nintedanib receive BSC upon treatment discontinuation. The ERG considers that the treatment pathway assumed by the company is questionable. Nintedanib received a positive NICE recommendation in people with a percent predicted FVC of 50 - 80% at baseline; therefore it is possible that nintedanib could be used following the discontinuation of pirfenidone if individuals maintain a percent predicted FVC > 50%. Similarly, in principle, pirfenidone could be used following the discontinuation of nintedanib. Clinical advisors to the ERG suggested that in practice, people initiating pirfenidone may switch to nintedanib upon discontinuation, and vice versa. This was acknowledged in the company's clarification response (see clarification response,[10] question B7), but the absence of sequences was justified by the company on the basis that a similar approach was used in the nintedanib appraisal.[10, 12]

The ERG notes that including treatment sequences within the economic model would require a complete restructuring of the model and the impact of their inclusion on the ICER is unclear.

5.2.2.5. ERG's comments regarding the assumption of proportional hazards

A key structural assumption in the company's model is that the treatment effect estimated at week 52 holds for the entire duration of the model (34 years) for both nintedanib and BSC against pirfenidone. The ERG considers the assumption of proportional hazards over the entire model duration to be overly optimistic and inadequately supported by the evidence for either pirfenidone against BSC or nintedanib.

The assumption of proportional hazards is somewhat justified by the company for the treatment effect between BSC and pirfenidone based on: (i) *post hoc* analyses conducted by the company (see CS,[4] Appendix 20) in the CAPACITY/ASCEND trials[34,49] which did not show a significant interaction between the treatment effect and time (see CS,[4] page 207 and clarification response,[10] question B12) for OS and PFS, and; (ii) inspection of the log-cumulative hazard between people initiating pirfenidone in the ASCEND/CAPACITY/RECAP trials[34,40,49] and (iii) data from three long-terms registries (Edinburgh, INOVA and EuroIPF).

The ERG has a number of concerns, which are discussed in turn:

- Despite there being no statistical evidence to contradict the assumption of proportional hazards between pirfenidone and BSC (up to 72 weeks for PFS and last follow-up for OS), the ERG notes that evidence from the CAPACITY-trials[49] reported a smaller treatment effect for OS between week 52 (HR: 0.49; 95% CI: 0.24 – 1.01) and week 72 (HR: 0.77; 95% CI: 0.47 – 1.28). Whilst the difference is not statistically significant, the ERG considers that the strong assumption of proportional hazards remains questionable. The ERG further observes a discrepancy in the company's argument in that the HRs estimated using data at 52 weeks are used in the company's base-case. As discussed in Section 4.7, the ERG considers that if the assumption of proportional hazards was valid, then the HR estimated at 72 weeks would be a more appropriate estimate as it incorporates more of the available data.
- The ERG also re-plotted the log-cumulative hazard plots for OS (using KM data available in the company's model) based on data from the ASCEND/CAPACITY/RECAP trials (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,[50] 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%). However, according to data included in the company's model, approximately 8% of people in the ASCEND/CAPACITY trials had a percent predicted FVC ≥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.

Figure 34: Log-cumulative hazard plots for OS within the ASCEND/CAPACITY/RECAP trials (Plot drawn by the ERG)

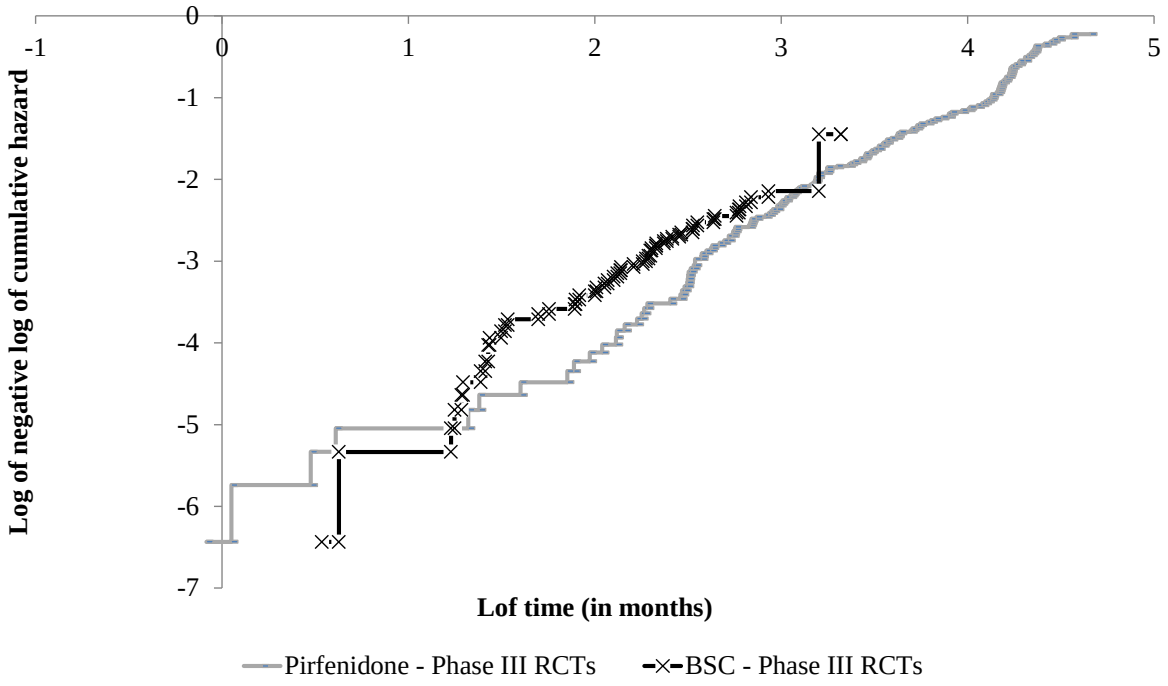
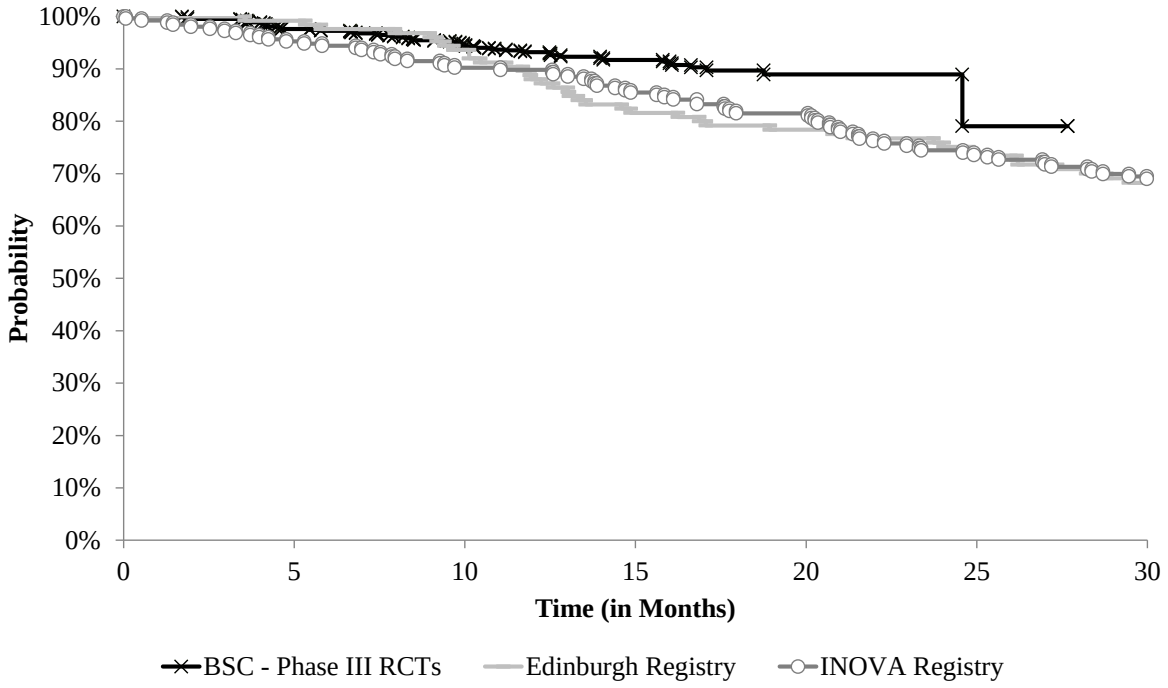


Figure 35: Plot of the OS for BSC from the ASCEND/CAPACITY trials and registries (Plot drawn by the ERG)



As a result, the ERG considers that the evidence presented by the company to support the assumption of proportional hazards for OS between BSC and pirfenidone in the long-term is inconclusive. The ERG notes however that the assumption of proportional hazards for PFS between BSC and pirfenidone appears more conclusive.

Clinical advisors to the ERG commented that it is possible that if a drug fundamentally alters the fibrosis pathway over the duration of a clinical trial, then with continued treatment it may be able to prevent declines over longer time periods. The ERG considers that this statement supports the possibility of continued effectiveness with long-term treatment but does not necessarily support a treatment effect for OS that is constant over the entire model duration.

As acknowledged by the company, the assumption of proportional hazard between pirfenidone and nintedanib is unclear. The ERG considers that assuming the treatment effect to hold for the entire model's duration is overly optimistic. The ERG notes that whilst indirect comparisons conducted by the ERG suggested (see Section 4.7) a slightly greater median treatment effect for pirfenidone using data up to 72 weeks (and excluding SP3) compared with nintedanib for OS (HR: 0.90; 95% CrI: 0.43 – 1.85), the differences were not statistically significant suggesting that the efficacy between nintedanib and pirfenidone could be similar. Results are also uncertain given the considerable heterogeneity between the population included in the trials for pirfenidone and nintedanib. As highlighted during the assessment for nintedanib by the CADTH, "*The two INPULSIS trials did not exclude people with normal lung function, while the ASCEND trial comparing pirfenidone against placebo imposed an upper limit on FVC. This resulted in a clinically meaningful difference in baseline per cent predicted FVC between the INPULSIS and ASCEND trials and suggested that patients in ASCEND may have had more advanced disease. This difference in baseline disease severity may have influenced the number of mortality events in the trials and impacted the ability to observe a mortality benefit with nintedanib.*"[91]

Consequently, the ERG considers that the company's base-case scenario provides a favourable estimate of the plausible ICERs for pirfenidone. Scenario analyses are presented by the company whereby the treatment effect is assumed to stop after 7, 10 and 14 years. The ERG notes that the ICER for pirfenidone compared with BSC for the ITT population increases from [redacted] when the treatment effect is assumed to stop after 7 years. The ERG considers that assuming the treatment effect to stop after 7 years is also arbitrary. The ERG notes that the treatment effect could stop earlier or later than 7 years, and therefore the ERG's preferred base-case are provided, in Section 6, using an optimistic (lifetime) and pessimistic assumption (treatment effect to stop at 2 years approximately at the end of the clinical evidence) regarding the duration of the treatment effect (lifetime to 2 year). This has been done because whilst the clinical advisors to the ERG considered it possible that there may be continued effectiveness with long-term treatment, the duration of persistence for any long-term treatment effect is

currently highly uncertain, particularly given that this is a heterogeneous condition and the mechanism of treatment is not fully understood at this time.

5.2.2.6. ERG's comments regarding the discontinuation with respect to progression

Within the company's model, people initiating pirfenidone and nintedanib could remain on treatment irrespective of progression status. Another structural assumption in the company' model is that the proportion of people who discontinue treatment would be the same irrespective of the progression status. The ERG considers that this is not adequately supported by the evidence. Nevertheless, the ERG notes that given the approach chosen by the company whereby OS, PFS and discontinuation are modelled separately, no impact is expected from this assumption as discontinuation is only used to calculate the treatment costs and treatment discontinuation has no impact on health outcomes.

5.2.3. Derivation of the baseline hazards of death, progression and discontinuation

This section focuses on the estimation of the baseline hazards of death, progression and discontinuation in people initiating pirfenidone. HRs/RRs are subsequently applied to the hazards from the pirfenidone arm to estimate the hazards of death, progression and discontinuation in people initiating BSC or nintedanib. These are discussed in Section 5.2.4. The source of data informing the KM curves for OS, PFS and time to discontinuation are summarised in Table 46.

Table 46: **Source of data informing KM curves (reproduced from clarification response, question B21,[10] Table 24)**

KM data	CAPACITY 1 & 2 (13 Jan 2009)*	ASCEND (14 Feb 2014)*	RECAP (30 June 2015)*
OS			
Pirfenidone – all	√	√	√
Pirfenidone - percent predicted FVC of 50 – 80%	√	√	√
Pirfenidone - percent predicted FVC >80%	√	√	√
BSC	√	√	
PFS			
Pirfenidone – all	√	√	
Pirfenidone - percent predicted FVC of 50 – 80%	√	√	
Pirfenidone - percent predicted FVC >80%	√	√	
BSC	√	√	
TTOT			
Pirfenidone – all	√	√	√
Pirfenidone - percent predicted FVC of 50 – 80%	√	√	√
Pirfenidone - percent predicted FVC >80%	√	√	√

* Date of data-cut

The baseline hazards of death and discontinuation (for reasons other than death and lung transplantations) in people initiating pirfenidone are estimated from IPD from the CAPACITY, ASCEND and RECAP trials[34,40,49] for all three modelled populations. Data from the ASCEND[34] and CAPACITY[49] trials, but not RECAP, are used to estimate the baseline hazards of progression in the company's model. The company justifies the exclusion of RECAP on the basis that progression data were not collected in this trial.

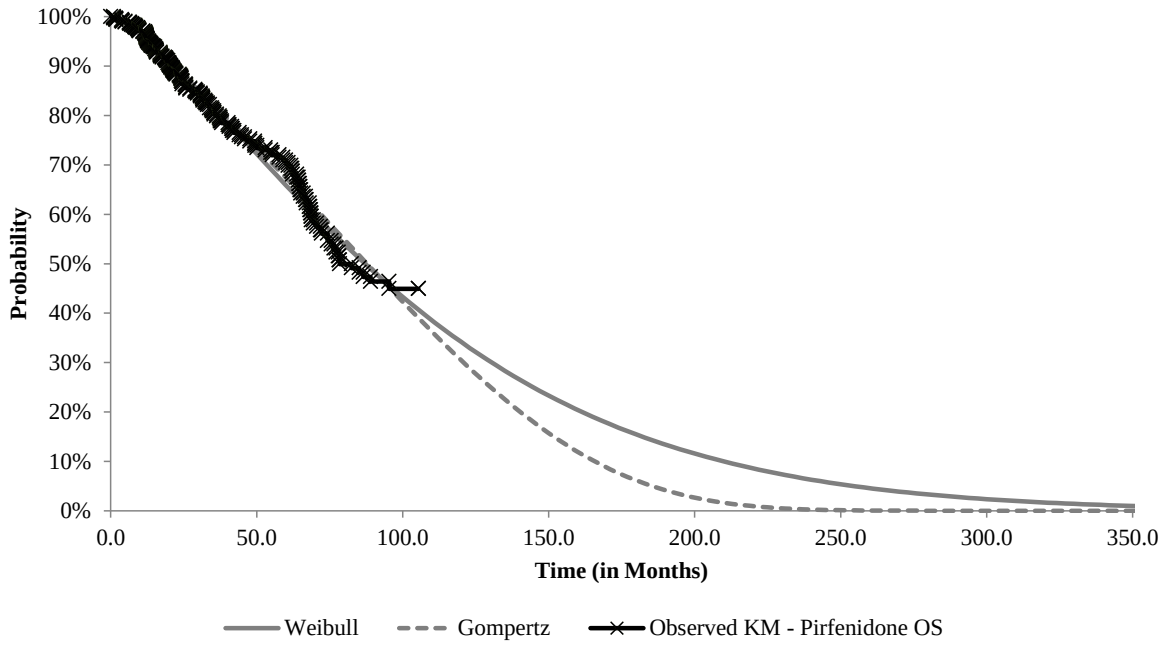
Progression is defined as per the ASCEND trial[34] definition and consists of confirmed ≥10% absolute decline in percent predicted FVC or confirmed ≥50m decline in 6MWD or death. This is principally justified by the company by the lack of data from the ASCEND trial on DLco.

A total of six single parametric functions were fitted to the observed KM curves: exponential, Weibull, Gompertz, log logistic, log normal and gamma. The Weibull distribution was selected for the base-case for the ITT population for all outcomes. This was justified in the CS based upon: (i) visual inspection of the fit during the observed period; (ii) statistical goodness of fit during the observed period (as measured by the Akaike Information Criterion [AIC] and Bayesian Information Criterion [BIC]), and; (iii) plausibility of the long-term extrapolation. Alternative parametric functions are examined in the sensitivity analyses.

5.2.3.1. ERG's comments regarding the estimating of overall survival

The ERG considers the process (i.e. assessing the fit to the observed data and assessing the plausibility of the long-term extrapolation) used by the company to select the most appropriate parametric distribution for OS to be generally appropriate. Nevertheless, the ERG considers the choice of the Weibull distribution in the company's base-case to be questionable and notes that the Gompertz distribution may provide a more clinically plausible extrapolation for OS.

Figure 36: Comparison of the observed KM for OS in people initiating pirfenidone against extrapolation using parametric distributions for the ITT population (Plot drawn by the ERG)



Of the six single distributions examined, the ERG considered the Weibull and Gompertz distributions to provide reasonable fits to both the observed period and a plausible long-term extrapolation in either the ITT population or people with a percent predicted FVC of 50 – 80% or >80% at baseline. Therefore these are the focus of comment in this section. The plot of the observed KM and Weibull and Gompertz

distributions are presented in Figure 36 for the ITT population (and Figure 46 and **Figure 47** in Appendix 1 in people with a percent predicted FVC > 80% at baseline and people with a percent predicted FVC of 50 - 80% at baseline, respectively). The ERG notes that only single parametric distributions are examined by the company and that the use of a piecewise distribution could potentially improve the fit.

5.2.3.1.1. Visual inspection of the fit to the observed period and goodness of fit.

The ERG considers that both the Weibull and Gompertz distributions provide a similar fit to the observed period and that it is difficult to differentiate between the two. The ERG notes that both curves provided a very similar visual fit to the observed period and had broadly similar BIC values (861.89 for Weibull vs. 869.44 for the Gompertz for the ITT population – see CS, Table 72, page 212). The ERG reiterates that goodness of fit criteria only provide an indication of the goodness of fit during the observed period and do not categorically indicate that one distribution should be preferred over alternative distributions.

5.2.3.1.2. Plausibility of the long-term extrapolation

Whilst the Weibull and Gompertz distributions provided a relatively similar fit during the observed period, these distributions provided different long-term extrapolations (Table 36). Contrary to the company, the ERG considers the Gompertz distribution to provide a more realistic long-term extrapolation for the following reasons:

i. A key argument from the company regarding the plausibility of the long-term extrapolation using the Weibull distribution relies on a comparison of the model prediction for BSC and registry data from the INOVA and Edinburgh cohorts. As described in Section 5.2.2.5, the ERG has a number of concerns with the survival observed in these registries compared with people initiating BSC that were enrolled in the ASCEND/CAPACITY trials. As shown in Figure 35, the survival from the registries did not validate the survival observed in people initiating BSC in the ASCEND/CAPACITY trials.[34, 49] The ERG further notes that the HR which is used to model the survival from BSC is taken from results from the NMA which uses data from the ASCEND/CAPACITY trials, and therefore, validating the model against registries is inconsistent when the registry data do not match the control data from the trials. The ERG considers that making inferences about the plausibility of the long-term extrapolation based on the modelled OS for BSC against registry data has limited relevance given the OS data from the registry do not match the placebo arm of the trials. The ERG further notes that both the modelled OS for BSC using the Weibull and Gompertz distribution provided a reasonable fit to the OS from the registries (Figure **37**). Therefore the argument made by the company does not categorically indicate that one distribution

should be preferred over the other one upon inspection of the fit of the modelled OS for BSC with registries.

ii. Second, the ERG notes that the OS curve from the trial and predicted in the model includes death from any cause (IPF and other causes) and that the Weibull distribution has a longer tail compared with the Gompertz distribution; consequently, the hazards of deaths at older ages may be underestimated. The UK life tables provide an estimate of the survival in the general population, in whom the average survival is expected to be greater than the survival observed in people with IPF who have a chronic progressive illness. For the Weibull or Gompertz distributions to be considered appropriate, a higher, or at least, equal hazard of death (compared with the general population life table estimates) should be observed. It can be seen from Figure 38 for the ITT population (and in appendix 2 for the subgroups in Figure 48 and **Figure 49**) that the use of the Weibull distribution in the model leads in some occasions to lower probabilities of death in people with IPF initiating pirfenidone compared with the probability of death from the general population. This is not considered by the ERG to be plausible. In contrast, the Gompertz distribution generates consistently greater probabilities of death when compared with the life tables in England.

As a result, the ERG considers that the Gompertz distribution provides a more plausible extrapolation of OS than the Weibull distribution.

Figure 37: Plot of the KM for OS from registries and modelled survival for BSC using the Weibull and Gompertz distribution (Plot drawn by the ERG)

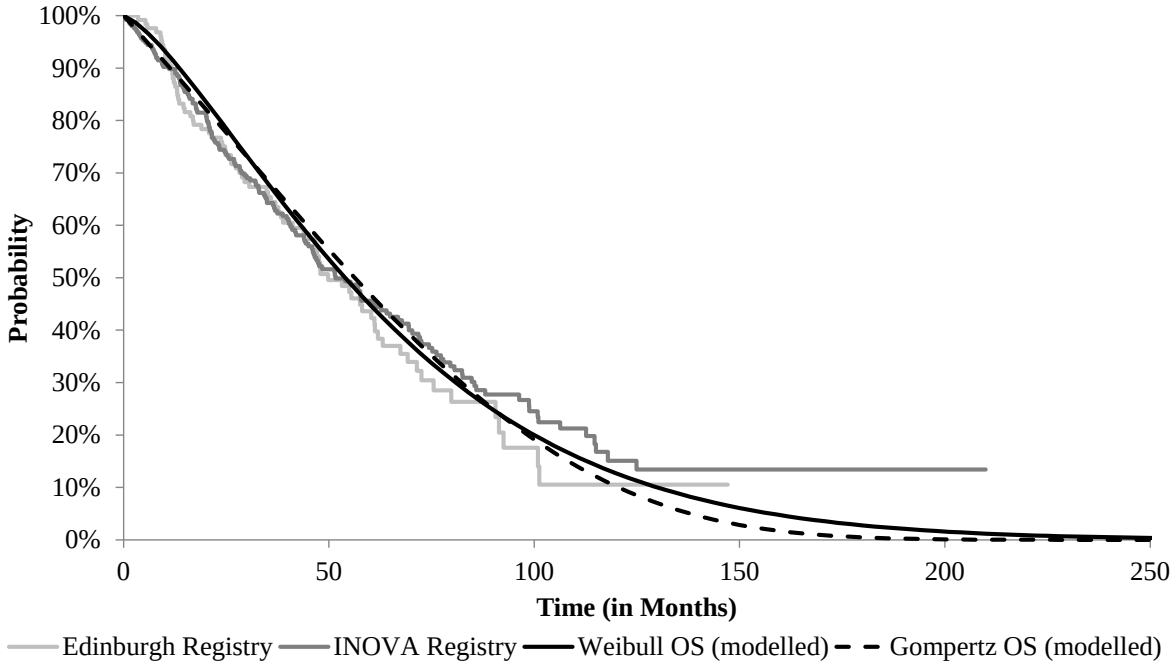
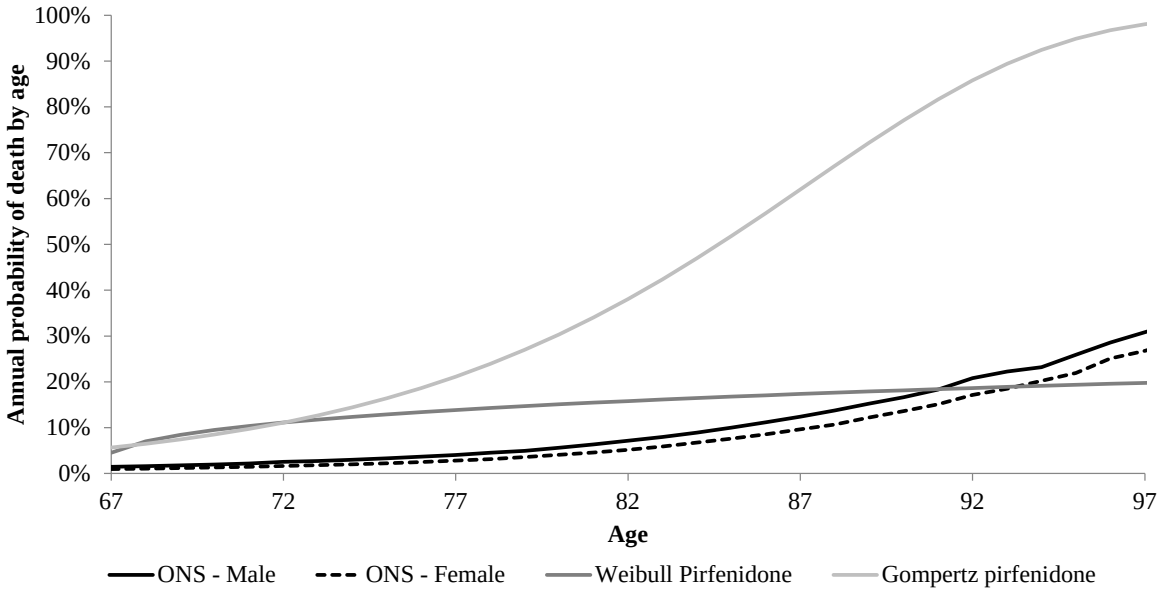


Figure 38: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and life tables in the UK in the ITT population (Plot drawn by the ERG)



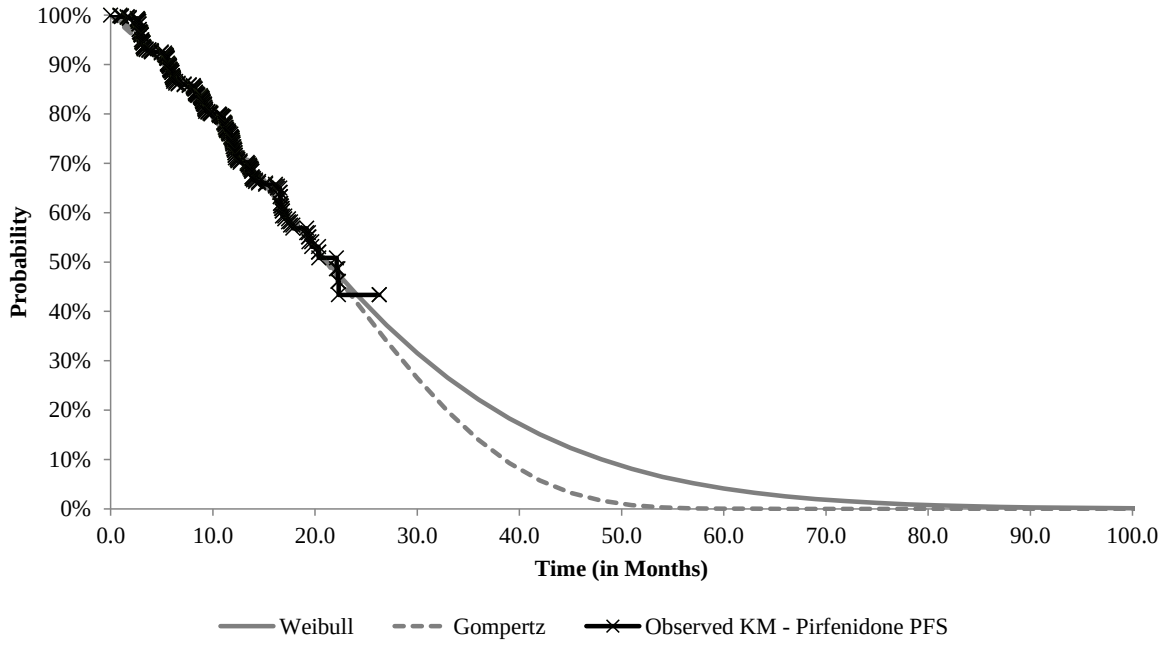
5.2.3.2. ERG's comments regarding the estimation of PFS

The ERG considers the definition of PFS used in the economic model to be largely appropriate. Discussion with clinical experts indicated that DLco is also considered to be clinically important but this is not as well accepted as a clinical trial endpoint (see Section 3) and was not included in the ASCEND trial.[34] Therefore, the ERG considered the definition used by the company based on the ASCEND trial[34] to be largely appropriate. Nevertheless, the ERG observes that PFS and its definition have only a minimal impact in the model and that the key driver of cost-effectiveness is OS.

The choice of the Weibull over the Gompertz distribution in the base-case is again questionable (Figure **39**). However, the impact on the ICER is minimal (increase from [redacted] per QALY gained using the Gompertz distribution – ITT population, against BSC), so any bias is likely to be small given the current model structure. However, the ERG notes that they would expect PFS to have a larger impact on the ICER if the relationship between disease progression, treatment discontinuation and treatment effect following discontinuation had been modelled in a more realistic manner.

The ERG further identified some inconsistencies in the approach used to estimate the hazard of progression for the subgroups. In response to a request for clarification from the ERG (see clarification response,[10] question B20), the company provided additional analyses and options in the economic model to use a more consistent methodology for PFS for the subgroups. Whilst the ERG expected the change to affect the subgroup analyses, the ERG is unclear why this also affects the results for the ITT population.

Figure 39: Comparison of the observed KM for PFS in people initiating pirfenidone against extrapolation using parametric distributions for the ITT population (Plot drawn by the ERG)

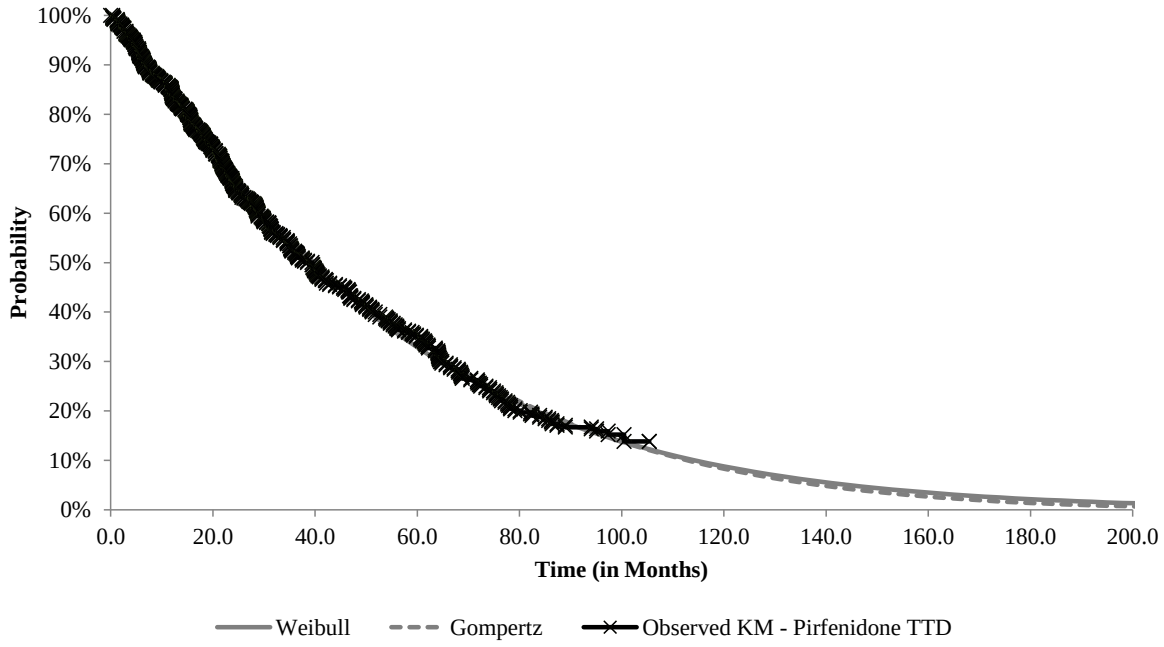


5.2.3.3. ERG's comments regarding the estimation of time to discontinuation

The ERG is satisfied with the approach used by the company to censor death and lung transplantation when estimating the time to discontinuation. However, the ERG notes that the censoring of lung transplantation may introduce bias as lung transplantation was not included in the base-case model structure; however, the impact is likely to be minimal given the small numbers discontinuing due to lung transplantation (see CS, Figures 4 to 6) in the ASCEND and CAPACITY trials.

The company's base-case uses the Weibull distribution for time to discontinuation based upon both visual and statistical goodness of fit to the observed portion of the pirfenidone curve. Alternative curve fits are explored as sensitivity analyses. The ERG considers that the choice between the Gompertz and Weibull distribution is questionable (Figure 40), but also that the impact on the ICER is again minimal (reduction in the ICER from [redacted] per QALY gained for ITT population for the comparison between pirfenidone versus BSC). However, as with PFS, the ERG would expect treatment discontinuation to have a larger impact on the ICER if the relationship between treatment discontinuation and treatment effect following discontinuation had been modelled in a more realistic manner. As detailed in Section 5.2.2.2, health outcomes are disconnected from costs, and therefore increasing the discontinuation rate leads to similar health outcomes at a lower costs and therefore an improved ICER for pirfenidone.

Figure 40: Comparison of the observed KM for discontinuation in people initiating pirfenidone against extrapolation using parametric distributions for the ITT population (Plot drawn by the ERG)



5.2.4. Treatment effects used in the company's base case for OS, PFS and time to discontinuation for pirfenidone vs. BSC and nintedanib

Treatment effects for pirfenidone against nintedanib and BSC are summarised in Table 47. The company's base-case analysis uses the treatment effects (HR) for pirfenidone against nintedanib and BSC (applied as inverse HR to the baseline pirfenidone curve) for the outcomes of OS and PFS reported in the Section 5.3 of the CS. The treatment effects are estimated from a random effects model which included all Phase II and Phase III trials using data up to 52 weeks (with the exception of SP2). Whilst only the OR are presented in the CS within the clinical section for the relative increase in discontinuation for nintedanib (compared with pirfenidone), ORs from the NMA are transformed into relative risks and used in the model subsequently. The company uses alternative models in scenario analyses including fixed effect models and data up to 72 weeks.

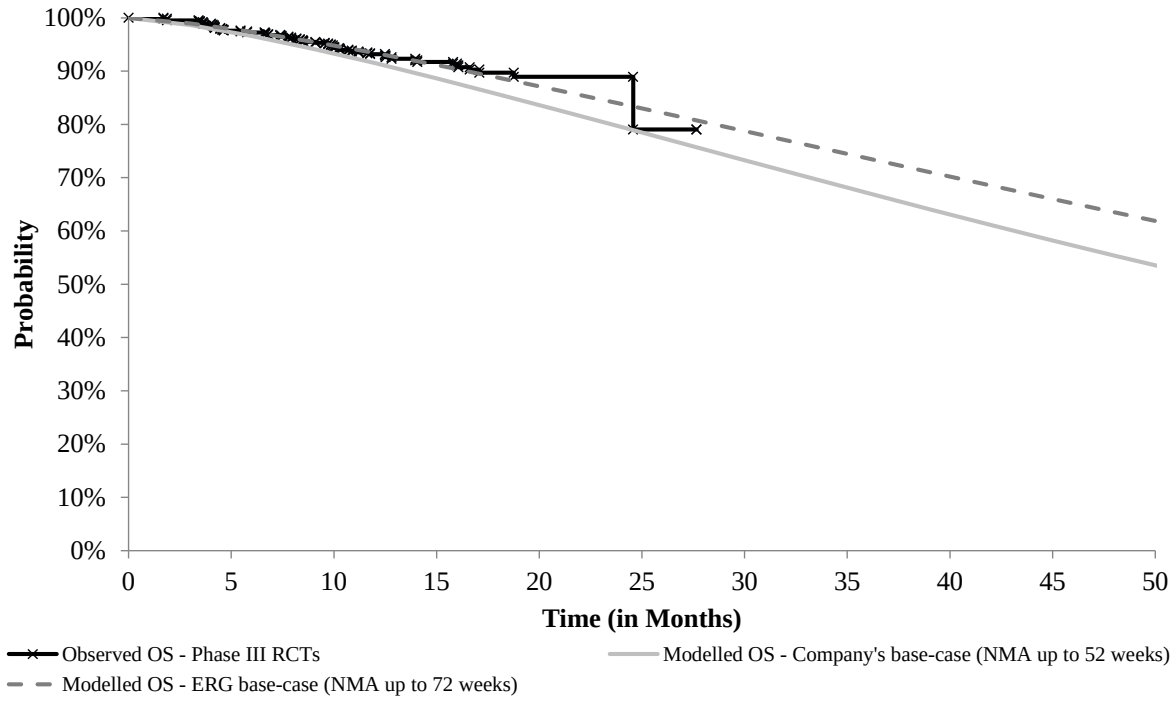
Table 47: Treatment effects used in the company's base-case

Treatment	Base-case HR (pirfenidone vs comparator)		Base-case RR (pirfenidone vs comparator)
	OS	**PFS**	**TTD**
Nintedanib	0.72	0.85	1.08
Best supportive care	0.52	0.63	NA

5.2.4.1. ERG's comments regarding the treatment effects used for OS in the company's model

As described in Sections 4.6 and 4.7, the CS reports results from a series of NMAs with varying strength and weaknesses, which are subsequently used in the company's model. In the company's base-case, the treatment effects (median) are estimated from a random effects model including all Phase II and III trials (referred to as the expanded network by the company) using data up to 52 weeks. As discussed in Section 4.7, the ERG considers that: (i) the treatment effects estimated using data up to 72 weeks are more appropriate and consistent with the company's assumption of proportional hazards; (ii) SP3 should be excluded from the base-case as this is a different (Japanese) population with a different dose and statistical adjustments were required as HRs were not reported, and; (iii) using the treatment effect at 52 weeks does not provide a reasonable fit to the observed KM for BSC (see Figure 41).

Figure 41: Fit of the modelled BSC using results from the NMA at 52 week (company's base-case) and 72 weeks (estimated by the ERG used in exploratory analyses)



The ERG further notes that the treatment effects taken from the NMA reported in the clinical section and subsequently used in the economic model, use posterior medians as point estimates, and associated 95% CrI. The ERG considers the use of the median in the economic model to be inappropriate and considers that the CODA samples (from the predictive distribution) should be used for the purpose of the modelling. As shown in section 6.1, using the median or mean point estimate could lead to inconsistent results.

5.2.4.2. ERG's comments regarding the treatment effects used for PFS in the company's model

The ERG notes that the treatment effect for PFS taken from the company's NMA using data up to 52 and 72 weeks are broadly the same between pirfenidone and BSC (HR: approximately 0.63) or nintedanib (HR: approx. 0.74), and therefore the company's decision to use the 52 week data instead of the 72 week data is unlikely to have a significant impact on the ICERs for pirfenidone. The ERG also notes that different definitions of PFS are used between trials included in the NMA. This is acknowledged in the CS (page 143). The ERG considers that this is likely to introduce some biases between pirfenidone and nintedanib but reiterates that PFS has a minimal impact on the ICER in the company's model. Although, as stated previously, it is expected that it would have a greater impact if progression was linked to treatment discontinuation and treatment effects were allowed to differ after discontinuation. Finally, as described above, the ERG considers that the CODA samples should be used in the model.

5.2.4.3. ERG's comments regarding the relative difference in treatment discontinuation in the company's model for pirfenidone and nintedanib

Although unclear from the CS, the OR estimated from the NMA were transformed into RR. The ERG notes that the relative risk for discontinuation (RR) is calculated for discontinuations for any reason but is applied in the model to people who discontinued treatment from reasons other than death and lung transplants. This may introduce bias if the rates of death or lung transplant differ between the trial arms compared in the NMA. The ERG further considers that the CODA samples should be used in the model.

5.2.5. Inclusion of costs associated with IPF-related mortality

The company's original base-case analysis assumes that 57.89% of deaths occurring in people initiating pirfenidone are IPF-related, based on the data from the CAPACITY/ASCEND trial (see CS, page 214), with the remaining deaths occurring due to causes unrelated to IPF. The proportion of deaths related to IPF in people initiating BSC and nintedanib was assumed to be greater compared with people initiating pirfenidone (72.22% and 68.57%, respectively). These figures were reported as being derived by applying data from the NMA, although the CS does not describe exactly how this was done.

In response to a request for clarification (question B10) regarding the source of the estimates for the proportion of deaths which are IPF-related, the company amended the methodology in the revised economic model by using the proportion of observed deaths that are IPF-related for each treatment, according to the company at 52 weeks from their respective trials (Table 48). The ERG observed that compared with the statement from the company, data at 72 weeks from the CAPACITY trials are used.

Table 48: Revised IPF-related mortality figures (reproduced from the clarification response,[10] question B10)

Intervention	Time point	n of IPF-related deaths	N of all-cause deaths	Proportion of death IPF-related
Pirfenidone	52 weeks*	17	32	53.13%
Placebo		35	50	70.00%
Nintedanib		26	42	61.90%
* contrary to the statement from the company data at week 72 from the CAPACITY trials are used				

A one-off cost of £9,996 is assigned in the model only for deaths attributable to IPF. This cost was taken from estimates provided in a report from the National Audit Office (2008)[101] and inflated to 2014 prices. This data source was also used in the nintedanib submission.[26] No costs are applied to deaths that are unrelated to IPF. Little justification is provided in the CS with the exception of the following statement "*costs associated with IPF are greatly increased in the last year of life due to increased resource use, home care and length of stay in hospital*" (see CS,[4] page 2014).

5.2.5.1. ERG's comments on the impact of IPF related-mortality on costs

The ERG has a number of concerns with the approach used by the company which included costs associated with end of life only in people dying from IPF-related causes.

The ERG considers the approach used by the company to estimate the proportion of death that are IPF-related in the revised economic model to be questionable and that ideally results from the NMA should be used. The ERG notes that whilst an NMA was used in the original submission to NICE, it was unclear on how this was done. The baseline proportion of death (whilst on BSC) was also unclear in the original submission. As a result, both of the approaches presented by the company are considered questionable.

No details are included within the CS on the source used to represent the costs associated with death from IPF-related causes. In brief, Hatziandreu *et al.* (2008)[101] is a modelling study which aimed to estimate the total costs of care provided to people in their last year of life for both cancer and organ failure (heart and pulmonary disease) in three settings; hospital, hospice and home. As with any modelling study, this economic analysis relied on a series of assumptions. Due to time and resource constraints, the ERG is not able to provide a complete assessment of this study. Nevertheless, the ERG notes that the company assumes the costs associated with end of life care in the last year of life in people dying from organ failure to be a reasonable approximation of the cost of care provided in the last year of life for people dying of causes related to IPF.

The ERG considers that deaths from causes other than IPF are also likely to be associated with costs that fall on the NHS. The exclusion of costs associated with death from other causes is likely to bias the cost-effectiveness estimate in favour of pirfenidone given the lower IPF-related mortality assumed in the company's model. The ERG asked the company to provide evidence which demonstrates that the costs in the last year of life for IPF-related deaths are higher than the costs in the last year of life for deaths from other causes (see clarification response,[10] question B14). In response, the company stated: "*The purpose of the submission was to assess the cost and clinical implications associated with pirfenidone for the treatment of IPF compared with current care. As a result of this, only costs borne by the condition have been considered in the analysis, with other costs deemed out of the scope of this analysis and unrelated to the decision problem. Consequently for end of life care, costs attributable to death from other causes than IPF are not included.*" The ERG considers the response from the company to be unsatisfactory and considers that the assumption of zero costs for people with IPF who die from other causes to be inadequately supported by evidence. The ERG further notes that Hatziandreu *et al.* (2008) states that "*It is undisputable that patients with organ failure who are at the end of their life have palliative care needs as severe and distressing as those with cancer. Patients suffering from other conditions of equal importance in terms of prevalence, and economic burden, such as dementia or renal failure are also subject to end of life care services*."[101]

Consequently, in the absence of evidence relating to the differential impact of deaths on resource use, the ERG considers that the cost associated with end of life should be applied to all deaths irrespective of the cause.

5.2.6. Incorporation of the impact of acute exacerbations on costs and quality of life in the company's model

As described in Section 5.2.2, the company's model includes the impact of acute exacerbations as a cost and HRQoL decrement during each model cycle which is applied according to the treatment

received (pirfenidone, BSC and nintedanib). Table 49 summarises the costs and utility decrements per model cycle assumed in the company's base-case in people receiving pirfenidone, BSC and nintedanib.

Table 49: Management costs and utility decrements (per model cycle) associated with acute exacerbations assumed in the company's model

	Pirfenidone	**Best supportive care**	**Nintedanib**
Decrement in utilities assumed per model cycle	- 0.00103	- 0.00165	- 0.00091
Management costs assumed per model cycle	£114.08	£226.34	£114.08

The utilities decrements associated with acute exacerbations are calculated from three components:
The incidence of acute exacerbations per model-cycle (3 months) by treatment. The rate for people treated with BSC is taken from the nintedanib submission.[26] The incidence is then adjusted in individuals receiving pirfenidone and nintedanib using ORs estimated from the NMA (see Table 50).

(i) The decrement in utilities associated with an acute exacerbation. The decrement is also taken from the nintedanib submission[26] and is calculated as a weighted average between the decrement in the first and subsequent months (see Table 51).

(ii) Assumptions regarding the duration over which HRQoL is reduced due to an exacerbation. This duration is assumed to be 3 months.

Table 50: Incidence of acute exacerbations (adapted from the CS, page 219, Table 80)

Treatment	Base-case OR vs BSC	Incidence of acute exacerbation assumed in the company's model
BSC	Baseline risk of 1.46% per model cycle	
PFN	0.62	0.91%
NTB	0.55	0.81%

Key: BSC, best supportive care; comp, comparator; NTB, nintedanib; OR, odds ratio; PFN, pirfenidone.

Table 51: Decrement in utility associated with an acute exacerbation (reproduced from CS,[4] Table 90, page 233)

Time frame	Utility [SE]	Reference
First month	-0.274 [0.059]	Nintedanib NICE company submission[26]
Subsequent months	-0.033 [0.053]	
Per model cycle	-0.113	First month + 2 * Subsequent months

Key: Dist, distribution; N/A, not applicable; NICE, National Institute for Health and Care Excellence; SE, standard error.

In contrast, the costs associated with the management of acute exacerbations per model cycle (see Table 52), are calculated from two components:

(i) The probability and duration of hospitalisations in people initiating BSC and pirfenidone. These are calculated using data from the CAPACITY trials.

(ii) The average cost associated with a hospital bed day. This is taken from the NHS Reference Costs for hospitalisations due to respiratory failure (HRG code: DZ27S, DZ27T, DZ27U) and is assumed to be £266.71 per day.

In the absence of comparable data, people treated with nintedanib are assumed to incur the same hospitalisation costs as people treated with pirfenidone.

Table 52: Calculation of hospitalisation cost (reproduced from CS,[4] Table 99, page 243)

	PFN	BSC
Number of cycle-length intervals observed [a]	3768	3771
Number of subjects with hospitalisation [b]	195	202
Rate of hospitalisation per cycle [c = a/b]	0.052	0.054
Probability of hospitalisation per cycle [d = 1-exp[c]]	0.050	0.052
Average length of stay in hospital [e]	8.48	16.27
Total cost of hospitalisation [f = e * cost of bed day]	£2,261.70	£4,339.37
Hospitalisation cost applied per cycle [g = d * f]	£114.08	£226.34

5.2.6.1. ERG's comments regarding the estimation of the costs and decrement in utilities associated with acute exacerbations

As described in Section 5.2.2.1, the inclusion of acute exacerbations (as implemented by the company in the economic model) has a minimal impact on the ICER, and therefore, only a brief critique is presented here. The ERG considered the lack of impact of acute exacerbations in the economic model to be an artefact of the model structure chosen by the company and not a reflection of the relevance of exacerbations in IPF.

The ERG also notes that the approach used by the company to include the impact of acute exacerbations and costs and HRQoL is inconsistent as different trial outcomes are used to estimate the impact of acute exacerbations on costs and QALYs. As a consequence of this inconsistency, pirfenidone is assumed to have a greater decrement in HRQoL associated with acute exacerbations compared with nintedanib during each model cycle, whilst the per cycle cost is assumed to be identical. The ERG considers the inputs for the decrement in utilities and costs for acute exacerbations to be broadly reasonable and notes that the majority of inputs are taken from the nintedanib submission and have a limited impact on the ICER.

The ERG further notes that data on hospitalisations used to represent the costs associated with acute exacerbations are not specific to hospitalisations due to acute exacerbations. The ERG sought clarification from the company on the inconsistencies in the approach to include the impact of acute exacerbations on costs and HRQoL (see clarification response,[10] question B15). In response, the company confirmed that data include hospitalisations from any causes and therefore the hospitalisation costs is broader than just the cost associated with acute exacerbations. The CS provides limited detail on the data used to inform the incidence and length of hospital stay. The ERG is unclear whether data on only respiratory-related, IPF-related or hospitalisations from any cause were used.

5.2.7. Incorporation of the impact of AEs on costs and HRQoL in the company's model

The impact of AEs on costs and HRQoL is applied during each model cycle according to the treatment currently received.

Table 53 summarises the costs and utility decrements per model cycle assumed in the company's revised base-case model in people receiving pirfenidone, BSC and nintedanib. In response to the ERG's clarification requests (see clarification response,[10] question A39 and the summary of model changes on page 38), the company amended the calculation of the incidence of AEs which led to a number of new errors being introduced into the model (described below). As can be seen from

Table 53, in the revised economic model the errors led to people on BSC experiencing greater costs and QALY impacts compared with people on pirfenidone or nintedanib.

Table 53: Costs and utility decrements (per model cycle) associated with AEs when averaged across the treated cohort (as applied in the revised company model) - prior to correction of errors by the ERG

	Pirfenidone	**Best supportive**	**Nintedanib**
Management costs	£93.79	£109.47	£32.18
Decrement in utility	-0.0040	-0.0052	-0.0015

The decrements in utilities and management costs associated with AEs per model cycle are calculated from:

(i) The incidence of AEs by treatment
(ii) The management costs associated with each AE
(iii) The utility decrement associated with AE
(iv) The assumed duration of the decrement for each AE.

Inputs and assumptions for each AE are summarised in Table 54. The company considered the same AEs included in the nintedanib submission;[26] namely: serious cardiac events, serious gastrointestinal event (which is subsequently replaced by the company by diarrhoea in the revised economic model), gastrointestinal perforation (nintedanib only), photosensitivity reaction (pirfenidone only) and rash (pirfenidone only in the original economic model and both pirfenidone and BSC in the revised economic model).

Table 54: Incidence, costs and decrement in utilities associated with each AE (per individual experiencing the event) included in the revised company model (prior to correction of errors by the ERG)

	Incidence			Decrement in utilities		Costs
	PFN	**BSC**	**NTB**	**Disutilities**	**Duration**	
Serious cardiac event	0.79%	1.05%	0.23%	-0.198	3 months	£2,200.15
Diarrhea	3.48%	4.52%	1.36%	-0.068		£1,910.91
Gastrointestinal perforation	-	-	0.08%	-0.118		£1,583.03
Photosensitivity reaction	2.32%	-	-	-0.032	15 days	£467.62
Rash	0.00%	1.82%	0	-0.03		£428.63

The incidence of AEs for rash, serious gastrointestinal events (assumed to be diarrhoea in the revised economic model) and serious cardiac events for BSC are calculated based on the average incidence for the placebo arm across the Phase III trials. Relative risks from the NMA submitted as part of the company's clarification response (see clarification response,[10] question A39) are then used to derive the incidence in people treated with pirfenidone and nintedanib. The incidence of AEs for gastrointestinal perforation and photosensitivity were taken from the nintedanib submission.[26]

The utility decrement associated with serious cardiac events, gastrointestinal events (replaced by diarrhoea in the revised economic model) and perforation were taken from the nintedanib submission[26] and are assumed to last 3 months. The utilities decrement associated with photosensitivity and rash were taken from Handorf *et al.* (2012)[102] and the NICE Centre for Clinical Practice, respectively, and are assumed to last 15 days. Costs associated with the management of AEs were taken from the NHS Reference Costs (2014-15)[103] using a similar approach to that used in the nintedanib submission.[26]

5.2.7.1. ERG comments on the inclusion of AEs in the economic model

As with acute exacerbations, the ERG notes that the inclusion of AEs (as implemented by the company in the economic model) has a minimal impact on the ICER, and therefore, only a brief critique is provided here.

The company revised their estimates of the incidence of AEs using results from the NMA in response to clarification on a separate issue (see clarification response,[10] question A39). The ERG considers this revised approach to be more appropriate. Nevertheless, the ERG identified a series of errors in the implementation of this within the company's revised model. The ERG notes that results from the NMA suggest that pirfenidone has a greater incidence of AEs (serious cardiac events, rash, diarrhoea) compared with placebo. However, in the economic model, the incidence of AEs used for pirfenidone is lower compared with BSC. The ERG notes that this is because the RRs from the NMA are applied incorrectly in the model. Furthermore, the ERG notes some discrepancies between results from the NMA reported in the clarification response (see clarification response,[10] question A39) and the data from the NMA used in the economic model.

The ERG further notes that in the original economic model, the company included serious gastrointestinal events; which was subsequently replaced by diarrhoea in the revised economic model. The ERG considers the inclusion of diarrhoea to be appropriate. However, whilst data on the incidence of diarrhoea appear to be used, the costs and utility decrements associated with serious gastrointestinal events are still used. If the costs and utility decrements associated with diarrhoea are lower than those associated with serious gastrointestinal events, then the approach used by the company is likely to be unfavourable to nintedanib which is considered to be associated with a greater incidence of diarrhoea compared with BSC and pirfenidone.

5.2.8. HRQoL

Table 55 summarises the health state utility values assumed within the company's model.

Table 55: Summary of health state utility values used in the company's model for the base-case and sensitivity analyses

Health state	Pirfenidone (Base-case)	Pirfenidone Trial (Alternative mapping)[95]	Panther & ACE	Ofev STA[26]
Progression-Free	0.847	0.791	0.82	0.777
Progression	0.7818	0.744	0.74	0.744
Transplant	Assumed to be the same as progression-free			

The CS includes details of a systematic review of studies which provide estimates of HRQoL for adult people with mild to moderate IPF (see CS, Section 5.4). As described in Section 5.1.1, the company undertook a single search in November 2015 to identify cost-effectiveness studies, HRQoL studies and resource use data. A total of 22 references were included in the HRQoL review, of which 5 references

(corresponding to 4 studies) reported EQ-5D data which are briefly described within the CS (CS, pages 228 to 230) with the remaining studies described in the CS Appendix 22. This included EQ-5D data collected in Richeldi *et al.* (2014)[72] and used in the 2015 nintedanib submission,[26] EQ-5D (measure using time trade off) data from a registry,[104] SF-36 and EQ-5D data from a RCT comparing bosentan and placebo,[105] and EQ-5D data from a RCT comparing sildenafil and placebo.[96]

In the base-case, the company obtained EQ-5D utility scores for the progression-free and progressed health states based on the mean SGRQ score collected in the CAPACITY trials for each health state. The mean SGRQ in people who are progression-free (37.31) and progressed (42.40) was estimated from the CAPACITY trials[49] using a generalised estimating equation (GEE) regression model to account for the correlation between measurements from the same individual at different time points. These mean scores were then mapped onto the EQ-5D using a mapping algorithm published by Freemantle *et al.* (2015) which was developed specifically for people with IPF.[106] A scenario analysis is also presented using a mapping algorithm published by Starkie et al;[95] this algorithm was estimated in people with Chronic Obstructive Pulmonary Disease (COPD).

Alternative utility values, identified from the company's systematic review, are used in additional scenario analyses. For the sensitivity analysis including lung transplantation, the utility value associated with lung transplantation was assumed to be similar to the value for the progression-free health state. The limitation of making such an assumption is acknowledged by the company.

5.2.8.1. ERG's comments regarding the estimation of health state quality of life

Limited detail is included on the methods for the systematic review of HRQoL data. Whilst the search strategies and list of excluded studies are provided in the CS Appendices 17 and 18, no information is provided on the outcomes that were eligible for this review, thereby making it difficult to assess why some studies were excluded. The ERG notes that the link between the systematic review of HRQoL literature and the evidence selected for use in the model is unclear. For example, the mapping study used to map from the SGRQ data from the CAPACITY trials to the EQ-5D[106] was excluded from the systematic review as it was deemed to include an ineligible patient population. However, the study by Freemantle was conducted in people with IPF (as described on page 225 of the CS) and therefore the reason for exclusion is inconsistent. The study by Nathan *et al.* (2015) was also excluded (see CS,[4] Appendix 18) due to the inclusion of an "ineligible population", however the population relates to people with diagnosed IPF and therefore seems relevant. The company also did not identify the study by Raghu *et al.* (2013)[75] which reported EQ-5D data. This study was reported in Loveman *et al.* (2014)[93, 94] which was included in the review of cost-effectiveness studies but appears to have been excluded from the HRQoL review. Therefore the ERG cannot be certain that all relevant HRQoL data have been identified and presented in the CS.

In the company base-case, EQ-5D utility scores are estimated from the trial. The ERG notes that the mean SGRQs are transformed into an EQ-5D score based on a linear mapping function published by Freemantle *et al.*[106] The ERG considers the general method of using aggregate data for the mean SGRQ to be appropriate in principle given the use of a linear mapping function. Nevertheless, the ERG notes that the mapping algorithm is unconstrained and therefore could predict values greater than 1.0 when the SGRQ is below 26 when applied at the individual level. In response to a request for clarification from the ERG (see clarification response,[10] question B18a), the company reported that 25.5% of individuals at baseline and 19.5% based upon the last observation had a SGRQ score below 25. The ERG considers that given the large proportion of people at the individual level with a SGRQ score below 25, utility values estimated at the aggregate level could be biased and that utility values should be estimated at the individual level and capped at 1.0. Within the company's clarification response (see clarification response,[10] question B18c), these estimates were provided by the company and led to a decrease in utility value in people without progression (from 0.8485 to 0.8185) and people who progressed (from 0.7835 to 0.7597). The impact is examined by the ERG in Section 6.

The company further assumes that HRQoL is constant within the respective health states. The ERG considers that HRQoL is likely to reduce over time within each health state, or at least, within the progressed disease health state given the progressive nature of the condition. However, as described, given the model structure chosen by the company, including a HRQoL decrement for further disease progression in the progressed state is challenging. The ERG considers this implication to likely overestimate the number of QALYs but the size and direction of the effect of incremental QALYs and therefore the ICER is unclear.

The ERG observes that alternative utility values used in scenario analyses have a moderate impact on the ICER. The ERG considers this to be an artefact of the model structure chosen by the company in that HRQoL is assumed to remain constant within health states and not a reflection of the impact of HRQoL in IPF. The ERG further observes that no adjustment for utility by age is assumed and therefore, people are assumed to maintain the same level of HRQoL irrespective of age. The ERG considers this implication to likely overestimate the number of QALYs and potentially the incremental QALY gain associated with pirfenidone. This is examined by the ERG in Section 6.

5.2.9. Drug acquisition costs and resource use associated with the management of IPF

This section discusses the drug acquisition and administration costs and the resources use and costs associated with the management of IPF included in the company's model.

Drug acquisition costs

Before accounting for dose reductions and interruptions, the company's model assumes that people treated with pirfenidone receive a total of nine 267mg capsules per day, at a daily cost of £71.70 (or a cost per 3-month model cycle of £5,730.62) using the list price.

The company calculates that in the CAPACITY and ASCEND trials,[34, 49] a fewer number of pills were given daily, with an average of 7.88 capsules per day. This is used in the company's base-case, leading to a daily cost of £62.80 (or a cost per 3-month model cycle of £5,730.62) using the list price.

The daily cost in people treated with nintedanib is assumed to be at parity with the daily cost calculated for pirfenidone. This is justified on the basis that this assumption was made in the nintedanib submission.[26]

No drug acquisition costs are assumed for people receiving BSC. This is justified by the company on the basis that BSC represents the placebo arm of the trials from which efficacy data were derived. Similarly, no concomitant medications are assumed in the company's model.

Administration costs

No administration costs were assumed for either treatment as both nintedanib and pirfenidone are taken orally.

Costs associated with the management of IPF and lung transplant

The CS presents the methods and results of a systematic review of studies with the aim to identify data on resource use and costs for adult patients with mild to moderate IPF. As described in Section 5.1.1, a single search strategy was conducted in November 2015 to identify cost-effectiveness studies, HRQoL studies and resource use data. A total of 7 references were included in its review for resource use, of which 2 were cohort studies,[107, 108] one was an economic evaluation alongside an RCT in patients with IPF,[109] one was a cost-effectiveness analysis in patients with cystic fibrosis[110] and three were technology appraisals by national bodies (NICE and the SMC).[111,42, 112]

The company separates treatment-specific monitoring costs (resource use at initiation and liver function tests) and the costs associated with the progression status. These are summarised in Table 56.

Table 56: Cost of resource use per model cycle (reproduced from CS,[4] Table 96)

Cost type		**PFN**	**BSC**	**NTB***
Cycle specific costs	Cycle 1	£969.38	£964.71	£969.38
	Cycle 2	£5.61	£0.94	£5.61
	Cycle ≥ 3	£1.87	£0.56	£1.87
Progression status specific costs	Pre-progression	£513.22	£513.22	£513.22
	Post-progression	£525.44	£525.44	£525.44

Key: BSC, best supportive care; NTB, nintedanib; PFN, pirfenidone.

* only for people with moderate IPF

Cycle : 3 month

Resource use by progression status, which is not specific to the treatment received is summarised in Table 57. This was estimated from discussion with a panel of UK clinicians (see CS, page 239). As shown in Table **57**, the difference in resource use between health states is assumed to be dependent on the individual's percent predicted FVC for oxygen, healthcare professional visits and GP visits. Therefore, the only difference in management costs between health states is due to the differences in FVC between people who are progression-free and those with progression. The frequency of liver function tests (that are specific to treatments) was taken from the pirfenidone and nintedanib SmPC.

The cost associated with lung transplantation (which is used only in a sensitivity analysis) was taken as the average cost of lung/heart transplant reported in a report published by the NHS Blood and Transplant (2013) uplifted to 2014/2015.

Table 57: Resource use assumed in the company's model (based on CS, Tables 94 and 95)

Resource use item	At treatmen	Subsequent MRU	Unit cost
Liver function test	TRUE	*	£1.87
Gas transfer	TRUE	every 4 months	£202.08
Lung volume study	TRUE	None	£170.54
Full pulmonary (covers	TRUE	every 4 months	£165.85
Field exercise test	TRUE	every 6 months	£177.13
Oxygen	FALSE	for all patients with <80% FVC	£206.08
Healthcare professional visit	TRUE	every 4 months if FVC >60%, every 3 months if FVC<60%	£248.17 £177.53
GP visit	FALSE	based upon FVC	£37.00
Key: FVC, forced vital capacity; GP, general practitioner; MRU, medical resource use; SmPC, summary of product			
** Liver function tests were administered as per the pirfenidone SmPC for pirfenidone and nintedanib patients (every month for the first 6 months of treatment, then every 3 months), and for BSC are administered according to clinician opinion (every 1.5 months for the first 6 months of treatment, then 0.3 times per model cycle).*			

5.2.9.1. ERG's comments

As with the HRQoL searches, few details are included in the CS regarding the methods for the systematic review of resource use. Whilst the search strategies and list of excluded studies are provided in the CS (Appendices 17 and 18), no information is provided on the outcomes that were eligible for this review, thereby making it difficult to assess why certain studies were excluded. The ERG notes that the link between the systematic review of resource use and the evidence selected for use in the model is also unclear. For example, the CS states that the estimates of unit costs from the nintedanib submission are preferred by the company over those identified from the literature (see CS, page 237). However, the ERG notes that the nintedanib submission is not included in the company's resource use review. The CS also does not present whether the costs are consistent across the different sources, thus it is unclear whether the data in the nintedanib submission reflect the data from other published sources. The ERG further notes that the company identified a study reporting the impact of pirfenidone in a real-world setting through the UK Named Patient Programme using a retrospective study design. Findings from this study are not used or compared with the resource used in the model. The study by Loveman et al[93, 94] which was included in the review of cost-effectiveness studies, is also not included in the company's resource use review. The ERG considers this to be inappropriate

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 IMPULSIS 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

not been reported in a sufficiently transparent manner. It is also unclear how any potentially divergent views between clinicians were accounted for.

Contrary to the statement from the company, oxygen was not included in the company's original model; this was amended in the revised model. However, the ERG notes that resources use by progression status (notably oxygen, healthcare professional visits and GP visits) are driven by the level of percent predicted FVC. The ERG observes that given the structure chosen by the company, which is based on progression status and not percent predicted FVC level, the implementation of these percent predicted FVC dependent costs in the model relies on a series of assumptions. Notably, the company uses the percent predicted FVC distribution at baseline (divided into 10% bands) to represent the distribution of percent predicted FVC in people without progression and assumes a shift of one band in percent predicted FVC in people with progressive disease. This is arbitrary and an artefact of the chosen model structure.

The ERG further observes some double-counting in the first cycle for the costs associated with the management of the condition. However, as this has been done consistently in the pirfenidone, BSC and nintedanib arms, the effect is cancelled out across the treatment arms and therefore has no impact on the ICER.

In addition, as with the modelling of HRQoL by health state, the company assumes that resource use is constant within the respective health states. The ERG considers that resource use is likely to increase over time within each health states, or at least, within the progressed disease health state given the progressive nature of the condition. The ERG considers this implication to likely underestimate the management costs associated with IPF. The size of the effect and direction of the ICER is unclear.

5.2.10. Summary of data used for the subgroups of people with a percent predicted FVC of 50% to 80% and >80%

The only data that are subgroup-specific are the baseline OS, PFS and discontinuation curves. For the subgroup analyses (percent predicted FVC of 50 – 80% or >80% at baseline), parametric functions were fitted to the observed KM using percent predicted FVC as covariates (percent predicted FVC<50%, 50%≤FVC<80%, FVC≥80%) for both OS and time on treatment. However, a different approach was taken for PFS. In this case, the KM data for each subgroup were used separately to fit the parametric distributions. Whilst unclear from the CS, the Weibull distribution was also selected for the subgroups of people with a percent predicted FVC of 50 to 80% or >80. The other model parameters applied in the subgroup analyses were the same as those applied in the ITT analysis.

5.2.10.1. ERG's comments

The company uses data by subgroup to derive the baseline hazards of death, discontinuation and progression in people initiating pirfenidone. The ERG considers this to be appropriate. As described in Section 5.2.3.1, the ERG considers the Gompertz distribution to provide a more plausible long-term extrapolation for OS compared with the Weibull distribution in both the ITT population and in the subgroups of people with a percent predicted FVC of 50% to 80% and >80% (see [redacted] and [redacted]).

The ERG further identified that a different approach was used for PFS compared with OS and time to discontinuation for the subgroup analyses. In their clarification response (see clarification response,[10] question B20), the company acknowledged the inconsistency and stated that using a consistent approach with OS and time to discontinuation had a minimal impact on the ICER (increase from [redacted] per QALY gained). Nevertheless, the ERG identified an error when the Gompertz and gamma distribution was used for PFS; this is corrected in the ERG's exploratory analyses (see Section 6).

The treatment effects from the ITT population are used to represent the treatment effect for the subgroup. The ERG considers this to be appropriate given the lack of stratification by subgroup in the trial, and the *post hoc* nature of the subgroup analysis. The ERG notes that approximately 75% of people enrolled in the trial had a percent predicted FVC ≤80%. The ERG notes that *post hoc* analyses provided by the company in response to clarification questions (question A31) showed numerically different treatment effects for OS, although it is unclear if the differences are real. The ERG further notes some apparent typographical errors in some of the values reported in Table 23 for OS in people with a percent predicted FVC > 80% from the ASCEND trial. As it is not possible to rule out with certainty a different treatment effect by subgroup, the ERG's exploratory analyses examine the impact of using the treatment effects by subgroups from the *post hoc* analyses.

The company further assumed that the impact of acute exacerbations, IPF-related mortality and AEs are the same between the subgroups of people with a percent predicted FVC of 50% to 80% and >80%. As described in Sections 5.2.5 to 5.2.7, the impact of different assumptions on the ICER is minimal and thus, the ERG is satisfied with using the same data by subgroup for the purpose of the model but highlights that the impact on costs and HRQoL could be different according to the subgroup examined.

Finally, the company assumed no differences in HRQoL and resource use for the progression-free and progression states in people with a percent predicted FVC of 50% to 80% and >80%. The ERG considers this to be inappropriate given that both HRQoL and resource use are a function of percent predicted FVC for the progression-free and progressed states. The impact on the ICER of making this assumption is uncertain given the chosen structure as progression is not modelled.

5.2.11. Cost-effectiveness results

Results for the ITT population and the subgroup of people with a percent predicted FVC >80% are presented in this report both using the list price and with the PAS for pirfenidone. Results for the subgroup of people with a percent predicted FVC of 50 – 80% are presented using the list price only in this report with results using the PAS for pirfenidone and the PAS for nintedanib available in a confidential appendix. Whenever possible, results reported here are taken either from the results provided at the clarification stage following amendments made by the company, or in the case of results incorporating the PAS, from the PAS submission template. On some occasions results had to be re-run by the ERG where there existed discrepancies between the model and values reported by the company. These are highlighted as being taken from ERG analysis. Finally, it should be noted that the base-case results presentd by the company and reproduced here exclude the stopping rule for pirfenidone, but include a stopping rule for nintedanib as this was the assumption used in the company's base-case.

- **ITT population – Mild to Moderate IPF**

Table 58 summarises the estimated health gains and costs for each strategy for the ITT population. Pirfenidone is estimated to result in an additional 1.87 QALYs at an incremental cost of [redacted] (using the list price) compared with BSC, over a life-time horizon. This corresponds to an ICER for pirfenidone versus BSC of [redacted] per QALY gained. It can be seen from Table 59 that when applying the PAS for pirfenidone, the incremental cost for pirfenidone versus BSC is £40,010 and the ICER is £21,387.

Table 58: Central estimates (based on point estimates of parameters) of cost-effectiveness for the ITT population – discounted results (list price)

	Costs	LY	QALYs	Inc Costs	Inc LY		Inc QALYs	ICER
BSC	[redacted]	5.38	3.80					
pirfenidone	[redacted]	8.67	5.67	[redacted]	3.29		1.87	[redacted]

BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted

Table 59 Summary of definitions used for progression-free survival

	Costs	LY	QALYs	Inc Costs	Inc LY	Inc QALYs	ICER
BSC	£26,627	5.38	3.80				
pirfenidone	£66,638	8.67	5.67	£40,010	3.29	1.87	£21,387

BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

- **People with a percent predicted FVC > 80% at baseline**

Table 60 summarises the estimated health gains and costs for each strategy in people with a percent predicted FVC > 80% at baseline. Pirfenidone is estimated to generate an additional 2.17 QALYs at an incremental cost of [redacted] (using the list price) compared with BSC, over a life-time horizon. This corresponds to an ICER for pirfenidone versus BSC of [redacted] per QALY gained.

It can be seen from Table 61 that when the PAS for pirfenidone is applied, the incremental costs for pirfenidone versus BSC are £52,480 and the ICER is £24,187.

Table 60: Central estimates (based on point estimates of parameters) of cost-effectiveness in people with a percent predicted FVC > 80% – discounted results (list price)

	Costs	LY	QALYs	Inc Costs	Inc LY	Inc QALYs	ICER
BSC	[redacted]	7.11	4.82				
Pirfenidone	[redacted]	11.26	6.99	[redacted]	4.15	2.17	[redacted]

Table 61: Central estimates (based on point estimates of parameters) of cost-effectiveness in people with a percent predicted FVC > 80% – discounted results (PAS price)

	Costs	LY	QALYs	Inc Costs	Inc LY	Inc QALYs	ICER
BSC	£31,729	7.11	4.82				
Pirfenidone	£84,209	11.26	6.99	£52,480	4.15	2.17	£24,187

- **People with a percent predicted FVC of 50 - 80% at baseline**

Table 62 summarises the results for the subgroup of people with a percent predicted FVC of 50 - 80% at baseline. The company's model suggests that, at the list price, pirfenidone is the most effective and most expensive option and BSC is the least effective and least expensive option. Based on an incremental analysis of the three options, nintedanib is expected to be ruled out due to extended dominance at the list price. When compared with BSC, pirfenidone is estimated to generate an additional 1.696 QALYs at an incremental cost of [redacted]. The corresponding ICER for pirfenidone versus BSC is estimated to be [redacted] per QALY gained at the list price. When applying the pirfenidone PAS, the ICER for pirfenidone versus BSC was £21,318. The results for pirfenidone versus nintedanib are in the confidential appendix.

Table 62: Central estimates (based on point estimates of parameters) of cost-effectiveness in people with a percent predicted FVC of 50 - 80% – discounted results (list price)

	Costs	QALYs	Incremental results versus BSC			ICER (incremental analysis)
			Costs	QALYs	ICER	
BSC	[redacted]	3.443	-	-	-	-
Nintedanib	£65,065	4.226	£40,197	0.783	£51,331	Extendedly dominated
Pirfenidone	[redacted]	5.138	[redacted]	1.696	[redacted]	[redacted]

Table 63: Central estimates (based on point estimates of parameters) of cost-effectiveness in people with a percent predicted FVC of 50 - 80% – discounted results (with PAS)

	Costs	QALYs	Incremental results versus BSC			ICER (incremental analysis)
			Costs	QALYs	ICER	
BSC	£24,868	3.44	-	-	-	-
Nintedanib	See confidential appendix					Extendedly dominated
Pirfenidone	£61,012	5.14	£36,145	1.70	£21,318	£21,318

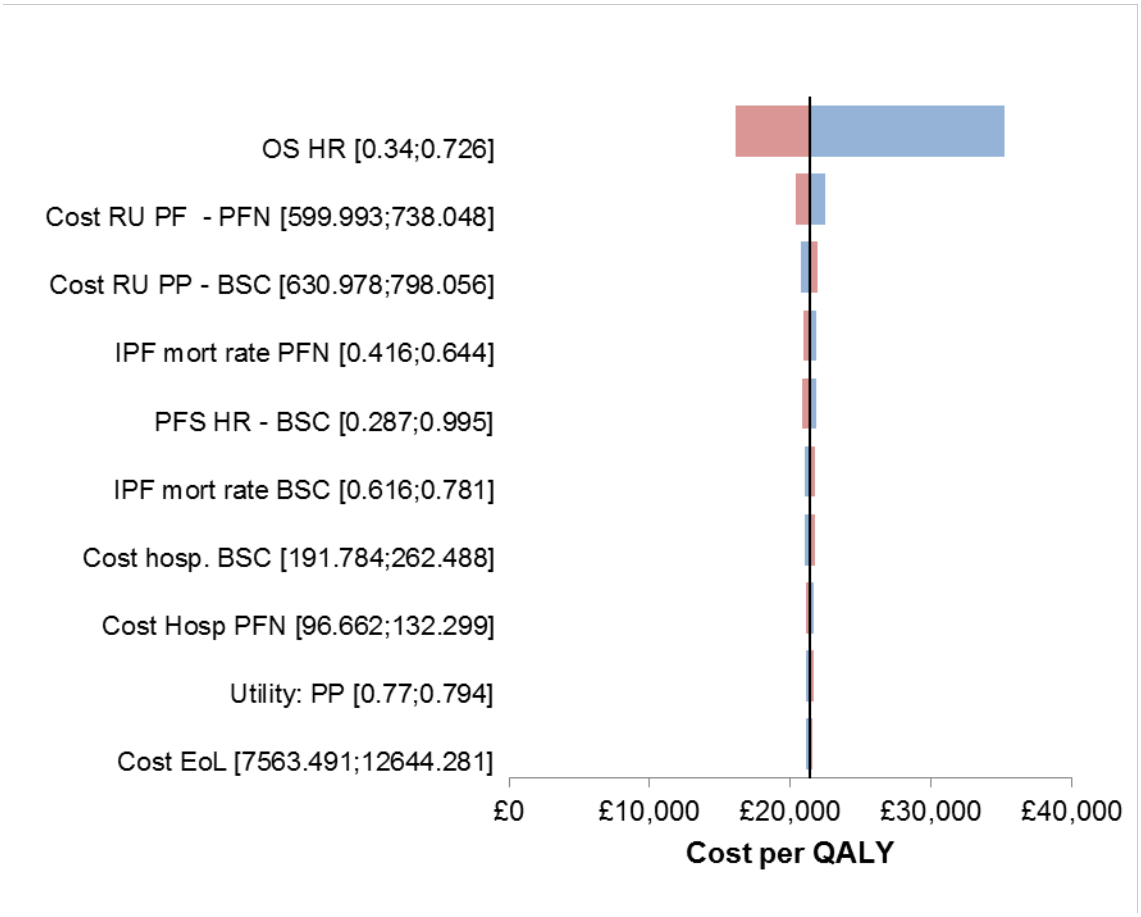
5.2.12. Sensitivity analyses

The company conducted a range of uncertainty analyses including probabilistic sensitivity analysis (PSA), deterministic univariate sensitivity analyses (DSA) and scenario analyses. Results for the ITT population for the DSA and PSA are reported here (taken from the company's clarification response[10] and the PAS submission template). Findings for scenario analyses and the subgroups of people with a percent predicted FVC of 50% to 80% and >80% are similar and therefore not reported here, but are available in the clarification responses[10] and the PAS submission template.

- **Deterministic one-way sensitivity analysis**

Figure 42Figure 42 shows the one-way DSA conducted by the company for the ITT population with the PAS applied. As recognised by the company, the ICER is most sensitive to the HR for OS.

Figure 42: Tornado diagram – ITT population, PAS price (reproduced from Figure 1 of the PAS submission template)



Key: BSC, best supportive care; EoL, end of life; Hosp, hospitalisation; HR, hazard ratio; IPF, idiopathic pulmonary fibrosis; mort, mortality; PF, progression-free; PFN, pirfenidone; PP, post-progression; QALY, quality-adjusted life year; RU, resource use.

- **Probabilistic sensitivity analysis**

The company reported results from the PSA for the ITT population and the subgroup of people with a percent predicted FVC of 50% to 80% and >80% in its clarification response.[10] These are summarised below in Table 64 when using the list price. Results when incorporating the PAS are presented in Table 65 for pirfenidone versus BSC but results for pirfenidone versus nintedanib for the moderate population are reported in the confidential appendix.

The company report PSA results which are close to the deterministic results for the ITT population. The ERG notes that results for the mild and moderate IPF subgroups for the deterministic and probabilistic analyses are also similar.

The cost-effectiveness acceptability curves (CEACs) and cost-effectiveness plane for the ITT population when incorporating the PAS are presented in Figure 43 and Figure 44, respectively.

Table 64: PSA results (list price) for ITT population and people with a percent FVC >80% at baseline

	Costs	**QALYs**	**Inc Costs (vs. BSC)**	**Inc QALYs (vs. BSC)**	**ICER**	**Probability that pirfenidone is optimal at £20,000 per QALY gained**	**Probability that pirfenidone is optimal at £30,000 per QALY gained**
ITT – mild to moderate IPF (see clarification response, Table 179)							
BSC	[redacted]	3.765					
Pirfenidone	[redacted]	5.68	[redacted]	1.91	[redacted]	[redacted]	[redacted]
People with a percent predicted FVC of 50 - 80% at baseline (see clarification response, Table 208)							
BSC	[redacted]	3.42					
Pirfenidone	[redacted]	5.15	[redacted]	1.74	[redacted]	[redacted]	[redacted]
People with a percent predicted FVC > 80% at baseline (see clarification response, Table 190)							
BSC	[redacted]	4.799					
Pirfenidone	[redacted]	7.00	[redacted]	2.21	[redacted]	[redacted]	[redacted]

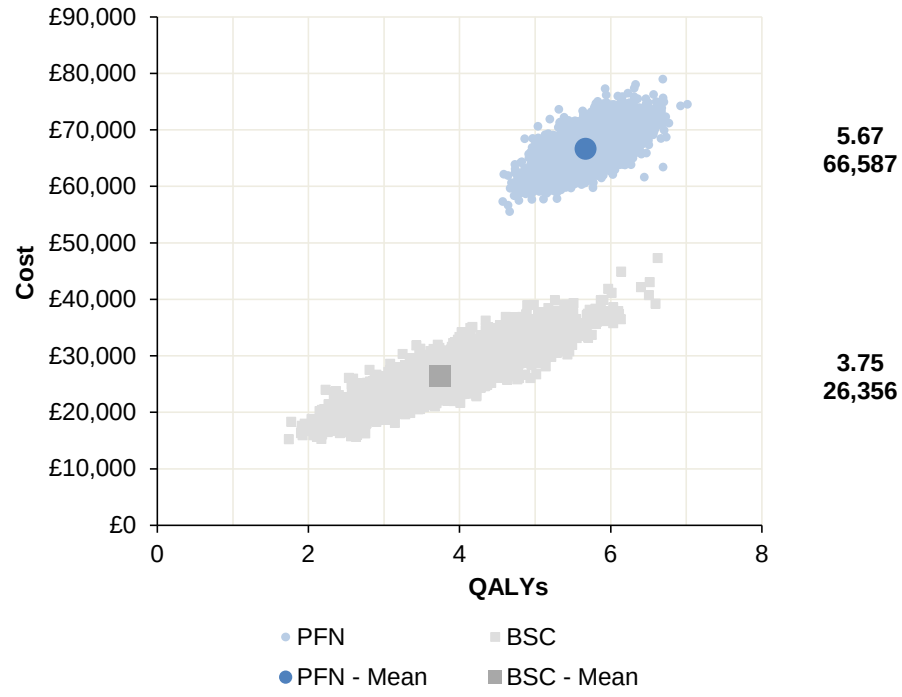
BSC – best supportive care

Table 65 PSA results (with PAS) for ITT population and people with a percent FVC >80% at baseline

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER	Probability that pirfenidone is optimal at £20,000 per QALY gained	Probability that pirfenidone is optimal at £30,000 per QALY gained
ITT – mild to moderate IPF (generated by ERG using company model submitted following clarification)							
BSC	£26,356	3.748					
Pirfenidone	£66,587	5.670	£40,231	1.92	£20,928	0.44	0.85
People with a percent predicted FVC of 50 – 80% at base-line (reproduced from PAS template, Table 15)							
BSC	£24,651	3.40					
Pirfenidone	£61,029	5.14	£36,378	1.74	£20,863	0.34	0.47
People with a percent predicted FVC > 80% at base-line (reproduced from PAS template, Table 14)							
BSC	£31,448	4.758					
Pirfenidone	£84,283	7.01	£52,835	2.25	£23,476	0.27	0.76

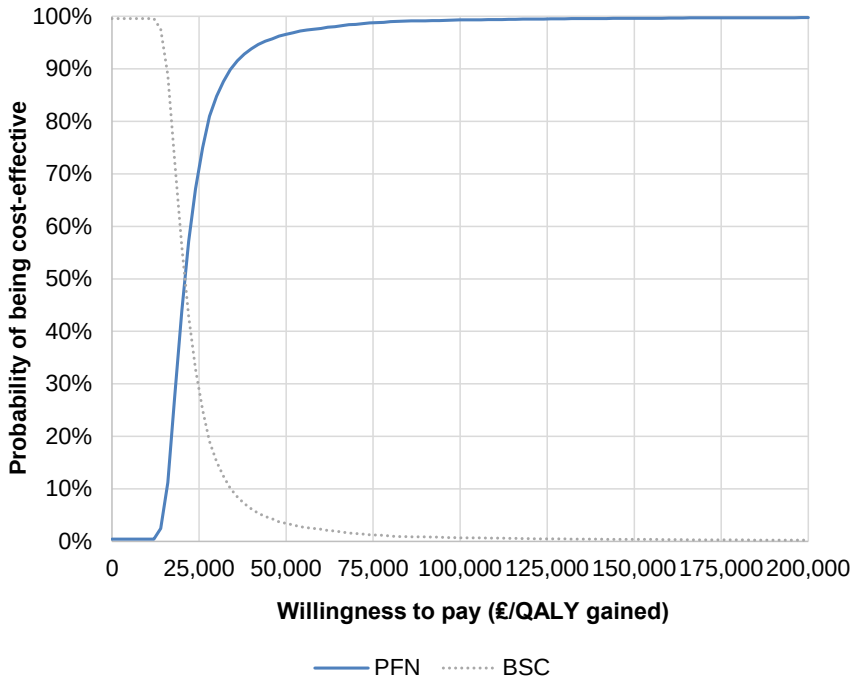
BSC – best supportive care

Figure 43: PSA scatterplot – ITT population, PAS price (generated by ERG)



Key: *BSC, best supportive care; PFN, pirfenidone; QALY, quality-adjusted life year*

Figure 44: Cost-effectiveness acceptability curve – ITT population, PAS price (generated by ERG)



Key: *BSC, best supportive care; PFN, pirfenidone; QALY, quality-adjusted life year*

ERG comments on PSA

The ERG has a number of concerns with the company's PSA. Notably, the majority of distributions appear to be arbitrary.

- The treatment effects from the NMA are varied using a gamma distribution based on the confidence interval assuming the HR to be normally distributed around the median. The ERG considers this to be inappropriate and that the CODA samples, using estimates from the predictive distributions should be used in the PSA. The ERG considers that the approach used by the company tend to underestimate the uncertainty in the treatment effects.
- PFS and OS are also modelled independently from each other and therefore no correlation is included. It is also possible in theory within the company's model, for the PFS curve to be greater than the OS curve as no constraint is added. However, the ERG notes that within the company's base-case assumptions, OS is consistently greater than PFS.
- The ERG considers the sampling of health utility values to be questionable and may underestimate uncertainty. The ERG observes that the mean SGRQ scores in people who are progression-free and with progression are sampled independently from each other assuming a normal distribution based on the mean score and standard errors estimated from the GEE model. This approach ignores the correlation between health states; the ERG considers that the variance-covariance matrix from the GEE should be used instead. The ERG further notes that the uncertainty in the mapping algorithm used to estimate the EQ-5D score is not accounted for within the company's model.
- The ERG further notes that the majority of distributions used to sample costs (resource use, management of AE, hospitalisation costs, and end of life) appear to be arbitrary. The company arbitrarily varied costs from a gamma distribution assuming an arbitrary variance of 20% around the mean despite having sufficient information to estimate the precision around some of these parameters.

- **Scenario analysis**

In addition to the DSA and PSA, the company reports cost-effectiveness results across nine groups of scenarios; these involved altering the model time horizon (10-year to 30-year), utility values, duration of the treatment effect, baseline hazard of death, progression and time to discontinuation, the studies included in the NMA for OS, PFS and exacerbations, implementation of the stopping rule and resource use. Results are available in the CS and clarification response. In brief, the ICERs were sensitive to the time horizon, assumption relating to the duration over which the treatment effect is assumed to remain constant, the parametric distributions for OS in people initiating pirfenidone, the treatment effects taken from the NMAs for OS only, and the inclusion of the stopping rule.

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 66: 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 National Jewish Health Interstitial Lung Disease database and not the US strand registry as suggested by the company. 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.

Table 67: Comparison of OS and PFS – ITT population (reproduced from CS,[4] Table 123)

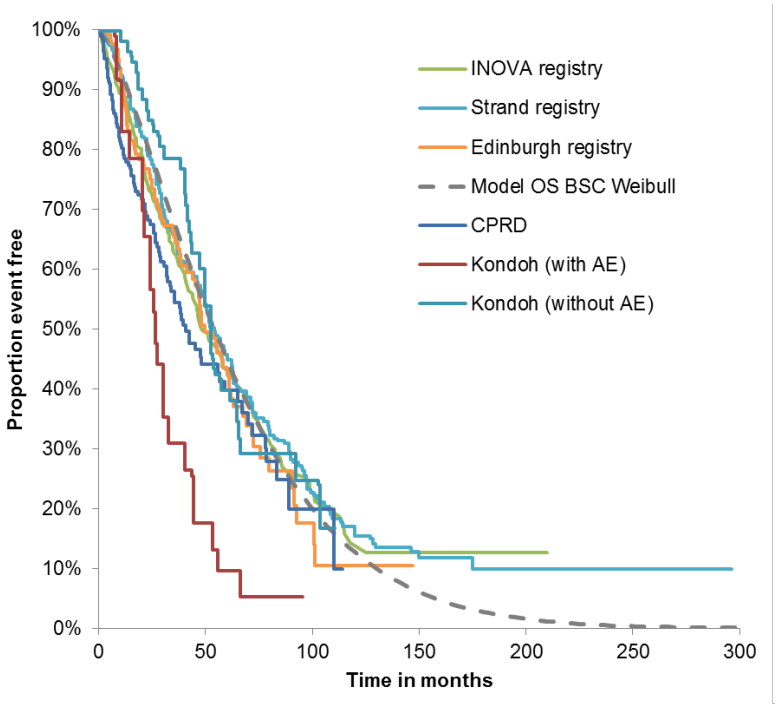
Outcome	This submission		Fisher *et al.*[64]		Roskell *et al.*[66]	
	BSC	PFN	BSC	PFN	BSC	PFN
Mean OS	5.38	8.67	6.10	9.29	5.25	9.26
Mean PFS	1.50	2.05	1.28	3.28	NR	NR

Key: IPF, idiopathic pulmonary fibrosis; LY, life year; NR, not reported; NTB, nintedanib; PFN, pirfenidone; QALY, quality-adjusted life year.

Overall, the company considers that their model generates estimates for OS and PFS (Table 67) and QALYs (Table 66) for both people initiating pirfenidone and BSC which are comparable to previous economic evaluations.

In addition, the company reports a comparison of the modelled OS in people initiating BSC predicted by the model against long-term registry data (see Figure 45). The company considers that the predicted OS for BSC in the model is consistent with long-term registries.

Figure 45: Long-term overall survival for BSC IPF people – ITT population



Number at risk	0	50	100	150	200	250	300
Strand	321	174	72	41	31	31	0
Edinburgh	125	44	6	0	0	0	0
CPRD	193	31	2	0	0	0	0
Kondoh with AE	23	4	0	0	0	0	0
Kondoh without AE	51	27	12	0	0	0	0
INOVA	234	82	22	8	8	0	0

Key: AE, acute exacerbation; BSC, best supportive care; OS, overall survival.

5.2.13.1. ERG's comments regarding the model validation of the company's model

The ERG observes that whilst the CS presents information regarding the external validity of the model, the CS does not describe any other forms of quality assurance such as:

- Validation of the model structure and key structural assumptions using clinical experts to ensure face validity;
- Peer review of the model by an independent health economist, or;
- Verification of the calculations within the model by an independent modeller.

As described in Section 5.2, the ERG has a number of concerns with the company's model regarding the conceptual representation of the condition, the representation of the treatment pathway in IPF, the implementation of the stopping rule and questionable structural assumptions including the assumption of a constant treatment effect over time. Based on these concerns, the ERG considers the company's model to lack face validity.

As part of its critical appraisal, the ERG checked the calculation to identify any programming errors and/or inconsistencies in the economic model. Inputs were also varied to establish if changes in inputs resulted in expected changes to the model outputs. Checks were also performed to ensure that the parameters presented in the CS and the company response to clarification correspond to those used in the economic model. No major programming errors were identified in the company's model during this process. The ERG identified however some minor programming errors and discrepancies, some of which were rectified in the revised economic model submitted by the company following responses to clarifications. These included:

- Lack of discounting for the cost of end of life (rectified in the revised model),
- Inclusion of the cost oxygen (rectified in the revised model following ERG comment),
- Double-counting of resource use in the first cycle,
- Discrepancies between results from the NMA presented in the clarifications responses for AEs and treatment effects used in the economic model,
- Miscalculation of the incidence of AEs –NMA outputs applied incorrectly in the model,
- Minor programming error for the PSA for cost for health professional visits for the progression-free health state,

The ERG further observes that in the revised model, an additional error was introduced by the company which was not present in the original model submitted to NICE. In brief, in the original submission to

NICE the cost associated with progression was correctly applied to patients in the progression health state irrespective of treatments. In the revised model, the company applied the costs for the progression-free health state to patient in the progressed health state, but only for the pirfenidone treatment arm. This change between the original model and the model submitted following clarification was not mentioned by the company.

The ERG also notes that whilst the PAS for pirfenidone was implemented correctly in the revised company model, the ERG had to correct the implementation of the PAS for nintedanib, in addition to setting the discount to its true confidential value, before generating the results presented in the confidential appendix. This was because the discount for nintedanib was being applied in addition to the discount for pirfenidone when calculating the nintedanib drug costs.

The company compares the model prediction for OS with estimates from two survival modelling studies reported in a power point presentation by Fisher *et al.* (2015) and a poster presented by Roskell *et al.* (2014). The ERG considers the comparison to be of limited relevance given that the same source of data is used for pirfenidone (RECAP OLE) and there are potential biases associated with the use of registry data as described in section 5.2.

The company also justifies its structure based on a previous economic evaluation conducted by Loveman et al[94] but does not provide a comparison of the results with this study. The ERG notes that Loveman *et al.* estimated the number of total QALYs to be 2.98 in people initiating BSC and 3.34 in people initiating pirfenidone compared with 3.15 and 4.46 in the company's model. These differences are not discussed by the company.

Finally, a key argument from the company regarding the model's validity relies on a comparison between the modelled OS for BSC and the OS observed in three long-term registries. As noted in Section 5.2, the ERG has several concerns with the survival observed in these registries compared with people initiating BSC that were enrolled in the ASCEND/CAPACITY trials in that the survival from the registry did not validate the survival from the BSC arm of the trial.

In conclusion, the ERG considers the validation undertaken by the company to be misleading and considers that the company's base-case may overestimate the benefit of pirfenidone compared with BSC. The ERG further considers the lack of reporting on the assessment of face validity for the model using clinical experts to be a matter of concern.

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, an unlicensed dose 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.
3. **Stopping the treatment effect after 2 years (approximately the end of follow-up of the clinical evidence for pirfenidone vs. BSC) compared with extrapolating the treatment effect to the entire model duration.** As described in Section 5.2.2.5, there is considerable uncertainty regarding the duration over which the treatment effects observed in the trials could be reasonably expected to persist. Consequently, the ERG present results using an optimistic scenario (treatment effect assumed to be constant over the entire lifetime – as assumed in the company's base-case) and a pessimistic scenario (treatment effect stop approximately after the end of follow-up of the clinical evidence at 2 years).
4. **Capping of utility estimates for individuals at 1.0 in the IPD used to derive average utilities for the progression-free and progressed state.** As described in Section 5.2.8.1, the company's base-case utility values were estimated from the aggregate mean SGRQ mapped onto EQ-5D using a mapping function. The ERG noted that when applied at the individual-level, the mapping function predicted values over 1.0. The ERG considers that the utility values estimated at the individual-level and capped at 1.0 is more appropriate leading to a mean utility value of 0.82 for progression-free and 0.76 for progressed disease (compared with 0.85 and 0.78 in the company's base-case).
5. **Adjustment of utility by age.** As described in Section 5.2.8.1, the company's base-case assumes utility values to be constant with respect to age or time. This has the effect of over-estimating the total number of QALYs. Whilst it is not possible within the company's model to explore the impact of progression in quality of life with respect to time, the ERG considers that utility values should at least be adjusted by age to avoid over-estimating QALYs. Health utilities were adjusted by age by the ERG based on the ratio of the change in utility values observed in the general population taken from an analysis conducted by Ara *et al.* (2010)[114] using data from the Health Survey for England (HSE).
6. **Including costs associated with end of life for all people irrespective of the cause of death.** The ERG considers that the company's assumption that only IPF-related mortality is associated with end of life costs is inadequately justified by the evidence. In the absence of evidence on the differential costs according to the cause of death, the one year cost assumed by the company for end of life care (£9,996) is applied to all deaths, irrespective of cause. The ERG notes that the impact will be slightly different between treatment arms due to discounting.
7. **Including titration in the first cycle based on data provided by the company at the clarification stage.** A different dose intensity is used between the first (3 months) and subsequent cycles. In response to clarification, the company provided the ICER for this

analysis. It should be noted that the ERG was not able to replicate the ICER provided by the company and therefore the ICER for this analysis presented by the ERG are inconsistent with those reported by the company in Table 19 of the response to clarification (see addendum to clarification response).

8. **Using compliance from IMPULSIS 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 IMPULSIS 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 applied to nintedanib in the company's base-case. A scenario analysis is presented in which the stopping rule is applied for both pirfenidone and nintedanib.

Within the ITT trial population (adults with mild to moderate IPF), the company's model estimates the ICER for pirfenidone against BSC to be [redacted] per QALY gained (probabilistic ICER: [redacted]) using the list price for pirfenidone and £21,387 per QALY gained (probabilistic ICER: £20,928) when incorporating the PAS. Within the subgroup of people with a percent predicted > 80% at baseline (considered to be mild IPF), the company's model estimates the ICER for pirfenidone against BSC to be [redacted] per QALY gained (probabilistic ICER: [redacted]) using the list price for pirfenidone and £24,187 per QALY (probabilistic ICER: £23,476) when incorporating the PAS. Within the subgroup of people with a percent predicted FVC of 50 - 80% at baseline (considered to be moderate IPF), nintedanib is ruled out due to extended dominance at the list price; the resulting ICER for pirfenidone versus BSC is estimated to be to be [redacted] per QALY gained (probabilistic ICER: [redacted]) using the list price and £21,318 (probabilistic ICER: £20,863) when incorporating the PAS. The results for pirfenidone versus nintedanib when incorporating the PAS for the moderate subgroup cannot be reported here and can be found in the confidential appendix.

The company presented a series of scenario analyses. The ICERs were mostly sensitive to the time horizon, assumptions regarding the duration over which the treatment effect is assumed to remain constant, the parametric distributions for OS in people initiating pirfenidone, the treatment effects taken from the NMAs for OS only, and the inclusion of the stopping rule.

The ERG critically appraised the company's health economic analysis and the model upon which this analysis was based. The ERG has a number of concerns regarding the model structure and parameterisation of the company's model. These include: (a) the inability of the model to capture the progressive nature of IPF; (b) the absence of stopping rule in the company's base-case; (c) the inadequacy of the partitioned survival approach when implementing a stopping rule; (d) the assumption that treatment effect is constant over the entire model duration, and; (e) estimation of the treatment effect. The ERG further observes that under the company's base-case assumption, there are discrepancies between the model's predictions of OS and the observed trial data in people initiating BSC. The company does not comment on these discrepancies and focus instead on a comparison with registry data which does not match the trial data in people initiating BSC.

Whilst a number of issues identified could not be addressed by the ERG without major restructuring of the economic model (particularly amending the implementation of the stopping rule), a number of analyses were undertaken by the ERG which informed the ERG preferred base-case. The main changes

include: (a) 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; (b) use of the Gompertz distribution for OS (rather than the Weibull); (c) exploring different durations over which the treatment effect is assumed to be maintained (ranging between 2 years to the entire time horizon); (d) capping utility estimates for individuals at 1.0; (e) adjusting utility by age; (g) including costs associated with end of life for all people irrespective of cause of death; (h) amending dose reductions/interruptions for pirfenidone and nintedanib, and; (i) correcting of errors in the economic model. The results of these exploratory analyses are summarised in Section 6.

6. IMPACT ON THE ICER OF EXPLORATORY ANALYSES UNDERTAKEN BY THE ERG

Section 6.1.1 summarises the impact of each individual change which forms part of the ERG preferred base-case. Section 6.1.2 presents the ERG preferred base-case.

6.1.1. Impact of each individual change which forms the ERG-preferred base-case assumptions

Table 68 and Table 69 presents the impact on the ICER of each individual change for the ITT population and subgroups of people with a percent predicted FVC of 50 – 80% and > 80% using the list price. Table 68 presents the results for the deterministic model, whilst Table 69 presents the results for the probabilistic model for the scenario where the efficacy estimates were updated to use the CODA samples. Results using the PAS are presented in Table 69 and Table 71 for the deterministic and probabilistic analysis. The ICERs for nintedanib versus BSC are not included in Table 68 as nintedanib is always extendedly dominated by pirfenidone and BSC at the list price. It can be seen from Table 68 and Table 69 that the ICERs are sensitive to four key assumptions: (i) the duration of extrapolation of the treatment effect; (ii) inclusion of the stopping rule; (iii) the treatment effect assumed and; (iv) the use of the Gompertz rather than the Weibull distribution for OS.

As expected, the ICERs are the most sensitive to the assumption around the extrapolation of the treatment effect. Assuming that the treatment effect does not persist beyond 2 years (compared with the company's base-case whereby the treatment effect is extrapolated over the entire model's duration) has the effect of increasing the ICER for pirfenidone against BSC from approximately [redacted] per QALY gained for the ITT population when using the list price. The ICER changed from approximately £21,000 to £73,000 when incorporating the PAS. This is because people initiating pirfenidone experience a shorter duration of benefits (2 years).

In contrast, the implementation of the stopping rule for pirfenidone has the effect of reducing the ICER for pirfenidone versus BSC from approximately [redacted] per QALY gained for the ITT population. The ICER changed from approximately £21,000 to £15,000 when incorporating the PAS. However, as described in Section 5.2.2.2, this is an artefact of the model structure whereby treatment discontinuation limits the cots but is disconnected from health outcomes. The ERG reiterates that the analyses using the stopping rule lacks face validity and provides a lower bound of the plausible ICER (i.e. most optimistic scenario).

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, has a moderate to large effect on the ICER for pirfenidone versus BSC with an increase from approximately [redacted] per QALY

gained for the ITT population (see Table 70). The ICER changed from approximately £21,000 to £29,000 when incorporating the PAS. This is because a lower treatment effect (higher HRs) leads to fewer health gains whilst on pirfenidone.

Finally, the use of the Gompertz distribution to represent the baseline hazard of death in people initiating pirfenidone leads to a moderate increase in the ICER for pirfenidone against BSC from approximately [redacted] per QALY gained for the ITT population. The ICER changed from £21,000 to £25,000 when incorporating the PAS. This is attributable to the fact that the Gompertz distribution has a shorter tail compared with the Weibull distribution, reducing the period over which treatment benefits can be accrued.

The impact of each change on the mild and moderate populations was similar to that for the ITT population, with the exception of the assumption that treatment effect stops after 2 years, where the impact was greater for the mild population (see Table 69).

Full incremental results for the four changes which had the biggest effect on the ICERs for pirfenidone versus BSC are presented in Appendix 3 for the analysis incorporating the PAS.

Table 68: 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 IMPULSIS	[redacted]	[redacted]	[redacted]
Correction of errors	[redacted]	[redacted]	[redacted]

[a] nintedanib is extendedly dominated

Table 69: **Summary of the impact of individual changes to the ICER for pirfenidone versus BSC[a], using the PAS price and mean parameter inputs (deterministic model)**

	ITT population	People with a percent predicted FVC of 50 - 80%	People with a percent predicted FVC > 80%
Company base-case	£21,387	£21,331[b]	£24,187
Inclusion of stopping rule for pirfenidone	£14,847	£15,197	£15,707
Treatment effect assumed to stop after 2 years	£72,599	£66,503	£112,214
Gompertz distribution for OS	£25,360	£24,855	£31,379
HRQoL capped at 1.0	£22,041	£21,983	£24,928
Adjustment of HRQoL by age	£22,716	£22,487	£26,129
End of life costs applied to death irrespective of causes	£21,957	£22,000	£24,606
Pirfenidone dose titration	£21,120	£21,060	£23,893
Correction of errors	£22,574	£22,501	£25,519

[a] results for pirfenidone versus nintedanib are presented in the confidential appendix

[b] generated by ERG after correcting error in calculation of days within drug costs

Table 70 Summary of the impact of individual changes to the ICER for pirfenidone versus BSC, using the list price for the probabilistic model [b]

	ITT population	People with a percent predicted FVC of 50 - 80%			People with a percent predicted FVC > 80%
	Pirfenidone vs. BSC	Nintedanib vs. BSC	Pirfenidone vs. nintedanib	Pirfenidone vs. BSC	Pirfenidone vs. BSC
Company base-case	[redacted]	Nintedanib extendedly dominated by pirfenidone		[redacted]	[redacted]
Treatment effect at 72 weeks (CODA sample)[b]	[redacted]	£40,436	[redacted]	[redacted]	[redacted]

[b] Run probabilistically in order to incorporate the CODA sample

Table 71: Summary of the impact of individual changes to the ICER for pirfenidone versus BSC, using the PAS price for the probabilistic model [b]

	ITT population	People with a percent predicted FVC of 50- 80%			People with a percent predicted FVC > 80%
	Pirfenidone vs. BSC	Nintedanib vs. BSC	Pirfenidone vs. nintedanib	Pirfenidone vs. BSC	Pirfenidone vs. BSC
Company base-case	£20,928	See confidential appendix		£20,863	£23,476
Treatment effect at 72 weeks (CODA sample)[b]	£28,922			£28,766	£33,060

[b] Run probabilistically in order to incorporate the CODA sample

6.1.2. ERG-preferred base-case ICERs

The ERG's preferred base-case, which combines individual changes detailed in Section 6.1, is presented in Table 72 assuming no stopping rule for either treatment and in Table 74 assuming the stopping rule for both treatments, for the list price. Equivalent results when incorporating PAS are reported in Table 73 and Table 75 (with the results for pirfenidone versus nintedanib in the moderate subgroup reported in the confidential appendix). The ERG's preferred base-case is presented using an optimistic and pessimistic assumption regarding the duration of the treatment effect (lifetime to 2 year). This has been done because whilst the clinical advisors to the ERG considered it possible that there may be continued effectiveness with long-term treatment, the duration of persistence for any long-term treatment effect is currently highly uncertain, particularly given that this is a heterogeneous condition and the mechanism of treatment is not fully understood at this time. Results are run probabilistically (5,000 iterations) to incorporate the CODA sample.

Based on the ERG's preferred base-case assumptions, no stopping rule and the pirfenidone list price, within the ITT population (adults with mild to moderate IPF), the ICER for pirfenidone versus BSC is expected to be in the range [redacted] per QALY gained. The inclusion of the stopping rule results in ICERs for pirfenidone versus BSC of [redacted] per QALY gained. When incorporating the PAS the ICERs range from £39,895 to £115,751 per QALY gained without the stopping rule and £27,124 to £75,121 per QALY gained with the stopping rule.

Based on the ERG's preferred base case assumptions, no stopping rule and the pirfenidone list price, within people with a percent predicted FVC > 80% (considered to be mild IPF), the ICER for pirfenidone versus BSC is expected to be in the range of [redacted] to [redacted] per QALY gained. When the stopping rule is assumed, the ICER for pirfenidone versus BSC is expected to be in the range of [redacted] to [redacted] per QALY gained. When incorporating the PAS the ICERs range from £49,921 to £186,260 per QALY gained without the stopping rule and £31,722 to £113,365 per QALY gained with the stopping rule.

Based on the list price, within people with a percent predicted FVC of 50 - 80% (considered to be moderate IPF), pirfenidone consistently produced greater QALYs compared with nintedanib at a lower cost, and therefore nintedanib was dominated by pirfenidone, irrespective of whether the stopping rule is included. Excluding the stopping rule, the expected ICER for pirfenidone versus BSC is expected to be in the range of [redacted] per QALY gained. When the stopping rule is assumed, the ICER for pirfenidone versus BSC is expected to be in the range of [redacted] per QALY

gained. When incorporating the PAS the ICERs range from £39,166 to £104,915 per QALY gained without the stopping rule and £27,432 to £70,234 per QALY gained with the stopping rule. The results for pirfenidone versus nintedanib when incorporating the PAS are reported in the confidential appendix.

Full incremental results for the scenairos presented in Tables 73 and 75 are provided in Appendix 3.

Table 72: ERG-preferred base-case assuming no stopping rule (ICER for pirfenidone versus BSC), analyses conducted, using the list price

	ITT population	**People with a percent predicted FVC of 50 - 80%**	**People with a percent predicted FVC > 80%**
Optimistic ERG base-case (life-time treatment effect) – probabilistic	[redacted]	[redacted][a]	[redacted]
Pessimistic ERG base-case (2 years of treatment effect) - probabilistic	[redacted]	[redacted][a]	[redacted]

[a] Nintedanib dominated by pirfenidone in ERG preferred base-case

Table 73: ERG-preferred base-case assuming no stopping rule (ICER for pirfenidone versus BSC)[a], analyses conducted, using the PAS price

	ITT population	**People with a percent predicted FVC of 50 – 80%**	**People with a percent predicted FVC > 80%**
Optimistic ERG base-case (life-time treatment effect) – probabilistic	£39,895	£39,166	£49,921
Pessimistic ERG base-case (2 years of treatment effect) – probabilistic	£115,751	£104,915	£186,260

[a] results for pirfenidone versus nintedanib are presented in the confidential appendix

Table 74: ERG-preferred base-case assuming the stopping rule to apply (ICER for pirfenidone versus BSC), analyses conducted using the list price

	ITT population	People with a percent predicted FVC of 50 - 80%	People with a percent predicted FVC > 80%
Optimistic ERG base-case (life-time treatment effect)	[redacted]	[redacted][a]	[redacted]
Pessimistic ERG base-case (2 years of treatment effect)	[redacted]	[redacted][a]	[redacted]

[a] Nintedanib dominated by pirfenidone in ERG preferred base-case

Table 75: ERG-preferred base-case assuming the stopping rule to apply (ICER for pirfenidone versus BSC)[a], analyses conducted using the PAS price

	ITT population	People with a percent predicted FVC of 50 - 80%	People with a percent predicted FVC > 80%
Optimistic ERG base-case (life-time treatment effect)	£27,124	£27,432	£31,722
Pessimistic ERG base-case (2 years of treatment effect)	£75,121	£70,234	£113,365

[a] results for pirfenidone versus nintedanib when incorporating the PAS are in the confidential appendix

7 END OF LIFE

The CS states that life-expectancy in people with IPF is 3 years from the time of diagnosis (CS, page 43). The ERG therefore does not consider that pirfenidone for the treatment of IPF meets the criteria laid out in the NICE methods guide for a 'life-extending treatment at the end of life', which is that the treatment is indicated for patients with a short life expectancy, normally less than 24 months.[43]

8 OVERALL CONCLUSIONS

The ERG had some concerns regarding the generalisability of the trial population to patients with IPF and comorbid obstructive airway disease. These patients were excluded from the three main RCTs comparing pirfenidone with placebo (ASCEND, CAPACIY 1 and CAPACITY 2), but according to clinical advisors to the ERG, these patients would be considered for treatment in current practice provided they have a percent predicted FVC of between 50% and 80%.

The meta-analysis of trial data for the outcome of PFS is considered to be subject to some uncertainty due the combination of data from trials which used different definitions of PFS and should be interpreted with caution.

The three main pirfenidone RCTs (ASCEND, CAPACITY 1 and CAPACITY 2) were considered by the ERG to be at low to moderate risk of bias, on account of inconsistencies between some protocol-specified outcomes and analyses and those reported in the CS, and the possible influence of uncontrolled variables such as rate of disease progression.

The ERG considers the data from SP2 and SP3 to be less relevant to the decision problem due to the use of a non-licensed dose of pirfenidone, and differences in the population, which was exclusively Japanese and was therefore considered to be less relevant to the population likely to be treated in England. These two studies were also assessed to be at higher risk of bias than the three main pirfenidone RCTs.

The ERG concludes that whilst the available evidence suggests that there is a statistically significant reduction in all-cause mortality for pirfenidone compared with placebo, there remains uncertainty regarding whether the size of the treatment benefit for overall survival is constant over time due to variation in the treatment effect estimated using data from 52 weeks and 72 weeks.

The ERG concludes that there is some evidence to support a statistically significant reduction in the decline in percent predicted FVC compared with placebo, but notes that a statistically significant treatment effect was not demonstrated in one of the RCTs (CAPACITY 1), which weakens the strength of the evidence for this outcome.

The ERG concludes that pirfenidone does not appear to have a significant effect in individual trials on other outcomes that are important to patients, such as disease specific health-related quality of life measures (SGRQ, UCSD SOBQ). The evidence for a statistically significant treatment effect on 6MWD was not consistent in the CAPACITY trials, and therefore the effect of pirfenidone on physical function, which is understood by the ERG to be an important driver of HRQoL, remains uncertain.

The ERG concludes that the AEs from the trials are consistent with those listed in the SmPC and that pirfenidone is generally well tolerated with most AEs experienced being mild to moderate.

A *post hoc* pooled analysis of ASCEND and CAPACITY 1 & 2 found no evidence for differential treatment effects according to disease severity, as assessed using three key efficacy outcomes; absolute ≥10% FVC decline, ≥50m 6MWD decline, and ≥20-point worsening of dyspnoea as measured by UCSD SOBQ. For these analyses disease severity was categorised according to baseline percent predicted FVC of 50 - 80% (moderate IPF) and >80% (mild IPF). In response to a clarification request from the ERG, the company also provided subgroup analyses according to disease severity for OS and PFS from the ASCEND and CAPACITY trials, although exact numbers within each subgroup in each trial arm were not reported. The findings did not suggest differential treatment effects according to disease severity for either outcome (as judged by the reported HR and 95% CI), however a treatment-by-subgroup interaction test was not reported so it is unclear if the difference between these subgroups was statistically significant. This subgroup analysis is particularly relevant to the decision problem as these groups had different comparator treatments and therefore separate analyses have been presented in the economic section for these subgroups. The ERG concludes that the evidence presented in the CS is not sufficient to support the use of subgroup specific treatment effects for these two groups.

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.

Based on the company model when using the list price, the ICER for pirfenidone against BSC is [redacted] per QALY gained within the ITT-trial population in adults with mild to moderate IPF. The ICER for people with a percent predicted FVC >80% at baseline was [redacted] per QALY against BSC when using the list price. The ICER for people with a percent predicted FVC of 50 - 80% at baseline is [redacted] per QALY gained against BSC when using the list price. Nintedanib was extendedly dominated when using the list price in patients with a percent predicted FVC of 50 - 80% at baseline.

Based on the company model when incorporating the PAS for pirfenidone, the ICER for pirfenidone versus BSC was £21,387 per QALY in the ITT population and £24,187 per QALY in the mild subgroup (percent predicted FVC >80% at baseline) and £21,318 per QALY in the moderate subgroup (percent predicted FVC of 50 - 80% at baseline). The results for pirfenidone versus nintedanib when incorporating the nintedanib and pirfenidone PAS (moderate subgroup) are reported in the confidential appendix.

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.

The ERG identified a number of concerns regarding the model structure and parameterisation of the company's model including (a) the inability of the model to capture the progressive nature of IPF, (b) the absence of stopping rule in the company's base-case; (c) the inadequacy of the partition-survival approach when implementing a stopping rule, (d) the assumption that treatment effect is constant over the entire duration of the model, and (e) estimation of the treatment effect. The ERG further observes that under the company's base-case assumption, there are discrepancies between the model's prediction for OS and observed trial data in people initiating BSC.

A number of analyses were undertaken by the ERG which informed the ERG preferred base-case. The ERG's exploratory analysis led to consistently higher ICERs for pirfenidone against BSC for all three populations (ITT, mild [percent predicted FVC>80%] and moderate [percent predicted FVC of 50 - 80%] subgroups), even under the company's optimistic base-case assumption that the treatment effect is assumed to hold for the entire duration of the model. Using the list price, the ICER for pirfenidone ranged from [redacted] per QALY against BSC in the ITT-trial population. In the mild subgroup, the ICER ranged from [redacted] per QALY against BSC when using the list price. In the moderate subgroup the ICERs ranged from [redacted] per QALY against BSC when using the list price. When incorporating the PAS the ICERs for pirfenidone versus BSC were above £27,000 per QALY in the ITT population, above £31,000 in the mild subgroup and above £27,000 in the moderate subgroup. Results for pirfenidone versus nintedanib in the moderate subgroup when incorporating the PAS are presented in the confidential appendix.

A key uncertainty is around the duration of the extrapolation of the treatment effect. In the company's base-case, the treatment effect is assumed to be constant over the model's entire duration. The ERG considered this to be overly optimistic and inadequately supported by the evidence and believes that the treatment effect could reduce over time; although there is a lack of data to support either assumption. Assuming a shorter duration of extrapolation for the treatment effect led to an increase in the ICERs.

An important limitation in the company's model implementation regards the implementation of the stopping rule. Despite the fact that the NICE recommendations for pirfenidone (TA379) and nintedanib (TA282) include identical stopping rules, the company's model structure does not accommodate the robust exploration of the impact of this stopping rule on the ICERs. Whilst a scenario analysis including the stopping rule for both treatments is presented in the CS, the ERG has a number of concerns with

this analysis as stopping treatment earlier led to a reduction in treatment costs, but left the gain in life years and QALYs unchanged. The ERG considers that results from these analyses need to be interpreted with caution and that no robust ICERs have been presented by the company when the stopping rule is implemented. The ERG considers that the ICERs presented by the company using the stopping rule could represent a lower bound of the true ICER when the stopping rule is implemented in clinical practice, as the life-time costs of treatment are reduced when the stopping rule is applied in the model, but the incremental QALYs are not reduced by the shorter duration of treatment.

8.1 Implications for research

IPF is a heterogeneous condition and there is natural variability in the rates of decline in percent predicted FVC. It is therefore difficult for clinicians to know if treatment is benefiting an individual patient as a patient who experiences stability on the drug may have had a low rate of decline in FVC without treatment and a patient who experiences a moderate rate of decline in FVC on treatment may have experienced a more rapid decline without treatment. Further research into biomarkers which predict the rate of disease progression or which predict response to treatment would be beneficial.

9 REFERENCES

1. Roche Products Ltd. Esbriet 267 mg Hard Capsules, Summary of Product Characteristics: eMC; 2015.
2. National Institute for Health and Care Excellence. Pirfenidone for treating idiopathic pulmonary fibrosis (TA282). London, UK: The National Institute for Health and Care Excellence; 2013.
3. National Institute for Health and Care Excellence (NICE). Pirfenidone for the treatment of idiopathic pulmonary fibrosis (review of TA282) [ID837]) - Final scope. London: NICE; 2015.
4. Roche Product Limited. Pirfenidone for the treatment of adults with idiopathic pulmonary fibrosis: Company evidence submission. London, UK; 2016.
5. Kim DS, Collard HR, King TE, Jr. Classification and natural history of the idiopathic interstitial pneumonias. *Proc Am Thorac Soc* 2006;3:285-92. http://dx.doi.org/10.1513/pats.200601-005TK
6. Schmidt SL, Tayob N, Han MK, Zappala C, Kervitsky D, Murray S, *et al.* Predicting pulmonary fibrosis disease course from past trends in pulmonary function. *Chest* 2014;145:579-85. http://dx.doi.org/10.1378/chest.13-0844
7. Behr J, Bendstrup E, Crestani B, Gnther A, Olschewski H, Skld M, *et al.*: Safety and tolerability of N-acetylcysteine (NAC) with pirfenidone in idiopathic pulmonary fibrosis (IPF): PANORAMA. *European Respiratory Journal* 2015;46. http://dx.doi.org/10.1183/13993003.congress-2015.OA3477
8. Behr J, Bendstrup E, Crestani B, Gunther A, Olschewski H, Skold M, *et al.* Pirfenidone and N-acetylcysteine (NAC) for the Treatment of Idiopathic Pulmonary Fibrosis (IPF): Exploratory Efficacy Results From the PANORAMA Study. *Chest* 2015;148. http://dx.doi.org/10.1378/chest.2335639
9. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, *et al.* An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med* 2015;192:e3-19. http://dx.doi.org/10.1164/rccm.201506-1063ST
10. Roche Products Ltd. Response to Clarification Letter.; 2016.
11. Nathan. Effect of Continued Treatment With Pirfenidone Following a Clinically Meaningful Decline in Percent Predicted Forced Vital Capacity in Patients With Idiopathic Pulmonary Fibrosis. ATS 2015 Annual Congress; Denver, USA, abstract no. 1.
12. National Institute for Health and Care Excellence. Nintedanib for treating idiopathic pulmonary fibrosis (TA379). London, UK; 2016.
13. Harari S, Caminati A. Idiopathic pulmonary fibrosis: from clinical trials to real-life experiences. *Eur Respir Rev* 2015;24:420-7. http://dx.doi.org/10.1183/16000617.0042-2015
14. Chaudhuri N, Duck A, Frank R, Holme J, Leonard C. Real world experiences: pirfenidone is well tolerated in patients with idiopathic pulmonary fibrosis. *Respir Med* 2014;108:224-6. http://dx.doi.org/10.1016/j.rmed.2013.11.005
15. Koschel D, Cottin V, Skold M, Tomassetti S, Azuma A, Giot C, *et al.* Pirfenidone post-authorization safety registry (PASSPORT)–Interim analysis of IPF treatment. *European Respiratory Journal* 2014;44:1904.
16. MIMS. Esbriet. Haymarket Media Group Ltd; 2016. URL: http://www.mims.co.uk/drugs/respiratory-system/cough-congestion-respiratory-distress/esbriet (Accessed 28/03/2016, 2016).
17. Boehringer Ingelheim Limited. Ofev 150 mg soft capsules, Summary of Product Characteristics; 2015.
18. Saketkoo LA, Mittoo S, Huscher D, Khanna D, Dellaripa PF, Distler O, *et al.* Connective tissue disease related interstitial lung diseases and idiopathic pulmonary fibrosis: provisional core sets of domains and instruments for use in clinical trials. *Thorax* 2013:thoraxjnl-2013-204202.
19. Nathan SD, Meyer KC. IPF clinical trial design and endpoints. *Current opinion in pulmonary medicine* 2014;20:463-71.

20. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *American journal of respiratory and critical care medicine* 2011;184:1382-9.
21. Swigris JJ, Kuschner WG, Jacobs SS, Wilson SR, Gould MK. Health-related quality of life in patients with idiopathic pulmonary fibrosis: a systematic review. *Thorax* 2005;60:588-94. http://dx.doi.org/10.1136/thx.2004.035220
22. Swigris JJ, Esser D, Conoscenti CS, Brown KK. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. *Health Qual Life Out* 2014;12. http://dx.doi.org/10.1186/s12955-014-0124-1
23. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;184:459-66. http://dx.doi.org/10.1164/rccm.201011-1790OC
24. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, *et al.* Long-term Course and Prognosis of Idiopathic Pulmonary Fibrosis in the New Millennium. *Chest* 2011;140:221-9. http://dx.doi.org/10.1378/chest.10-2572
25. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, *et al.* Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *American journal of respiratory and critical care medicine* 2011;183:1231.
26. Boehringer Ingelheim Limited. Nintedanib for the treatment of idiopathic pulmonary fibrosis: Company evidence submission. The National Institute for Health and Care Excellence; 2015.
27. InterMune. Protocol PIPF-016: InterMune Inc; 2012.
28. InterMune. Clinical Study Report No. PIPF-006: Appendix 16.1.1. Protocol and Protocol Amendments: InterMune Inc; 2006.
29. InterMune. Clinical Study Report No. PIPF-004: Appendix 16.1.1. Protocol and Protocol Amendments: InterMune Inc; 2006.
30. Roche Products Ltd. Addendum to the response to Clarification Letter. 2016.
31. Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, *et al.* The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respiratory medicine* 2012;106:1447-55.
32. Moher D, Liberati A, Tetzlaff J, Altman DG, Group. TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009;6:e1000097. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097
33. ASCEND Clinicaltrials.gov. Efficacy and Safety of Pirfenidone in Patients With Idiopathic Pulmonary Fibrosis (IPF). In; 2016.
34. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2083-92. http://dx.doi.org/doi: 10.1056/NEJMoa1402582.
35. CAPACITY 2 ClinicalTrials.gov. Three-Arm Study of the Safety and Efficacy of Pirfenidone in Patients With Idiopathic Pulmonary Fibrosis. In; 2016.
36. CAPACITY 1 ClinicalTrials.gov. Safety and Efficacy of Pirfenidone in Patients With Idiopathic Pulmonary Fibrosis. In; 2016.
37. Noble PW, Albera C, Bradford WZ, Costabel U, du Bois RM, Fagan EA, *et al.* Pirfenidone for idiopathic pulmonary fibrosis: analysis of pooled data from three multinational phase 3 trials. *Eur Respir J* 2016;47:243-53. http://dx.doi.org/doi: 10.1183/13993003.00026-2015
38. Taniguchi H, Ebina M, Kondoh Y, Ogura T, Azuma A, Suga M, *et al.* Pirfenidone in idiopathic pulmonary fibrosis. *Eur Respir J* 2010;35:821-9. http://dx.doi.org/doi: 10.1183/09031936.00005209.
39. Azuma A, Nukiwa T, Tsuboi E, Suga M, Abe S, Nakata K, *et al.* Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005;171:1040-7.

40. Costabel U, Albera C, Bradford W, Hormel P, King TJ, Noble P, *et al.* Analysis of lung function and survival in RECAP: An open-label extension study of pirfenidone in patients with idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2014;31:198-205.
41. National Jewish Health. Participation Program for Pulmonary Fibrosis - P_3F Registry. In; 2016.
42. InterMune. Single technology appraisal: Pirfenidone for the treatment of idiopathic pulmonary fibrosis: Manufacturer's submission. Brisbane, CA, USA; 2011.
43. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013; 2013.
44. Centre for Reviews Dissemination. Systematic review: CRD's guidance for undertaking reviews in health care. CRD, University of York; 2009. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]; 2011.
45. Aravena C, Labarca G, Venegas C, Arenas A, Rada G. Pirfenidone for Idiopathic Pulmonary Fibrosis: A Systematic Review and Meta-Analysis. *Plos ONE* 2015;10:e0136160. http://dx.doi.org/10.1371/journal.pone.0136160
46. Atkins CP, Loke YK, Wilson AM. Outcomes in idiopathic pulmonary fibrosis: A meta-analysis from placebo controlled trials. *Respir Med* 2014;108:376-87.
47. Jiang C, Huang H, Liu J, Wang Y, Lu Z, Xu Z. Adverse Events of Pirfenidone for the Treatment of Pulmonary Fibrosis: A Meta-Analysis of Randomized Controlled Trial. *PLoS ONE* 2012;7:e47024.,
48. Huang H, Dai HP, Kang J, Chen BY, Sun TY, Xu ZJ. Double-Blind Randomized Trial of Pirfenidone in Chinese Idiopathic Pulmonary Fibrosis Patients. *Medicine (Baltimore)* 2015;94:e1600. http://dx.doi.org/doi: 10.1097/MD.0000000000001600.
49. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, *et al.* Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011;377:1760-9. http://dx.doi.org/doi:10.1016/S0140-6736(11)60405-4.
50. Kondoh Y, Taniguchi H, Katsuta T, Kataoka K, Kimura T, Nishiyama O, *et al.* Risk factors of acute exacerbation of idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2010;27:103-10.
51. European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. 2011. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf (Accessed 9 March, 2016).
52. Shinogi. SP3 study protocol. Date of preparation: 16 October 2006.
53. Shinogi. SP2 study protocol. Date of preparation: 9 October 2001.
54. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000;161:646-64.
55. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis [Supplementary Appendix]. *N Engl J Med* 2014;370:2083-92.
56. Roche Products Ltd. Cost-effectiveness / cost-utility AUC model of Esbriet (pirfendione) compared to BSC and Ofev (nintedanib) in mild to moderate IPF. 2016.
57. FDA. U.S. Food and Drug Administration. Pirfenidone Capsules [NDA 22-535]. Sponsor's Briefing Document for the FDA Pulmonary-Allergy Drugs Advisory Committee Meeting Scheduled for 09 March 2010. 2010. URL: http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PulmonaryAllergyDrugsAdvisoryCommittee/UCM203083.pdf (Accessed 9 March, 2016).
58. Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. *BMC Med Res Methodol* 2010;10.
59. Sahn SA, Albera C, Du Bois RM, Bradford W, Costabel U, King TE. The effect of treatment with pirfenidone on progression-free survival in patients with idiopathic pulmonary fibrosis (IPF): Exploratory analysis of outcomes using novel criteria for disease progression.

American Journal of Respiratory and Critical Care Medicine 2011;183:(Meeting Abstracts 1):[A3810].
60. Kondoh Y, Taniguchi H, Ebina M, Azuma A, Ogura T, Taguchi Y, *et al.* Risk factors for acute exacerbation of idiopathic pulmonary fibrosis-Extended analysis of pirfenidone trial in Japan. *Respir Investig* 2015;53:271-8. http://dx.doi.org/doi: 10.1016/j.resinv.2015.04.005.
61. Nathan SD. Effect of Continued Treatment With Pirfenidone Following a Clinically Meaningful Decline in Percent Predicted Forced Vital Capacity in Patients With Idiopathic Pulmonary Fibrosis. ATS 2015 Annual Congress; 15-20 May; Denver, USA, abstract no. 22.
62. Nathan SD, Albera C, Bradford WZ. Effect of continued treatment with pirfenidone following clinically meaningful declines in forced vital capacity: Analysis of data from three phase 3 trials in patients with idiopathic pulmonary fibrosis. In; 2016 (Unpublished, AiC).
63. Kreuter M. Pirfenidone: an update on clinical trial data and insights from everyday practice. *Eur Respir Rev* 2014;23:111-7. http://dx.doi.org/doi:10.1183/09059180.00008513.
64. Fisher M. Disease-progression modeling in idiopathic pulmonary fibrosis: Prediction of time to disease progression and life expectancy with pirfenidone. Paper presented at: BTS 2015 Annual Congress2-4 Dec, London, UK.
65. PIPF-002 ClinicalTrials.gov. Safety Study of Oral Pirfenidone in Patients With Pulmonary Fibrosis/Idiopathic Pulmonary Fibrosis. In; 2016.
66. Roskell N, Saunders O, Lee D, Fisher M. Long-term survival analysis: Pirfenidone compared to standard care for the treatment of patients with idiopathic pulmonary fibrosis. *European Respiratory Journal* 2014;44:A1905.
67. Navaratnam V, Fleming KM, West J, Smith CJ, Jenkins RG, Fogarty A, *et al.* The rising incidence of idiopathic pulmonary fibrosis in the U.K. *Thorax* 2011;66:462-7. http://dx.doi.org/doi: 10.1136/thx.2010.148031
68. Genentech I. Safety Study of Oral Pirfenidone in Patients With Pulmonary Fibrosis/Idiopathic Pulmonary Fibrosis. 2015, https://clinicaltrials.gov/ct2/show/NCT00080223.
69. Albera C. Pirfenidone Is Efficacious in Patients With Idiopathic Pulmonary Fibrosis (IPF) and Mild Restrictive Disease: Benefit of Early Intervention. Paper presented at: ATS 2015 Annual Congress; Denver, USA, 15-20 May.
70. Behr J, Demedts M, Buhl R, Costabel U, Dekhuijzen RP, Jansen HM, *et al.* Lung function in idiopathic pulmonary fibrosis--extended analyses of the IFIGENIA trial. *Respir Res* 2009;10. http://dx.doi.org/doi: 10.1186/1465-9921-10-101
71. Demedts M, Behr J, Buhl R, Costabel U, Dekhuijzen R, Jansen HM, *et al.* High-dose acetylcysteine in idiopathic pulmonary fibrosis. *N Engl J Med* 2005;353:2229-42.,
72. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370:2071-82.
73. Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, *et al.* Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *N Engl J Med* 2011;365:1079-87. http://dx.doi.org/doi: 10.1056/NEJMoa1103690
74. Martinez FJ, de Andrade JA, Anstrom KJ, King TE, Raghu G. Randomized Trial of N-acetylcysteine in Idiopathic Pulmonary Fibrosis. *The New England journal of medicine* 2014;370:2093-101. http://dx.doi.org/10.1056/NEJMoa1401739
75. Raghu. The Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 2012;366:1968-77.
76. Chhabra SK. Forced vital capacity, slow vital capacity, or inspiratory vital capacity: which is the best measure of vital capacity? *J Asthma* 1998;35:361-5.
77. The Idiopathic Pulmonary Fibrosis Clinical Research Network. Prednisone, Azathioprine, and N-Acetylcysteine for Pulmonary Fibrosis. *The New England journal of medicine* 2012;366:1968-77. http://dx.doi.org/10.1056/NEJMoa1113354
78. National Institute for Health and Care Excellence. Idiopathic pulmonary fibrosis - nintedanib. Final appraisal determination. December 2015, Available on: http://www.nice.org.uk/guidance/indevelopment/gid-tag491/documents [date accessed: Jan 2016].

79. Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 2002;64:583-616. http://dx.doi.org/Doi 10.1111/1467-9868.00353
80. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol* 2012;41:818-27. http://dx.doi.org/10.1093/ije/dys041
81. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making* 2013;33:657-70. http://dx.doi.org/10.1177/0272989X13485155
82. Plummer M. JAGS User Manual. In. 3.3.0 ed; 2012.
83. R Developement Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing; 2013.
84. Su Y-S, Yajima M. R2jags: Using R to Run 'JAGS'. In; 2015.
85. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 2000;10:325-37. http://dx.doi.org/Doi 10.1023/A:1008929526011
86. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *J Stat Softw* 2005;12:1-16. <Go to ISI>://WOS:000232806500001
87. Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville J. Sensitivity of a Search Filter Designed to Identify Studies Reporting Health State Utility Values; 2015.
88. Centre for Reviews of Dissemination. Search strategies. York; 2014.
89. Cooper K, Mendes D, Picot J, Loveman E. Pirfenidone for the treatment of idiopathic pulmonary fibrosis : A Single Technology Appraisal; 2012.
90. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996;313:275-83. http://www.ncbi.nlm.nih.gov/pubmed/8704542
91. Canadian Agency for Drugs and Technologies in Health. CDEC Final Recommendation - Nintedanib. Ottwa, Canadia: Canadian Agency for Drugs and Technologies in Health; 2015.
92. Canadian Agency for Drugs and Technologies in Health. CDEC Final Recommendation - Pirfenidone Resubmission. Ottwa: Canadian Agency for Drugs and Technologies in Health. 2015, Available from: https://www.cadth.ca/pirfenidone-12. .
93. Loveman E, Copley VR, Colquitt JL, Scott DA, Clegg AJ, Jones J, *et al.* The effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: systematic review, network meta-analysis and health economic evaluation. *BMC Pharmacol Toxicol* 2014;15:63. http://dx.doi.org/10.1186/2050-6511-15-63
94. Loveman E CV, Colquitt J, Scott DA, Clegg A, Jones J, et al. The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation. *Health Technology Assessment* 2015;19:1-336.,
95. Starkie HJ, Briggs AH, Chambers MG, et al. Predicting EQ-5D Values Using the SGRQ. *Value in Health* March–April 2011;Volume 14:Pages 354–60.
96. Zisman DA SM, Anstrom KJ, Collard HR, Flaherty KR, Hunninghake GW. . A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *New England Journal of Medicine* 2010;363:620-8.,
97. National Institute for Health and Clinical Excellence. Updated information and analyses to support the Single Technology Appraisal submission of pirfenidone for the treatment of idiopathic pulmonary fibrosis (IPF) - TA282. 2012, Available at: https://www.nice.org.uk/guidance/ta282/resources/idiopathic-pulmonary-fibrosis-pirfenidone-intermune-additional-analyses2.
98. National Institute for Health and Care Excellence. Idiopathic pulmonary fibrosis - Pirfenidone. Final appraisal determination. *NICE* April 2013, Available at: https://www.nice.org.uk/guidance/ta282/documents/idiopathic-pulmonary-fibrosis-pirfenidone-final-appraisal-determination-document2.
99. Albera C. Challenges in idiopathic pulmonary fibrosis trials: the point on end-points. . *EurRespirRev* 2011;20.

100. National Institute for Health and Care Excellence (NICE). Assessment group report (TA257) - Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2. June 2012, Available at http://publications.nice.org.uk/lapatinib-or-trastuzumab-in-combination-with-an-aromataseinhibitor-for-the-first-line-treatment-of-ta257.
101. Hatziandreu E. The potential cost savings of greater use of home- and hospice based end of life care in England (2008). *NAO* 2008, Available at: https://www.nao.org.uk/wp-content/uploads/2008/11/07081043_potentialsavings.pdf.
102. Handorf EA, McElligott S, Vachani A, et al. Cost Effectiveness of Personalized Therapy for First-Line Treatment of Stage IV and Recurrent Incurable Adenocarcinoma of the Lung. *Journal of Oncology Practice* 2012;8:267-74.
103. Department of Health. NHS reference costs 2013 to 2014. 2015. URL: https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 (Accessed 1 March, 2016).
104. Pittrow D KJ, Kreuter M, Hoeper MM, Wirtz H, Koschel D, et al. . Symptom burden and health related quality of life in patients with idiopathic pulmonary fibrosis in clinical practice: Insights-IPF registry. *Value in Health* 2014;17:A600.
105. King TE J, Brown KK, Raghu G, du Bois RM, Lynch DA, Martinez F, et al. . BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine* 2011;184:92-9.
106. Freemantle N, Wilson A, Fisher M. Mapping The St George's Respiratory Questionnaire To The Euroqol 5 Dimensions: A Study In Patients With Idiopathic Pulmonary Fibrosis. *Value in Health* 2015;18:A503.
107. Parfrey H LC, Gibbons MA, Armstrong E, Harris E, Frank R, et al. . Healthcare utilisation by patients with idiopathic pulmonary fibrosis; observations from the UK pirfenidone named patient programme. *Thorax* 2013;68:A164-A65.,
108. Navaratnam V FA, Glendening R et al. . The Increasing Secondary Care Burden of Idiopathic Pulmonary Fibrosis: Hospital Admission Trends in England From 1998 to 2010. *Chest* 2013;143:1078-84.
109. Wilson E, Shulgina L, Cahn A, Chilvers E, Parfrey H, Clark. AB., *et al.* Treating idiopathic pulmonary fibrosis with the addition of co-trimoxazole: an economic evaluation alongside a randomised controlled trial. *Pharmacoeconomics* 2014;31:87-99.
110. Whiting P AM, Burgers L, Westwood M, Ryder S, Hoogendoorn M, et al. . Ivacaftor for the treatment of patients with cystic fibrosis and the G551D mutation: a systematic review and cost-effectiveness analysis. *Health Technology Assessment* 2014;18:1-106.
111. National Institute for Health and Care Excellence. Costing report: Idiopathic pulmonary fibrosis. London: NICE; 2013. 2013, Available from: https://www.nice.org.uk/guidance/cg163/resources/costing-report-190080397. .
112. Scottish Medicines Consortium. Detailed advice for nintedanib (Ofev®) [SMC no. 1076/15]. 2015, Available on: http://www.scottishmedicines.org.uk/SMC_Advice/Advice/1076_15_nintedanib_Ofev/nintedanib_Ofev [date accessed: Dec 2015].
113. Corte T, Bonella F, Crestani B, Demedts M, Richeldi L, Coeck C, *et al.* Safety, tolerability and appropriate use of nintedanib in idiopathic pulmonary fibrosis. *Respir Res* 2015 Sep 24;16:doi: 10.1186/s12931-015-0276-5.,
114. Ara R, Brazier J. Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value in Health* 2010;13:509-18.

10. APPENDICES

Appendix 1: Comparison of the observed KM for OS in people initiating pirfenidone against extrapolation using parametric distributions for people with a percent FVC > and 50 - 80% at baseline

Figure 46: Comparison of the observed KM for OS in people initiating pirfenidone against extrapolation using parametric distributions for people with a percent FVC of 50 - 80% at baseline (Plot drawn by the ERG)

Figure 47: Comparison of the observed KM for OS in people initiating pirfenidone against extrapolation using parametric distributions for people with a percent FVC of 50 - 80% at baseline (Plot drawn by the ERG)

Appendix 2: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and life tables in UK in people with a percent FVC of 50 – 80% and > 80% at baseline

Figure 48: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and life tables in UK in people with a percent FVC >80% at baseline (Plot drawn by the ERG)

Figure 49: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and life tables in UK in people with a percent FVC of 50 - 80% at baseline (Plot drawn by the ERG)

Appendix 3: Full incremental analysis for key sensitivity analyses and ERG base-case

Table 76 to Table 79 present full incremental results for the four key sensitivity analyses identified in Table 69 on the main report. These results are all deterministic and incorporate the PAS. The results for pirfenidone versus nintedanib can be found in the confidential appendix.

Table 76: Treatment effect assumed to stop after 2 years – deterministic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£33,798	5.215			
Pirfenidone	£66,638	5.667	£32,840	0.452	£72,599
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£31,180	4.690			
Nintedanib	See confidential appendix				
Pirfenidone	£61,035	5.138	£29,854	0.449	£66,503
Mild population (percent predicted FVC >80%)					
BSC	£40,671	6.606			
Pirfenidone	£84,209	6.994	£43,539	0.388	£112,214

Table 77: Gompertz distribution for OS – deterministic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£25,996	3.687			
Pirfenidone	£64,362	5.200	£38,366	1.513	£25,360
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£24,430	3.374			
Nintedanib	See confidential appendix				
Pirfenidone	£59,276	4.776	£34,846	1.402	£24,855
Mild population (percent predicted FVC >80%)					
BSC	£30,124	4.520			
Pirfenidone	£79,543	6.094	£49,420	1.575	£31,379

Table 78: Inclusion of stopping rule for pirfenidone – deterministic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£26,627	3.797			
Pirfenidone	£54,360	5.664	£27,733	1.868	£14,847
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£24,868	3.443			
Nintedanib	See confidential appendix				
Pirfenidone	£50,596	5.136	£25,728	1.693	£15,197
Mild population (percent predicted FVC >80%)					
BSC	£31,729	4.824			
Pirfenidone	£65,740	6.989	£34,011	2.165	£15,707

Table 79: Treatment effect at 72 weeks (incorporating CODA samples) – probabilistic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£29,694	4.393			
Pirfenidone	£66,685	5.672	£36,991	1.279	£28,922
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£27,683	3.995			
Nintedanib	See confidential appendix				
Pirfenidone	£61,097	5.157	£33,414	1.162	£28,766
Mild population (percent predicted FVC >80%)					
BSC	£35,220	5.520			
Pirfenidone	£84,133	6.999	£48,913	1.480	£33,060

Table 80 to Table 83 below present full incremental results for the ERG-preferred base-case under both optimistic and pessimistic assumptions regarding the duration of treatment effect, both with and without the stopping rule applied, when incorporating the PAS.

Table 80: ERG-preferred base-case assuming no stopping rule with optimistic assumption regarding duration of treatment effect (life-time effect) – probabilistic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£30,972	3.968			
Pirfenidone	£69,560	4.935	£38,589	0.967	£39,895
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£29,220	3.64	-	-	-
Nintedanib	See confidential appendix				
Pirfenidone	£64,325	4.53	£35,106	0.90	£39,166
Mild population (percent predicted FVC >80%)					
BSC	£35,053	4.747			
Pirfenidone	£84,735	5.742	£49,682	0.995	£49,921

Table 81: ERG-preferred base-case assuming no stopping rule with pessimistic assumption regarding duration of treatment effect (2 years) – probabilistic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£34,439	4.617			
Pirfenidone	£69,352	4.918	£34,913	0.302	£115,751
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£32,032	4.18	-	-	-
Nintedanib	See confidential appendix				
Pirfenidone	£63,604	4.48	£31,571	0.30	£104,915
Mild population (percent predicted FVC >80%)					
BSC	£39,060	5.498			
Pirfenidone	£84,712	5.743	£45,652	0.245	£186,260

Table 82: ERG-preferred base-case assuming stopping rule with optimistic assumption regarding duration of treatment effect (life-time) – probabilistic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£30,947	3.964			
Pirfenidone	£57,216	4.932	£26,269	0.968	£27,124
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£29,225	3.64	-	-	-
Nintedanib	See confidential appendix				
Pirfenidone	£53,790	4.53	£24,565	0.90	£27,432
Mild population (percent predicted FVC >80%)					
BSC	£35,035	4.757			
Pirfenidone	£66,796	5.759	£31,761	1.001	£31,722

Table 83: ERG-preferred base-case assuming stopping rule with pessimistic assumption regarding duration of treatment effect (2 years) – probabilistic incorporating PAS

	Costs	QALYs	Inc Costs (vs. BSC)	Inc QALYs (vs. BSC)	ICER (vs. BSC)
ITT population					
BSC	£34,430	4.610			
Pirfenidone	£57,048	4.911	£22,618	0.301	£75,121
Moderate population (percent predicted FVC of 50 – 80%)					
BSC	£32,081	4.20	-	-	-
Nintedanib	See confidential appendix				
Pirfenidone	£53,249	4.50	£21,169	0.30	£70,234
Mild population (percent predicted FVC >80%)					
BSC	£39,063	5.501			
Pirfenidone	£66,794	5.745	£27,731	0.245	£113,365

Appendix 4: Errors identified and corrected by the ERG

Error identified	Description of the error	Change by the ERG
Discrepancies in the NMA AE outputs between those used in the model and those reported in clarification response	Result from Table 16 in clarification response does not match NMA outputs in sheet "NMA" in the economic model in cells E108-110 & Cells K108-110.	Values reported in Table 16 in clarification response are used in the economic model by the ERG
Incorrect application of outputs of NMA for AEs	In sheet "Model Inputs", RR are applied to the incidence of AEs on BSC. However, the inverse of RR are calculated in Sheet "NMA" and used.	The inverse of the RR used in sheet "NMA" are used for AE in sheet "Model Inputs"
Parameters for PFS using the Gompertz and Gamma for the subgroup analyses	In sheet "PFS Parameters", cells H25:26 for Gompertz are linked to incorrect Cells. Same for Gamma distribution.	In sheet "PFS Parameters", in cell H25, replace D138 by D115. In sheet "PFS Parameters", in cell H26, replace D140 by D116.
Calculation of the drug acquisition cost per cycle.	In sheet "Model Inputs" in Cell E200, assume 365 days instead of 365.25 (as used throughout the rest of the model).	Replace 365 by 365.25
Calculation of the cost for healthcare professional visits in PSA	In sheet "Costs" in Cell G64. Calculation use the deterministic cost for health care professional (Cell F53)	In sheet "Costs" in Cell G64. Replace F53 by H53
Use of the cost for progression-free for the progressive health state in people initiating pirfenidone	In sheet "Esbriet" column BJ, use of "c_dm_pir_pre" instead of cost for progression health state	In sheet "Esbriet" column BJ, replace "c_dm_pir_pre" by "c_dm_pir_post"

Implementation on nintedanib discount in PAS analyses	Nintedanib discount applied in addition to pirfenidone discount	Nintedanib PAS discount applied to nintedanib list price. Corrected as follow: [([2151.11/60] x 2) x (365.25/4)] x (1 – PAS discount) Also corrected error in days (from 365 by 365.25) for both drugs