Evidence Review Group Report commissioned by the NHS R&D HTA Programme on behalf of NICE

Pirfenidone for the treatment of idiopathic pulmonary fibrosis

Updated CIC/AIC September 2012

This document summarises the manufacturer's submission received in December 2011 and should be read in conjunction with the supplementary document 'ERG overview of manufacturer's additional analyses', which summarises additional data from the manufacturer received in August 2012

Produced by Southampton Health Technology Assessments Centre

Authors Keith Cooper, Senior Research Fellow, SHTAC
Diana Mendes, Research Fellow, SHTAC
Jo Picot, Research Fellow, SHTAC
Emma Loveman, Senior Research Fellow, SHTAC

Correspondence to [redacted]

Date completed 21st September 2012

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 09/135/01.

Acknowledgements

We are very grateful to Dr Katherine O'Reilly, Consultant Physician, University Hospital Southampton NHS Foundation Trust and Professor Athol Wells, Consultant Physician, Royal Brompton and Harefield NHS foundation Trust who offered clinical advice to the ERG.

We also thank: Karen Welch, Information Scientist, SHTAC for commenting on the manufacturer's search strategy and checking searches, Dr Jeremy Jones, Principal Research Fellow, SHTAC for assistance with interpreting the electronic model, and Dr Jill Colquitt, Senior Research Fellow, SHTAC for acting as internal editor for the ERG report.

Declared competing interests

None from the Authors. Both clinical advisors have been unpaid members of the Intermune advisory board. Dr O'Reilly has received research travel support from Intermune.

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:

Cooper K, Mendes D, Picot, J, Loveman E. Pirfenidone for the treatment of idiopathic pulmonary fibrosis: A Single Technology Appraisal. SHTAC, 2012.

Contributions of authors

K Cooper (Senior Research Fellow) critically appraised the health economic systematic review and the economic evaluation and drafted the report; D Mendes (Research Fellow) critically appraised the health economic systematic review and the economic evaluation and drafted the report; J Picot (Research Fellow) critically appraised the clinical effectiveness systematic review and drafted the report; E Loveman (Senior Research Fellow) critically appraised the clinical effectiveness systematic review, drafted the report and project managed the review.

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

6MWD	6-Minute Walk Distance
6MWT	6-Minute Walk Test
AE	Adverse Event
ANCOVA	Analysis of Covariance
BSC	Best Supportive Care
CEAC	Cost-Effectiveness Acceptability Curve
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CRD	Centre for Reviews and Dissemination
CRDQ	Chronic Respiratory Disease Questionnaire
DLco	Carbon monoxide Diffusing Capacity
DSA	Deterministic Sensitivity Analysis
ERG	Evidence Review Group
EQ-5D	European QoL 5 dimensions
FDA	Food and Drug Administration
FVC	Forced Vital Capacity
HJC	Hugh-Jones Classification
HR	Hazard Ratio
HRQoL	Health Related Quality of Life
ICER	Incremental Cost Effectiveness Ratio
IPF	Idiopathic Pulmonary Fibrosis
ITT	Intention to Treat
L	Litres
LYG	Life Years Gained
MS	Manufacturer's Submission
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
OS	Overall Survival
PFS	Progression-Free Survival
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
ROW	Rest of the World
QALY	Quality Adjusted Life Year
QoL	Quality of Life
RCT	Randomised Controlled Trial
RR	Relative Risk
SAE	Serious Adverse Event
SD	Standard Deviation
SGRQ	St. George's Hospital Respiratory Questionnaire
SPC	Summary of Product Characteristics
SpO2	Oxygen saturation by pulse oximetry
TEAE	Treatment Emergent Adverse Event
UCSD SCBQ	University of California, San Diego Shortness of Breath Questionnaire
VBA	Visual Basic for Applications
VC	Vital Capacity
WHO QoL	World Health Organization Quality-of-Life (Questionnaire)

SUMMARY

Scope of the manufacturer submission

The manufacturer's submission (MS) does not fully reflect the scope of the appraisal issued by the National Institute for Health and Clinical Excellence (NICE). One comparator specified in the scope issued by NICE ('triple therapy': prednisolone, azathioprine, N-acetylcysteine) was not included in the MS assessment of clinical effectiveness or cost effectiveness as a systematic search for data did not identify any eligible studies. As a result the MS focuses on the efficacy of pirfenidone relative to placebo, for which four relevant randomised controlled trials (RCTs) were included, and an economic evaluation of pirfenidone compared to best supportive care (BSC).

Summary of submitted clinical effectiveness evidence

The clinical effectiveness evidence in the MS comes from:

- Four RCTs comparing pirfenidone against placebo. Two of these RCTs were undertaken by the manufacturer (the CAPACITY trials 1 and 2) and compared pirfenidone 2403mg per day to placebo (CAPACITY 2 also had a third arm using 1197mg pirfenidone). These are the primary source of evidence. The other two studies, undertaken by Shionogi, the manufacturer sponsoring pirfenidone in Japan, Taiwan and South Korea (trials SP2 and SP3), compared pirfenidone 1800mg with placebo (SP3 also had a third arm using 1200mg pirfenidone).
- Two single-arm, open label studies which were included primarily to assess the safety of pirfenidone alongside the evidence from the included RCTs.

The primary outcome differed according to the trial. In the two CAPACITY trials the primary outcome was change in percentage predicted forced vital capacity (FVC). In the SP2 trial the primary outcome was change in the lowest oxygen saturation by pulse oximetry (SpO_2) reached during the six minute walk test (6MWT), and in the SP3 trial the primary outcome was change in vital capacity (VC).

The mean decline in percentage predicted FVC was found to be statistically significantly different between pirfenidone and placebo in the CAPACITY 2 trial at 72 weeks, but not in the CAPACITY 1 trial. When analysed together (an a priori analysis) the pooled data show a statistically significant difference between pirfenidone and placebo. In the SP3 and SP2 trials pirfenidone treatment resulted in a statistically significant difference in mean decline in VC, litres, compared with placebo. Categorical analyses of the extent of decline (or improvement) in

percentage predicted FVC in the CAPACITY trials were undertaken which showed a statistically significant treatment effect for pirfenidone in the pooled analysis and CAPACITY 2, but not in CAPACITY 1.

A secondary outcome of progression free survival (PFS) assessed in the CAPACITY trials showed a similar pattern, with a statistically significant benefit seen in CAPACITY 2 but not in CAPACITY 1, however, when data from the two trials were pooled together a statistically significant benefit in terms of PFS was observed with pirfenidone. PFS was also shown to be statistically significantly improved with pirfenidone compared to placebo in the SP3 trial. The SP2 trial did not report PFS. For 6MWT the CAPACITY 1 study showed a statistically significant effect of pirfenidone whereas the CAPACITY 2 study did not. The pooled analysis for this outcome was statistically significant in favour of pirfenidone.

The manufacturers expected the proportion of participants who died to be low in the two CAPACITY trials[2] and for this reason an exploratory survival analysis was pre-specified. Although there was a reduction in the risk of death with pirfenidone in CAPACITY 2 this was not statistically significant compared to the placebo group. Risk of death in CAPACITY 1 was similar in each group and the pooled analysis showed no statistically significant difference in risk of death between the groups. IPF-related mortality (which was not defined) was used in the economic model. This did indicate a statistically significant difference in favour of pirfenidone.

For the other secondary outcomes: dyspnoea; exacerbations; respiratory hospitalisations; and [redacted] there were no statistically significant improvements with pirfenidone compared to placebo in the studies which measured them. Some adverse events were more frequently reported in those treated with pirfenidone but there do not appear to be any significant safety issues overall.

A series of meta-analyses, some of which included data from the SP2[3] and SP3[4] trials, was also presented.

The evidence seen suggests that pirfenidone appears to offer short-term benefit to patients, although the clinical significance of improvements in these outcomes is unclear.

Summary of submitted cost-effectiveness evidence

The MS includes:

i) A review of published economic evaluations of pirfenidone for IPF,

ii) An economic evaluation undertaken for the NICE STA process. The cost effectiveness of pirfenidone is compared to BSC for the treatment of IPF in adult patients with mild to moderate disease.

A systematic search of the literature was conducted by the manufacturer to identify economic evaluations of pirfenidone for the treatment of IPF. The review did not identify any relevant studies.

The economic evaluation uses an individual patient level micro-simulation model with a Markov framework developed in Microsoft Excel. The cost effectiveness model compared the total costs and quality adjusted life years (QALYs) of pirfenidone to BSC in a hypothetical cohort of patients with mild to moderate IPF. The model adopted a lifetime horizon to capture lifetime costs and health outcomes, with 24-week cycles. In the model, patients progress to more severe disease through deterioration in their clinical symptoms, i.e. FVC and 6MWT distance (6MWD). These surrogate outcomes (FVC and 6MWD) are used to estimate the risk of IPF-related mortality, the risk of hospitalisation and patients' health related quality of life (HRQoL).

Results are presented for lifetime costs and QALYs and incremental cost-effectiveness ratios (ICERs) for a cohort representing patients with mild and moderate IPF. For the base case, an ICER [redacted] is reported. Results are also presented for a subgroup for patients with a % predicted FVC < 80%.

The manufacturer's univariate sensitivity analysis showed the base case ICER was most sensitive to the discount rates for costs and outcomes, and the number of pills per day of pirfenidone. The probabilistic sensitivity analysis estimates there is [redacted] probability of pirfenidone being cost effective at a willingness-to-pay threshold of £20,000 - £30,000 per QALY gained.

Commentary on the robustness of submitted evidence

Strengths

- The MS conducted a systematic search for clinical and cost-effectiveness studies of pirfenidone. It appears unlikely that the searches missed any additional clinical effectiveness or cost-effectiveness studies that would have met the inclusion criteria.
- The RCTs comparing pirfenidone against placebo were of reasonable methodological quality, and measured a range of outcomes that are relevant to the decision problem.
- The MS appears to present unbiased estimates of the primary outcome for pirfenidone versus placebo.
- The economic model presented in the model used an appropriate approach for the disease area.
- The cost-effectiveness analysis meets the requirements of the NICE reference case.

Weaknesses and Areas of uncertainty

- A wide range of surrogate outcomes were reported across the included RCTs, and results of these varied. The clinical significance of these outcomes is uncertain with the MS offering limited discussion of the issue of clinical significance. Overall the MS provides a limited interpretation of the clinical evidence.
- Meta-analyses of all outcomes were undertaken and in the primary outcome statistical heterogeneity was observed, which despite multivariate testing by the manufacturer could not be fully explained. The methods chosen for the meta-analysis meant that the ERG could not check the data presented.
- The population within the included RCTs may not be generalisable to those presenting to secondary care in England and Wales. Based on baseline FVC scores participants in the trials were likely to be of milder IPF and few participants had the types of comorbidities expected to be seen in clinical practice.
- The MS does not provide an estimate of the clinical effectiveness of pirfenidone in relation to triple therapy (a scoped comparator intervention) in this population owing to limitations in the evidence base.
- Triple therapy was not used as comparator in the economic analysis. The MS have not attempted to use any other type of evidence that may be available for triple therapy, nor discussed how they could have included triple therapy within the model even though the evidence may not have been robust. The MS does not discuss the limitations of not fulfilling the scope, nor discuss the relevance of the triple therapy comparator to current clinical practice.

- The economic model has been coded as an individual patient simulation in Visual Basic for Applications (VBA) which has made it less accessible and more difficult to interpret and critique. It is uncertain whether the bootstrapping of the baseline characteristics and the individual patient simulation were adequately combined in order to accurately perform the cost effectiveness analysis (CEA).
- The MS has not included all model parameters in either the univariate or probabilistic sensitivity analyses and so the full uncertainty around the model results has not been shown. In particular key parameters associated with overall survival, hospitalisations, and HRQoL have been omitted.
- There is some uncertainty around the discontinuation rates reported in the MS, where the rates reported within the MS differ. The reason for these differences is unclear. Furthermore, it is unclear what the long term discontinuation rates for patients on pirfenidone treatment would be.
- The average length of stay in hospital is significantly lower in the pirfenidone group than in BSC group. The reasons for the differences between the two groups are unclear and are not discussed in the MS.

Summary of additional work undertaken by the ERG

The ERG conducted the following additional analyses:

a) Variation of the regression coefficients used to estimate treatment effect
b) Variation of the hazard ratio for IPF-related mortality
c) Variation of patients' HRQoL
d) Assuming the same hospital length of stay for both treatment arms

Results show the substantial sensitivity of the ICER to the variation of parameters for patients' survival, hospitalisation, and HRQoL. The ICER ranged from [redacted] gained (when the upper limit of all coefficients of the regression equations were used) to [redacted] gained (using the upper limit of the IPF-related mortality hazard ratio).

1 INTRODUCTION TO ERG REPORT

This report is a critique of the manufacturer's submission (MS) to NICE from InterMune UK and Ireland on the clinical effectiveness and cost effectiveness of pirfenidone for idiopathic pulmonary fibrosis (IPF). It identifies the strengths and weakness of the MS. Clinical experts were consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the MS was requested from the manufacturer by the ERG via NICE on 22nd December 2011. A response from the manufacturer via NICE was received by the ERG on 20th January and this can be seen in the NICE evaluation report for this appraisal.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

The background appears to be a clear and accurate overview of the disease and the population is relevant to the NHS. The population is restricted to people with mild to moderate IPF as per the NICE scope (although the MS incorrectly suggests the scope is any IPF) which clinical advice to the ERG suggests are the most appropriate group for this type of treatment.

2.2 Critique of manufacturer's overview of current service provision

Based on clinical advice to the ERG a clear and accurate overview of current service provision is provided in the MS.

2.3 Critique of manufacturer's definition of decision problem

Population

The population described in the decision problem is appropriate for the NHS.

Intervention

The description of pirfenidone is reflective of its use in the UK. The clinical advice to the ERG is that not all patients would tolerate the full dose as defined in the license, and that at least 10-15% of individuals would stop taking pirfenidone owing to side effects.

Comparators

Two comparators are noted in the NICE scope, triple therapy (prednisolone, azathioprine, N-acetylcysteine) and best supportive care (BSC). The manufacturer comments that no other treatments for IPF have been subject to rigorous testing through RCTs. Clinical advice to the ERG suggests that triple therapy is unlikely to be a main comparator. There has been previous consensus in the clinical community suggesting it has only limited use, and a recent trial[1] has found an increased risk of mortality and side effects compared to placebo and has terminated this intervention arm of the study. N-acetylcysteine (NAC) is often used as a treatment on its own, however this was not in the NICE scope and is not discussed by the manufacturer. Clinical advice is that BSC is an appropriate comparator.

Outcomes

The outcomes reported appear to be valid. The primary outcome of forced vital capacity (FVC) is the best current surrogate for patient outcome. A threshold for clinically significant change in the FVC has been stated by the manufacturer as being the decline in percentage predicted FVC of ≥ 10% from baseline. Clinical advice to the ERG concurs that this is appropriate for change seen in an individual patient (see below for further discussion). The ERG clinical advice also suggests that progression free survival (PFS) is a useful measure as it combines a number of surrogate outcomes. Although it should be noted that PFS definitions (i.e. which outcomes are combined and what thresholds used) may differ between studies.

Economic analysis

The economic evaluation appears to be appropriate for the NHS.

Other relevant factors

The manufacturer's note that post-hoc subgroup analyses are presented. These are critiqued in the sections below.

There are no special considerations related to equity or equality.

3 CLINICAL EFFECTIVENESS

3.1 Critique of manufacturer's approach to systematic review

3.1.1 Description of manufacturer's search strategy

The manufacturer's literature searches were checked by an information scientist and are considered to be reasonably comprehensive, fit for purpose and reproducible. The narrative description of the approach to the literature searches is good. The documentation of the search strategies contain a balance of descriptor and free text terms with adequate truncation, mapping to subject headings, correctly linked sets and comprising acceptable randomised controlled trial (RCT), non- RCT, adverse event (AE), quality of life (QoL), and cost search filters.

Supplementary searching of in-house company databases and reference checking is also recorded by the manufacturer. Multi-file cross searching was undertaken rather than searching the databases separately. This results in some repetition in the indirect mixed treatment comparisons search section but this is unlikely to affect the results.

The ERG ran additional searches on Medline, Medline In Process (MEIP) and Embase to establish whether there were any further RCTs for triple therapy (prednisolone, azathioprine, N-acetylcysteine) of relevance (see below).

The MS makes reference to on-going trials from the clinicaltrials.gov database, however no overt documentation of a strategy or sources used to identify these trials was identified in the MS. The ERG ran searches on controlled-trials.com and UKCRN Portfolio Data checking for further on-going trials data (see below).

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection

To be included in the systematic review, trials had to meet the eligibility criteria that were provided in MS Table B5.2.1 (MS p51). There are some differences between the criteria in MS Table B5.2.1, those of the decision problem, and the licensed indication:

- The population described in the final scope (NICE), the decision problem in the submission and the licensed indication for pirfenidone the population is restricted to people with mild to moderate IPF. The eligibility criteria for the MS systematic review are stated as people with IPF.
- The comparator was not used to determine eligibility for the SR.

- The total maximum daily dose of pirfenidone recommended in the licensed indication is 2403 mg/day but dose is not noted in the NICE scope, the decision problem addressed, or the inclusion criteria.
- Study design was provided as an inclusion criterion for the MS systematic review (limited to RCTs with open label extensions with parallel design, or comparing different doses or schedules of the drug also considered). It is not usual to state study design in the decision problem.

No limits were placed on inclusion criteria relating to the quality of RCTs (both blinded and non-blinded RCTs eligible) or other study types.

A flow diagram to show the numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions is provided (MS Figure B5.2.1, p52). However there appear to be some errors in this:

- the diagram shows that 188 titles and abstracts failed to meet inclusion criteria but the sum of studies provided for each reason comes to 192. It could be that studies were excluded for more than one reason but this is not stated
- if 188 of the 207 studies were excluded as indicated in the diagram this would have left 19 studies. Then adding the further 6 items identified from the InterMune database would have given a total of 25 full publications. However 26 full publications are noted in the flow chart and this does correspond to the number of primary and secondary sources in MS Table B5.2.2.2, p52
- the chart shows that three studies identified from the InterMune database were excluded. It is unclear whether these three were among the six additional items noted above, or whether they were an additional three

Two pirfenidone studies initiated by the original developer of pirfenidone (Marnac Inc) were excluded from the MS (MS p56/57). It is not clear whether these appear in the flow chart of studies. Both studies were double blind randomised studies each of which had enrolled a total of 52 participants. In one RCT the comparison was pirfenidone 40 mg/kg/day up to 3600 mg/day versus prednisone 0.33 mg/kg/day.

[redacted]

[redacted] The other RCT was a comparison of pirfenidone 40 mg/kg/day up to

3600 mg/day versus placebo.

[redacted] The manufacturer does not discuss whether there was any potential for bias by excluding these studies.

3.1.3 Identified studies

Three full publications were identified in the MS and these encompassed four RCTs which are included together with multiple secondary publications. The key features of the four included RCTs are shown in Table 1. There is also one abstract for an extension study from two of the RCTs. The manufacturer provided electronic copies of the RCT reports. Summary details of the included RCTs were included in a number of separate tables:

- intervention, population and outcome descriptions are in MS Table B5.3.1, p60, with additional details on eligibility criteria in MS Table B5.3.3, p63
- patient numbers are reported in MS Table B5.3.2 and in the patient flow diagrams MS pages 76 to 78, p62
- statistical analyses (including information on sample sizes and power calculations) for the individual trials are reported in MS Table B5.3.7, p63

Table 1: Summary of the key features of the four included RCTs

	Study arms	Number enrolled	Primary outcome measure	Length of follow up
CAPACITY-1 (PIPF-006)[2a]	Pirfenidone 2403 mg/day	171	Change in % predicted FVC from baseline to week 72	72 weeks
	Placebo	173		
CAPACITY-2 (PIPF-004)[2a]	Pirfenidone 1197 mg/day	87	Change in % predicted FVC from baseline to week 72	72 weeks
	Pirfenidone 2403 mg/day	174		
	Placebo	174		
SP2[3]	Pirfenidone 1800 mg/day	72	Change in lowest SPO2 reached during the 6MWT	36 weeks
	Placebo	35		
SP3[4]	Pirfenidone 1200 mg/day	55	Change in vital capacity (VC[b]) from baseline to week 52	52 weeks
	Pirfenidone 1800 mg/day	108		
	Placebo	104		

[a]Hereafter referred to as CAPACITY-1 and CAPACITY-2 when discussed separately, or CAPACITY trials when discussed together.
[b]In restrictive lung disease such as IPF vital capacity and forced vital capacity can be treated as the same outcome.

The two CAPACITY RCTs[2] were analysed by the intention to treat (ITT) principle, for the other two included trials (SP2[3] and SP3[4]) it is not stated that analyses were ITT. The flow chart for SP3 (Fig B5.3.3 MS p78) suggests analyses were not ITT, for SP2 it is not clear (Fig B5.3.4, MS p78). In addition, outcomes from the two CAPACITY trials were pooled. Meta-analyses were conducted which are described in this report, section 3.1.7.

The CAPACITY trials[2] were sponsored by the manufacturer InterMune (who are conducting clinical development for the rest of the world other than Japan) and SP2[3] and SP3[4] were sponsored by Shionogi (who licenced the development rights for Japan, Taiwan and South Korea).

In addition to the RCTs, non-RCTs were also identified using a separate search strategy (provided in MS Appendix 9.6) however no flow chart was provided for identification of studies from the search results in either the MS or Appendix 9.6. The literature search identified two published non-RCTs, a prospective open-label phase 2 study (Raghu et al 1999[5]) and an open label compassionate use study (Nagai et al 2002[6]). Neither of these published non-RCTs contributes data to the MS. Two ongoing non-RCTs were also identified (although not explicitly stated, it is presumed that these studies were identified from InterMune's own database). These ongoing studies were the open label extension study RECAP[7] (PIPF-012) for the CAPACITY trials (only an abstract published), and a non-controlled study of tolerance (PIPF-002[8]). MS Table 5.2.6, p58, provides the justification for inclusion of these non-randomised studies, with both included because they provide additional safety data on the use of pirfenidone. Limited safety data are however presented from these non-RCTs (see Section 3.3.11 of this report).

There are differences between the sections of the MS that present the non-RCT evidence. MS section 5.2.7 (MS p 57) lists the four relevant non-RCTs noted above. In contrast, MS section 5.8, p145, only describes the two studies (RECAP PIPF-012[7] and PIPF-002[8]) included in the MS section on adverse events.

The baseline data from the RCTs presented in the MS have been checked against the trial publications. A few minor errors were found but there were no major problems. From a visual inspection of MS Table B5.3.5, p68, participants and control groups in each study appear broadly similar. The paper[2] which describes the CAPACITY trials states that there were no pronounced baseline imbalances between treatment groups within each study (the paper does

not report statistical testing). The SP2 publication[3] describes the groups as similar, and the SP3 paper[4] states there were no significant differences in baseline characteristics between groups except for smoking history.

The MS (p 65) states there were some differences within each study but the ERG assumes this is an error and that this should read differences between trials because it is differences between trials that are then discussed in the MS and which the ERG has summarised in the following paragraphs.

The MS highlights some differences in patient characteristics between the CAPACITY-1 and CAPACITY-2 trials[2] (MS p65). There were differences between the trials in the proportion of participants enrolled from the US (approximately 66% of the CAPACITY-2 trial but approximately 87% of the CAPACITY-1 trial). The difference in proportions of US participants led to other differences between the trials because the US participants differed from EU and rest-of-the-world (ROW i.e. non EU/US) participants in several aspects:

- weight. Men and women in the US weigh more than men and women in the ROW
- lung biopsy. This was more common among participants in the USA than participants in the ROW but the MS speculates that this was due to differences in reimbursement since the additional biopsies were not required for study entry
- more than 4 times as many USA participants used supplemental oxygen at baseline compared to participants from ROW.
- in CAPACITY-1 participants in the USA walked less in the six-minute walk test (6MWT) than those in the ROW whereas in CAPACITY-2 the mean 6MWT distance was similar for USA participants and ROW participants

These differences were reported in the MS but are not reported in the CAPACITY trials publication.[2]

Other differences in patient characteristics which the ERG has observed between trials reported in MS Table B5.3.5 are:

- the majority of participants in the CAPACITY trials were white (over 95%) whereas it is presumed that the majority of participants in SP2 and SP3 were East Asian since the SP2 and SP3 studies were conducted in Japan.

- fewer participants had never smoked in the SP2 and SP3 trials than in the CAPACITY trials (range across study arms of 6-22% versus 29-37% respectively)
- fewer participants had been diagnosed with IPF for a year or less in the SP2 and SP3 trials than in the CAPACITY trials (range across study arms of 17-39% versus 47-62% respectively)
- the mean baseline carbon monoxide diffusing capacity (DLco) was higher in the SP2 and SP3 trials than in the two CAPACITY trials
- fewer surgical lung biopsies were reported in the SP2 and SP3 trials than in the two CAPACITY trials (this may be related to the sites of these studies i.e USA vs non-USA)

All the included RCTs appear to meet the inclusion criteria and the ERG's view is that it is likely that all relevant RCTs have been identified on searches.

In addition to the completed and published RCTs the MS has identified three ongoing studies. An ongoing RCT and an ongoing non-randomised study are reported in MS Table B5.2.4 p56. The ongoing RCT is ASCEND (PIPF-016), an Intermune sponsored study, which will compare pirfenidone with placebo in patients with IPF. The severity of IPF is not stated but the ERG presume this will be mild to moderate IPF in accordance with the licenced indication of pirfenidone. This study aims to enrol 500 participants and the estimated primary completion date is December 2012. The ongoing non randomised study, RECAP[7] (PIPF-012), is the open-label extension study for participants from the two CAPACITY trials. This began in August 2008 and is due to end in July 2012. Interim results for safety are reported in the MS p158.

A further ongoing non randomised study, PIPF-002,[8] is listed in MS Table B5.2.6 (p58) as a relevant non-RCT. This is a non-controlled study of tolerance and interim safety results are included in the submission.

Searches for ongoing studies conducted by the ERG have identified one further ongoing study, recently received (December 30th 2011) by the ClinicalTrials.gov register. The title of this phase II study is 'A multicentre, randomised double-blind placebo-controlled trial for the safety and efficacy of pirfenidone in the treatment of idiopathic pulmonary fibrosis (IPF)" and it is listed as

being due to start in January 2012 in China (expected enrolment not stated). The study sponsor is given as the Beijing Kawin Technology Share-Holding Company Ltd.

3.1.4 Description and critique of the approach to validity assessment

The MS quality assessed the included trials using the NICE criteria and presents a summary in MS Table B5.3.9, p79, with further detail available in MS Appendix 3. CAPACITY-1 and CAPACITY-2[2] were assessed together as one, presumably because the same methods were applied in both studies although this is not explicitly stated in the MS.

The ERG agreed with the industry assessment of the CAPACITY trials in the MS (Table 2). For the SP2 and SP3 trials the ERG assessment differed from the industry assessment for questions 2 (concealment of allocation), 4 (blinding) and 7 (ITT analysis). From the information provided in the trial publications the ERG were not certain that the intervention to be allocated could not have been known in advance. Although SP2 and SP3 are described as double blind it is not certain whether all involved were blinded and the analyses described for SP2 and SP3 do not conform to the strict definition of ITT because some randomised participants were omitted. However, the proportion of randomised participants omitted from the analyses was less than 2% and 3% in SP2 and SP3 respectively and the risk of bias is therefore likely to be low.

Table 2: Manufacturer and ERG assessment of trial quality

		CAPACITY-1[2]	CAPACITY-2[2]	SP2[3]	SP3[4]
1. Was randomisation carried out appropriately?	MS:	Yes		Yes	Yes
	ERG:	Yes	Yes	Yes	Yes
Comment:					
2. Was concealment of treatment allocation adequate?	MS:	Yes		Yes	Yes
	ERG:	Yes	Yes	Uncertain	Uncertain
Comment: Information about concealment of treatment allocation is not available in the published papers for SP2 and SP3. The information in the MS (MS Table B5.3.1, MSp61) is not sufficient to determine whether allocation concealment was adequate.					
3. Were groups similar at outset in terms of prognostic factors?	MS:	Yes		Yes	Yes
	ERG:	Yes	Yes	Yes	Yes*
Comment: * Smoking history differed but overall groups appear similar.					
4. Were care providers, participants and outcome assessors blind to treatment allocation?	MS:	Yes		Yes	Yes
	ERG:	Yes	Yes	Uncertain	Uncertain
Comment: Outcome assessors are not explicitly mentioned for any of the trials. For the CAPCITY trials the ERG presumes that care providers/investigators (who were blinded) were assessing outcomes. For SP2 which is described as double blind the trial publication[3] states that the radiologists and Study Coordinating Committee were blinded for outcomes based on high resolution computed tomography (HRCT) scans but					

no further information is provided about blinding. For SP3 the trial publication[4] describes the study as double blind but no additional details to describe this are provided.					
5. Were there any unexpected imbalances in drop-outs between groups?	MS:	No		No	No
	ERG:	No	No	No	No
Comment: Although there were some differences these were unlikely to have been unexpected. In general there were more adverse events in the pirfenidone groups in all the studies, whereas in the placebo groups there were more exacerbations (SP2[3]), disease progression (SP3[4]) or deaths (CAPACITY studies[2])					
6. Is there any evidence that authors measured more outcomes than reported?	MS:	No		No	No
	ERG:	No	No	No	No
Comment:					
7. Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	MS:	Yes		Yes	Yes
	ERG:	Yes	Yes	Uncertain	No
Comment: Trial SP2 excluded 2 randomised participants (one from each group) from the efficacy analyses because they had violated inclusion criteria (no further details provided). In trial SP3[4] eight participants were excluded from the analysis for reasons described as 'No medication' and 'No data available'. No methods are provided in the published paper for SP2 to describe how missing data were accounted for. However the MS indicates that the principle of last observation carried forward (LOCF) was used and this is also the method described for SP3. The MS itself (MS p143) indicates that use of LOCF in accounting for missing data could lead to favouring the treatment arm with earlier drop outs in a progressive disease such as IPF.					

3.1.5 Description and critique of manufacturer's outcome selection

The outcomes selected by the manufacturer seem appropriate and match the NICE scope/decision problem.

The primary outcomes differed according to the trial:

- in CAPACITY-1 and CAPACITY-2 trials[2] the primary outcome was change in percentage of predicted FVC from baseline to week 72
- in the SP2 trial[3] the primary outcome was change in the lowest oxygen saturation by pulse oximetry (SpO_2) reached during the 6MWT
- in the SP3[4] trial the primary outcome was change in vital capacity (VC) from baseline to week 52

Secondary outcomes reported by the trials also varied both in the specific outcomes reported and the way in which the outcome was reported (e.g. change from baseline, mean value at time point) or defined (e.g. PFS). Types of secondary outcome commonly reported include measures of pulmonary function, exercise tolerance, disease progression, and QoL.

Some outcomes reported by the trial publications were not reported in the MS, or not reported in as much detail (although these do not appear to be key outcomes). For CAPACITY-1 and CAPACITY-2 trials[2] the following differences were identified:

- the time frames for reporting mortality outcomes differ (MS p105)
- some adverse event reporting in the published paper[2] is presented for the two trials separately, whereas the MS only reports the pooled analysis
- the MS does not report on substantial laboratory abnormalities (grade 4 or a shift of 3 grades) which are reported in the published paper[2] (p1766)

For the SP3 trial the following differences were identified:

- the results of the comparison between low-dose pirfenidone and placebo on changes in VC are not reported in as much detail in the MS as in the published paper
- the published paper reports more detail for the low dose pirfenidone comparisons for PFS, and lowest SpO2 during 6MWT than are reported in the MS
- the adverse events leading to discontinuation are specified in the published paper (Table 2 of the publication - although the numbers in this table don't match numbers in text) but are not reported in the MS which reports the overall n (%) discontinuing in each group due to AEs

For the SP2 trial the differences identified were:

- the MS does not report standard deviations (SDs) for primary end point results (but otherwise reports the same data as the published paper)
- data reported for secondary outcomes are less detailed in the MS. The MS does not report SDs for decline in VC and doesn't reproduce Fig.2 from the published paper which is the categorised analyses of the lowest SpO_2 during the 6MWT.
- the MS does not report on the correlations between changes in the lowest SpO2 and the SpO2 area and changes in VC, total lung cancer, and DLco, nor on the correlation between change in lowest SpO2 and change in SpO2 area during 6MWT
- the MS states that no specific QoL measures were used in the SP2 study (MS p72). However the ERG has found that in the trial publication for SP2[3] the Chronic Respiratory Disease Questionnaire (CRDQ) score and the Hugh-Jones

Classification (HJC) score were measured to assess participants' perceived quality of life during the study

The trial outcomes/data not reported by the MS appear not to be key outcomes and are unlikely to affect the overall conclusions of the MS.

Adverse events are reported for all trials whereas QoL was reported in the MS only for the CAPACITY trials.[2] These trials used the St George's Respiratory Questionnaire (SGRQ) and the World Health Organisation Quality of Life (WHO-QOL) instrument. At the time of the CAPACITY trials there was not a disease-specific QoL measure. As the measures used are not disease specific they may not fully capture HRQOL for IPF patients. The SGRQ appears to be validated in chronic obstructive pulmonary disease (COPD) (but not IPF), the WHO-QoL is also validated but is not a disease specific measure. The MS states that no specific QoL measures were used in the SP2 and SP3 studies (MS p72). However, as noted above, the ERG has found that in the trial publication for SP2[3] the CRDQ score and the HJC score were used. The CRDQ is validated (for COPD) and the HJC appears to be a clinician rated classification.

3.1.6 Description and critique of the manufacturer's approach to trial statistics

The MS reports trial results for the relevant outcome measures. The units of measurement, size of effect, numbers in analysis and whether an ITT analysis or not plus a discussion/justification of clinically important differences are reported for the majority of outcomes. The 95% confidence intervals (CI) (or an alternative such as SDs), however, are frequently not reported, particularly for mean % change outcomes. The 95% CIs are reported for hazard ratios (HR) and some SDs are reported for the mean % change outcomes from SP3. In a few cases SDs which are reported in a trial publication are not reported in the MS. The MS discusses the definition for a 'treatment adherent population' on page 93, however, the ERG were not able to identify any outcome data presented for the treatment adherent population.

Some interim data on adverse events are reported for the ongoing non-randomised studies RECAP (PIPF-012)[7] and PIPF-002.[8] It is clear in the MS that these data are from interim analyses.

3.1.7 Description and critique of the manufacturer's approach to the evidence synthesis

A narrative review of evidence is provided in the MS, together with an overview of two published meta-analyses (Noble et al.[9] and King et al.[10]). In general the tabulated data and the narrative data presented in the MS reflect the data presented in the trial publications. In the summary of outcomes (Section 3.3) the ERG note any issues identified on checking the data.

For the CAPACITY trials[2] the MS presents data for each trial separately and for an apriori pooled analysis. The pooled analyses were also presented in the trial publication. In a response to a request for clarification over the methods used for the pooled analysis, the manufacturer provided the rationale and detailed methods for these analyses. No discussion has been given in the MS of the rationale for having two separate trials however.

Various meta-analyses are provided, the outcomes of which are all marked CIC. These generally combined outcomes in two ways:

1) data from three studies (CAPACITY 2;[2] CAPACITY 1;[2] and SP3[4])
2) data from two studies (CAPACITY 2;[2] CAPACITY 1[2]) as the MS states that these were used for the cost effectiveness model (see Section 4.2.2 however).

The CAPACITY 1[2] and CAPACITY 2[2] studies were very similar and are therefore appropriate for meta-analysis. The SP2[3] and SP3[4] studies were undertaken in Japanese participants and the dose of pirfenidone used was different from the CAPACITY trials.[2] However, these doses should be largely comparable because of the different mean weight of the populations involved (the dose in the UK licensed indication was calculated by the manufacturer after adjusting the doses used in the Japanese studies to account for greater mean weight). One of the previously published meta-analyses included all four studies.[10] However, the manufacturer has not undertaken meta-analyses including the SP2[3] study with the exception of the analysis of mortality. No explanation has been provided by the MS for not including SP2[3] although the ERG assume this was related to the difference in the length of follow-up.

Outcomes pooled were change in FVC/VC (the MS also included a meta-analysis of CAPACITY 2[2] and SP3[4] as a 'low-dose' comparison); change in 6MWT; PFS; mortality (CAPACITY 1;[2] CAPACITY 2;[2] SP3[4] and SP2[3] combined); QOL and AEs. Other outcomes (dyspnoea and worst

SpO2) were meta-analysed with data presented in Appendix 21 (Figures 13 and 15 respectively) of the MS.

For FVC/VC the manufacturer found [redacted] and substantial heterogeneity when combining the two CAPACITY studies (I^2 = 71%). The manufacturers have explored the possible reasons for this statistical heterogeneity. They discuss the fact that the selection criteria, baseline characteristics, adherence and use of concomitant medications were similar between the two studies, and have undertaken a multivariable rank ANCOVA in an attempt to identify if any study or covariates on FVC change were potentially relevant. They also undertook a meta-analysis of the FEV1/FVC ratio for the two studies. The results of these analyses [redacted] explain little of the variance found and study variables remained an independent predictor of FVC change. The manufacturer states that the heterogeneity is possibly related to the fact that the CAPACITY 2 study met its primary end-point but the CAPACITY 1 did not. The MS states that the reasons for this are likely to be multifactorial and include some baseline imbalances in rates of FVC decline. [redacted] Although not discussed there does not appear to be any statistical heterogeneity for any of the other outcomes pooled.

Both random and fixed effects models were used. The MS states that for comparisons including either of the SP2[3] or SP3[4] studies random effects models were most appropriate [redacted] However, fixed effect models were also presented. For the meta-analysis of CAPACITY 1[2] and CAPACITY 2[2] the manufacturer states fixed effect models were appropriate as the trials had identical selection criteria. However random effects models were also presented.

Results were presented with relative differences (standardised mean differences for FVC/VC, 6MWT, QoL, Risk ratios for PFS, mortality, adverse events). No discussion was provided in the MS as to the choices of measures used in the meta-analysis and the ERG requested further clarification from the manufacturer. The response from the manufacturer explained that a rank ANCOVA was used for the statistical analyses of the primary outcome data in the CAPACITY trials.[2] The manufacturer reports that there are two key difficulties with using the mean change from baseline from these trials which would have the potential to skew the data. Any deaths

during follow-up would be applied at 72 weeks with a 0% predicted FVC, and also there would be no adjustment for baseline % predicted FVC. The manufacturer therefore used the least squares mean from the rank ANCOVA for the measure of change in % FVC for the CAPACITY trials in the meta-analysis. No attempt to test alternative approaches to these analyses, in order to see what effects these would have on the analysis, was reported in the MS. The ERG has been unable to check the data presented for the CAPACITY trials in the various meta-analyses and would suggest that using the data provided as mean change from baseline for the CAPACITY trials may reduce the effect sizes seen in the various meta-analyses.

The method of analysis for SP3 used the last observation carried forward and in this case the actual change from baseline was able to be used in an ANCOVA where data were available. The manufacturer therefore used the adjusted mean change from baseline in litres for SP3 in the meta-analysis of FVC/VC. It is not clear whether the manufacturer would have sufficient data for the % predicted VC to be meta-analysed. The mean % predicted change in VC for SP3 was presented in the FDA briefing although without data for the SDs.

Treatment effects were presented in a series of forest plots which included individual point estimates and confidence intervals from the included studies, and the summary estimate including confidence intervals.

There is no direct discussion of any sensitivity analysis in the MS, however, the use of different combinations of studies is essentially a sensitivity analysis. There were no subgroups analysed.

The ERG asked for clarification on the processes undertaken for assessing study eligibility, data extraction or quality assessment for the main evidence review as this was not presented in the MS. The processes described in the subsequent response (see NICE evaluation report) from the manufacturer appear to be appropriate.

Indirect comparison

The manufacturer attempted to identify studies which could be used for an indirect comparison. A CONSORT style flow-chart for the identification of studies is presented which indicates that there were 11 potentially eligible studies but six of these were excluded (summaries are provided for these together with justifications for exclusion) leaving five potential studies. Inclusion criteria for the potential indirect comparison are provided however these do not have a

stated intervention. The eligible comparators were reported to be BSC or triple therapy (prednisolone, azathioprine, N-acetylcysteine). The stated population and outcomes are relevant to the NICE scope. The MS reports the processes for assessing studies for inclusion.

The justification for exclusion of five of the six excluded studies appear to be based upon criteria that are not defined (e.g. dose of pirfenidone used, low number of participants, short-term study) although the data presented on these studies (MS Table B5.7.2, p141) suggest that these would not meet inclusion criteria for the interventions used (although the MS did not have one). The remaining excluded study is an ongoing RCT which included a triple therapy arm versus placebo, however this study has been terminated and no results are presented in the publications cited.

Five publications were stated to be eligible. Three of these are the pirfenidone RCTs (four studies within three publications: CAPACITY trials,[2] SP2,[3] SP3[4]) included in the main evidence summary, the remaining two publications relate to the IFIGENIA study[11] which compared triple therapy with 'double therapy' corticosteroids and azathioprine. There was no placebo comparison in this study and therefore it would not be appropriate for indirect comparison with the pirfenidone studies. The manufacturer does not explicitly state this is the case.

3.2 Summary statement of manufacturer's approach

The quality of the MS based on CRD criteria[12] for a systematic review as assessed by the ERG is shown in Table 3.

Table 3: Quality assessment (CRD criteria) of MS review

CRD Quality Item: score Yes/ No/ Uncertain with comments	
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Yes. Inclusion and exclusion criteria are clearly stated.
2. Is there evidence of a substantial effort to search for all relevant research? i.e all studies identified	Yes.
3. Is the validity of included studies adequately assessed?	Unclear. The MS presents information regarding key aspects of quality of the four included studies, using NICEs questions. No discussion is made as to the quality of the studies however and quality assessments in the MS are not considered further in relation to the synthesis and interpretation of the findings of the included studies. The validity of non-RCT studies was not assessed.

4. Is sufficient detail of the individual studies presented?	Yes.
5. Are the primary studies summarised appropriately?	Yes.

The systematic review is of reasonable quality according to CRD criteria and the submitted evidence appears to generally reflect the decision problem defined in the MS with one exception. There is no evidence for the comparator of triple therapy. However, the ERG note that it is unlikely there are any completed RCTs of triple therapy and the only ongoing trial of relevance has had the triple therapy arm terminated.

Overall therefore the risk of systematic error in the systematic review is low.

3.3 Summary of submitted evidence

In this section of the report, the ERG concentrates on the main outcomes of the included RCT evidence of pirfenidone treatment. Data have been checked by the ERG and summarised for each of the key outcomes below. For many outcomes the MS reports data at interim time points, however, the ERG have presented the data from the study endpoints only. The ERG report data from study arms using the standard doses of pirfenidone only (2403mg in CAPACITY trials, 1800mg SP2 and SP3 trials). There were a few differences between the data presented in the MS and the data in the study publications; however these were generally minor discrepancies. The data presented in the tables below are the ERG checked data and data presented in italics have been estimated by the ERG. The MS also presented limited data from the non-RCT extension studies (see MS page 158-159).

The MS presents summary data from all four included RCTs (and the pooled CAPACITY trial data[2]) in a summary table (B5.5.1, pp 81-92) and then subsequently presents data by study (data from the two CAPACITY trials[2] and the pooled data; data from SP3[4]; data from SP2[3]). The MS then presents a series of meta-analyses. The ERG presents data by study outcome, including any meta-analyses reported.

3.3.1 Summary of results for FVC

The two CAPACITY trials[2] report data for % predicted FVC and the SP3[4] and SP2 present VC, litres (L). Data are presented in Table 4.

The mean change from baseline in % predicted FVC was seen to favour pirfenidone compared to placebo at 72 weeks in the CAPACITY 2[2] study (p-value 0.001). In the CAPACITY 1[2] study no statistically significant difference was observed. The pooled CAPACITY study data[2] however showed a statistically significant difference in % predicted FVC. Statistically significant differences between groups were observed in the mean change in VC in the SP3[4] study and the SP2[3] study at 52 weeks and 36 weeks respectively. Data for the % predicted FVC change are presented in a figure in the MS (figure B5.5.7) for the SP3 trial[2] however, this does not appear to give the same data as reported in the text of the MS (this may be due to an error in the charting of the graph as it doesn't appear to align with the axes). Overall the differences in mean changes of FVC observed between pirfenidone and placebo were small and it is not clear whether these differences translate to a clinically significant effect (discussed in Section 3.4).

Meta-analyses of the mean change in FVC/VC were presented in the (MS p126) and the summary results from these are shown in Table 5 and Table 6 below.

[redacted]

Results for the meta-analysis of data from only the two CAPACITY trials[2] was seen to favour pirfenidone only when a fixed effect model was used. However, heterogeneity was seen to be substantial. As stated in section 3.1.7 above, the MS undertook a number of analyses in an attempt to understand the cause of this heterogeneity. Results suggest that this was because the CAPACITY 2[2] study met its primary endpoint but the CAPACITY 1 study did not. The MS point out that the reasons for this are likely to be multifactorial and despite extensive analysis of the data and discussion with IPF experts no clear explanation for this can be identified.

The ERG note that the mean change values entered for the two CAPACITY trials[2] come from the least squares mean generated from a rank ANCOVA rather than the mean change in FVCs presented in the trial reports and the ERG have therefore been unable to check these data.

[redacted]

Table 4: FVC/VC taken from MS, trial reports and EPAR, *or estimated by ERG*

Study	Baseline FVC	FVC study end	Mean change	Difference in mean change pirfenidone vs. placebo[a], p-value
CAPACITY2[2] study endpoint: 72 weeks				
Pirfenidone 2403mg/day, n=174	74.5% (SD 14.5)	66.6% (SD 21.77)	-8.0% (SD 16.5)	*4.4%*, p=0.001
Placebo, n=174	76.2% (SD 15.5)	63.9% (SD 26.31)	-12.4% (SD 18.5)	
CAPACITY1[2] study endpoint: 72 weeks				
Pirfenidone 2403mg/day, n=171	74.9% (SD 13.2)	65.9% (SD 23.53)	-9.0% (SD 19.6)	*0.6%*, p=0.501
Placebo, n=173	73.1% (SD 14.2)	63.6% (SD 25.06)	-9.6% (SD 19.1)	
Pooled CAPACITY1 trials[2] study endpoint: 72 weeks				
Pirfenidone 2403mg/day, n=345	*74.70%*	*66.25%*	-8.5%	*2.5%*, p=0.005
Placebo, n=347	*74.65%*	*63.75%*	-11.0%	
SP3[4]	Baseline VC, L	52 weeks VC, L	Adjusted mean change, L	
Pirfenidone 1800 mg/day, n=108	2.40 (SD 0.64)	2.36 (SD 0.73)	-0.09 (SE 0.02)	*0.07L*, p=0.042
Placebo, n=104	2.47 (SD 0.70)	2.42 (SD 0.75)	-0.16 (SE 0.02)	
SP2[3]	Baseline VC, L	36 weeks VC, L	Mean change in VC, L	
Pirfenidone 1800mg/day, n=72			-0.03	*0.1L*, p=0.037
Placebo, n=35			-0.13	

[a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

Table 5: Outcomes from meta-analysis FVC CAPACITY trials;[2] SP3[4]

[redacted]				

[a]a standardised mean across studies giving the average difference in standard deviations for the different measures of VC.

Table 6: Outcomes from meta-analysis FVC CAPACITY trials[2]

	Standardised mean difference[a]	95% Confidence interval	P value for test of overall effect	I^2
Random effects	0.20	-0.08, 0.48	0.16	71%
Fixed effect	0.20	0.05, 0.35	0.010	71%

[a]a standardised mean across studies giving the average difference in standard deviations for the different measures of VC.

3.3.2 Summary of results for PFS

Progression-free survival was reported by the two CAPACITY trials[2] and the SP3[4] trial. Data are presented in Table 7. The definition of PFS varied slightly between the CAPACITY trials and the SP3 trial. In the CAPACITY trials PFS was defined as the time to confirmed ≥10% decline in % predicted FVC, ≥15% decline in % predicted DLco or death. In the SP3 trial PFS was defined as time to decline from baseline VC ≥10% or time to death.

In the CAPACITY 2[2] trial pirfenidone 2403mg/day improved PFS at 72 weeks by reducing the risk of death or disease progression by 36% in comparison to placebo (p-value 0.0235). In contrast no statistically significant difference in PFS was observed between the 2403mg/day pirfenidone and placebo groups in the CAPACITY 1[2] trial (p-value 0.355). When results from the two CAPACITY trials were combined in the pooled analysis a statistically significant difference in favour of the 2403mg/day pirfenidone groups was seen (risk reduced by 26%, p-value 0.025). PFS was also reported for the SP3[4] trial and was found to favour 1800mg/day pirfenidone in comparison to placebo at 52 weeks (risk reduced by 55%, p-value 0.028). The MS presents Kaplan-Meier estimates of PFS for the individual trials and for the pooled analysis of the CAPACITY trials. Censored data are not marked on any of these figures and the methods of censoring are not provided, although they are described as 'appropriate' (MS p73).

The summary results from the meta-analyses of PFS that were presented in the MS are shown in Table 8 and Table 9 below.

[redacted]

Results for the meta-analysis of data from only the two CAPACITY trials[2] were also seen to favour pirfenidone when [redacted] random effects model was used and the level of statistical heterogeneity identified was very low (I^2 3%).

Table 7: PFS outcomes taken from MS

Study	Death or Disease progression	Disease progression - decline in	Disease Progression - decline in	Death before disease progression	HR (95% CI) p-value

		FVC	DLco		
CAPACITY2,[2] study endpoint: 72 weeks					
Pirfenidone 2403mg/day, n=174	45/172[a] (26.2%)	28 (16.3%)	9 (5.2%)	8 (4.7%)	0.64 (0.44 to 0.95) p=0.0235
Placebo, n=174	62/173[a] (35.8%)	39 (22.5%)	9 (5.2%)	14 (8.1%)	
CAPACITY1,[2] study endpoint: 72 weeks					
Pirfenidone 2403mg/day, n=171	54/170[b] (31.8%)	31 (18.2%)	10 (5.9%)	13 (7.6%)	0.84 (0.58 to 1.22) p=0.355
Placebo, n=173	60/172[b] (34.9%)	41 (23.8%)	9 (5.2%)	10 (5.8%)	
Pooled[c] CAPACITY trials[2] study endpoint: 72 weeks					
Pirfenidone 2403mg/day, n=345	99/342 (28.9%)	59 (17.3%)	19 (5.6%)	21 (6.1%)	0.74 (0.57 to 0.96) p=0.025
Placebo, n=347	122/345 (35.4%)	80 (23.2%)	18 (5.2%)	24 (7.0%)	
SP3,[4] study endpoint: 52 weeks					
Pirfenidone 1800 mg/day, n=108					0.45 (0.11 to 0.79) p=0.028
Placebo, n=104					

[a] 2 participants in the pirfenidone group and 1 in the placebo group were excluded due to no post-baseline FVC or DLco. [b] 1 patient in each group excluded due to no post-baseline FVC or DLco. [c] Pooled analysis: participants censored pirfenidone group 243/342 (71.1%), placebo group 223/345 (64.6%)

Table 8: Outcomes from meta-analysis of PFS reported in the MS for CAPACITY trials[2] and SP3[4]

	[redacted]	[redacted]	[redacted]	[redacted]
Random effects	0.70	0.56 to 0.88	0.002	0%
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 9: Outcomes from meta-analysis of PFS reported in the MS for CAPACITY trials[2]

	Standardised mean difference	95% Confidence interval	P value for test of overall effect	I^2
Random effects	0.74	0.56 to 0.97	0.03	3%
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

3.3.3 Summary of results for overall survival

An exploratory survival analysis was conducted using data from the two CAPACITY trials.[2] Data are presented in Table 10 and Table 11 below. In the narrative the MS refers to different time points for the calculation of survival data (up to 72 weeks or beyond 72 weeks) and it is unclear to the ERG at what period the data presented in the MS are calculated (this is also not stated in the figure legends in the MS). Fewer than 10% of participants in either treatment group of the two CAPACITY trials[2] died (Table 10). In CAPACITY 2,[2] although there was a 39% relative reduction in the risk of death in the pirfenidone 2403mg/day group compared to placebo the difference in overall survival was not statistically significant (p=0.191). Risk of death in

CAPACITY1,[2] was very similar in each study group (9.4% vs. 9.8% deaths in the pirfenidone 2403mg/day vs. placebo groups respectively). As expected from the results of the individual trials the pooled analysis also showed no statistically significant difference between the groups (p=0.315).

The MS presents Kaplan-Meier estimates of overall survival for the individual trials and for the pooled analysis of the CAPACITY trials. Censored data are not marked on any of these figures and the methods of censoring are not provided, although they are described as 'appropriate' (MS p73).

The MS points out that after week 72 the majority of participants remaining at risk in both studies were lost to follow-up. For this reason, the HR at 72 weeks for IPF-related mortality was used in the economic model (reported in the MS text as HR 0.53 [0.288 – 1.028; p = 0.0606). IPF-related mortality was not defined. The MS states that the rates of IPF-related mortality were estimated as 7.2% and 3.9% in the pirfenidone and placebo arms, respectively but the ERG believes these data have been presented the wrong way round that that IPF-related mortality was estimated as 3.9% in the pirfenidone arm and 7.2% in the placebo arm. The MS points out that the HR for IPF-related mortality differs to that reported in the trial publication for the CAPACITY trials[2] (HR 0.62) because it is based on deaths occurring up to week 72, whereas the publication[2] includes mortality in an unspecified time period after the 72 week endpoint.

The MS presents additional survival data considering only treatment emergent deaths (Table 11, treatment-deaths defined in table footnote). The time span for these reported deaths is unclear in the MS. When all treatment-emergent deaths are considered the analyses find no statistically significant difference between the pirfenidone 2403mg/day groups and the placebo groups in either the individual CAPACITY trials[2] or the pooled analysis. Additional analyses also reported IPF-related treatment-emergent deaths. No details are provided regarding how IPF-related treatment-emergent deaths were distinguished from non-IPF-related treatment-emergent deaths. In the analyses of the individual CAPACITY trials although there are fewer IPF-related treatment-emergent deaths in the pirfenidone 2403mg/day groups than in the placebo groups the difference between the groups is not statistically significant (CAPACITY2,[2] p=0.129, CAPACITY1[2] p=0.121). However when the data are combined in the pooled analysis the risk of death is statistically significantly lower in the pirfenidone 2403mg/day group in comparison to the placebo group (HR 0.48 (95% CI 0.24 to 0.95) p=0.030).

Table 10: Overall Survival (data from MS Fig B5.5.5) *(numbers in italics estimated by the ERG)*

Study	Patient deaths, n (%)	Hazard ratio (95% CI) p-value
CAPACITY 2,[2]		
Pirfenidone 2403mg/day, n=174	11 (6.3%)	0.61 (0.28 to 1.29)[a] p=0.191[b]
Placebo, n=174	17 (9.8%)	
CAPACITY 1,[2]		
Pirfenidone 2403mg/day, n=171	*16* (9.4%)	0.95 (0.48 to 1.87)[a] p=0.872[b]
Placebo, n=173	*17* (9.8%)	
Pooled[c] CAPACITY trials,[2]		
Pirfenidone 2403mg/day, n=345	*27* (7.8%)	0.77 (0.47 to 1.28)[a] p=0.315[b]
Placebo, n=347	*34* (9.8%)	

[a] HR from Cox proportional hazard model. [b] Log-rank test. [c] the incidence of IPF-related deaths was lower in the pirfenidone 2403mg/day group than in the placebo group (12 versus 25 deaths) HR 0.62 (95% CI, 0.35 to 1.13).
Period of follow-up for these data is unclear in the MS

Table 11: Overall Survival: treatment-emergent deaths in CAPACITY trials[2]

	All treatment-emergent deaths[a]			IPF-related treatment-emergent deaths		
Study	Patient death, n (%)	Patients censored, n(%)	HR (95% CI) p-value	Patient death, n (%)	Patients censored, n(%)	HR (95% CI) p-value
CAPACITY 2[2]						
Pirfenidone 2403mg/day, n=174	10 (5.7)	164 (94.3%)	0.71 (0.31 to 1.60) p=0.404	5 (2.9%)	169 (97.1%)	0.45 (0.16 to 1.30) p=0.129
Placebo, n=174	14 (8.0%)	160 (92%)		11 (6.3%)	163 (93.7%)	
CAPACITY 1[2]						
Pirfenidone 2403mg/day, n=171	9 (5.3%)	162 (94.7%)	0.59 (0.26 to 1.36) p=0.212	7 (4.1%)	164 (95.9%)	0.50 (0.20 to 1.23) p=0.121
Placebo, n=173	15 (8.7%)	158 (91.3%)		14 (8.1%)	159 (91.9%)	
Pooled[c] CAPACITY trials[2]						
Pirfenidone 2403mg/day, n=345	19 (5.5%)	326 (94.5%)	0.65 (0.36 to 1.16) p=0.141	12 (3.5%)	333 (96.5%)	0.48 (0.24 to 0.95) p=0.030
Placebo, n=347	29 (8.4%)	318 (91.6%)		25 (7.2%)	322 (92.8%)	

[a] Treatment emergent deaths defined as deaths that occurred after the first dose and within 28 days of the last dose of study treatment.
Period of follow-up for these data is unclear in the MS
ERG note that the HR for IPF-related treatment-emergent deaths is different from the HR for IPF-related deaths reported in the narrative of the MS (HR 0.62 (95% CI 0.35, 1.13) page 105). This may be related to the different definitions used, the actual number of deaths reported are the same but a different time period for the calculation of the HR may have been used. This is not discussed in the MS.

The MS also reports a series of meta-analyses of mortality, which include data from the and SP3[4] trials. Data from these can be seen in

Appendix 1 of this report.

3.3.4 Summary of results for 6MWT

The 6MWT was reported in the CAPACITY trials[2] and the worst SpO_2 observed during the 6MWT was reported in the CAPACITY trials[2] the SP3[4] trial and the SP2[3] trial. These data can be seen in Table 12 and Table 14 below.

In the CAPACITY 2[2] trial no statistically significant difference in the change from baseline 6MWT was observed between the pirfenidone and placebo groups. The CAPACITY 1[2] trial showed a statistically significant change from baseline favouring the pirfenidone group ($p<0.001$), and this was also reflected in the pooled analysis presented in the CAPACITY trial publication.[2] In the pooled analysis the difference in distance walked between those treated with pirfenidone and those treated with placebo was 24 metres. It is unclear whether this represents a clinically significant effect (see Section 3.4 for more discussion).

Meta-analyses of the mean change in 6MWT are presented in the MS and the summary results are shown in Table 13 below. Results suggest that pirfenidone led to a favourable change in 6MWT when compared to placebo. No statistical heterogeneity was identified. The ERG note that the mean change values entered in the meta-analysis come from the least squares mean generated from a rank ANCOVA rather than the mean change in 6MWT presented in the trial reports, and the ERG are unable to check these data.

The worst SpO_2 observed during the 6MWT was not found to be statistically significantly different between the pirfenidone treated participants and the placebo treated participants (Table 14).

Table 12: 6MWT distance taken from MS, trial reports and EPAR, *or estimated by ERG*

Study	Baseline 6MWT mean (SD) metres	6MWT 72 weeks mean metres	Mean change, mean (SD) metres	Difference in mean change (metres) pirfenidone vs. placebo[a], p-value
CAPACITY 2,[2]				
Pirfenidone 2403mg/day, n=174	411.1 (91.8)	*350.7*	-60.4 (120.61)	*16.4*, p=0.171
Placebo, n=174	410.0 (90.9)	*333.2*	-76.8 (135.4)	
CAPACITY 1,[2]				
Pirfenidone 2403mg/day, n=171	378.0 (82.2)	*332.9*	-45.1 (139.81)	*31.8*, p<0.001
Placebo, n=173	399.1 (89.7)	*322.2*	-76.9 (127.5)	
Pooled CAPACITY trials,[2]				
Pirfenidone 2403mg/day, n=345	394.6	*341.8*	-52.8	*24*, p<0.001
Placebo, n=347	404.6	*327.8*	-76.8	

[a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

Table 13: Outcomes from meta-analysis of 6MWT CAPACITY trials[2]

	Standardised mean difference	95% Confidence interval	P value for test of overall effect	I^2
Random effects	0.21	0.06, 0.36	0.006	0%
Fixed effect	0.21	0.06, 0.36	0.006	0%

Table 14: Worst SpO_2 during the 6MWT *(numbers in italics estimated by the ERG)*

	Mean change from baseline	Difference in mean change pirfenidone vs. placebo[a], p-value
CAPACITY 2[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=174	-1.5%	*0.8%*, p=0.087
Placebo, n=174	-2.3%	
CAPACITY 1[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=171	-1.9 %	*-0.6%*, p=0.893
Placebo, n=173	-1.3%	
SP3[4], study endpoint: 52 weeks		
Pirfenidone 1800 mg/day, n=108	-1.70% (SD 0.35)	*-0.17%*, p=0.739
Placebo, n=104	-1.53% (SD 0.35)	
SP2[3], study endpoint: 36 weeks		
Pirfenidone 1800mg/day, n=72	0.47% (SD 3.88)	*1.41%*, p=0.0722
Placebo, n=35	-0.94% (SD 3.35)	

[a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

3.3.5 Summary results for dyspnoea

Ratings of a participant's dyspnoea by the University of California, San Diego Shortness of Breath Questionnaire (UCSD SOBQ) was not found to be statistically significantly different between those treated with pirfenidone and those treated with placebo in the CAPACITY trials[2] (Table 15).

3.3.6 Summary results for exacerbations

Exacerbation rates were reported for the CAPACITY trials[2] as part of the composite outcome of 'time to worsening of IPF' (also included IPF-related death, lung transplantation, or respiratory hospitalisation). Data on exacerbation rates alone were not however provided in the MS for these two trials.[2]

Data on exacerbation rates from the SP3[4] trial and SP2[3] trial can be seen in Table 16. No statistically significant differences were observed between treatment groups in the SP3[4] trial. A statistically significant difference in exacerbation rate was observed between groups in the SP2[3] trial, with placebo participants demonstrating more exacerbations than the pirfenidone group.

3.3.7 Summary results respiratory hospitalisations

The mean number of days alive without a respiratory hospitalisation was reported from the CAPACITY trials[2] in the MS and was shown not to be different between treatment groups (Table 17). The ERG have been unable to check these data as they were not in the published trial reports.

The MS also undertook a post hoc analysis of the number of hospitalisations for respiratory and non-respiratory reasons which were found to be similar between the pirfenidone and placebo groups. The post hoc analysis also found that the mean number of days spent in hospital was around 50% less for the pirfenidone group (see MS Table B5.5.13). As this is a post hoc analysis the ERG have not replicated these data as it is unclear how reliable these data are. In addition, the ERG is unable to check these data as they were not in the published reports.

Table 15: Dyspnoea scores (UCSD SOBQ) *(numbers in italics estimated by the ERG)*

	Mean change from baseline	Difference in mean change pirfenidone vs. placebo[a], p-value
CAPACITY 2;[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=174	12.1	*3.1*, p=0.509
Placebo, n=174	15.2	
CAPACITY 1;[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=171	11.9	*2.0*, p=0.604
Placebo, n=173	13.9	

Scores range from 0 to 120, with higher numbers indicating more breathlessness. [a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

Table 16: Acute exacerbation rates *(numbers in italics estimated by the ERG)*

	Acute exacerbation rate	Difference in rates pirfenidone vs. placebo[a], p-value
SP3[4], study endpoint: 52 weeks[1]		
Pirfenidone 1800 mg/day, n=108	5.6%	*-0.8*, States not significant.
Placebo, n=104	4.8%	
SP2[3], study endpoint: 36 weeks		
Pirfenidone 1800mg/day, n=72	0	*14*, p=0.0031
Placebo, n=35	14%	

[1]SP3 data continue beyond the primary point of analysis (up to 28 days beyond 52 weeks). [a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

Table 17: Number of days alive without a respiratory hospitalisation *(numbers in italics estimated by the ERG)*

	Mean number of days	Difference in mean days pirfenidone vs. placebo[a], p-value
CAPACITY 2;[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=174	481.0	*14.3*, p=0.263
Placebo, n=174	466.7	
CAPACITY 1;[2] study endpoint: 72 weeks		
Pirfenidone 2403mg/day, n=171	477.1	*2.9*, p=0.735
Placebo, n=173	474.2	
Pooled CAPACITY trials,[2]		
Pirfenidone 2403mg/day, n=171	479.1	*8.7*, p=0.292
Placebo, n=173	470.4	

[a]A positive difference = pirfenidone is favoured, a negative difference = placebo is favoured

3.3.8 Summary of Health related quality of life

The MS reports data on health related quality of life in the two CAPACITY trials[2] as measured by the SGRQ and the WHO QOL [redacted] The ERG has been unable to check the data presented in the MS as these do not appear in the published trial reports.

A meta-analysis of the trial results for the SGRQ was undertaken in the MS [redacted]

Table 18: St George's Hospital Respiratory Questionnaire (SGRQ) *(numbers in italics estimated by the ERG)*

Table 19: WHO Quality of life questionnaire (WHOQOL) *(numbers in italics estimated by the ERG)*

Table 20: Outcomes from meta-analysis of SGRQ for CAPACITY trials[2]

3.3.9 Sub-group analyses results

The MS reports data from a number of subgroup analyses for the % predicted FVC from the CAPACITY trials.[2]

A subgroup analysis by baseline characteristics is presented (MS figure B5.5.2, p97-8). It is unclear whether these analyses were stated apriori or were sufficiently powered given the sample size, and as such the ERG have not reproduced these data as caution is required in the interpretation of the results. The results generally favoured pirfenidone however the confidence intervals were often wide, crossing the line of no effect indicating no treatment difference. The narrative of the MS (page 96) reports that pirfenidone was favoured when participants with predicted FVC <80% were analysed. However, no data for this subgroup were presented. Data from two subgroups (% predicted FVC <70% and % predicted FVC 70-80%) may have been combined to calculate this subgroup but these data are not presented as one group in the MS (MS figure B5.5.2, pp97-98 shows the individual subgroups, and shows that in the pooled analysis the individual subgroups confidence intervals cross the line of no effect).

The MS also presents a categorical assessment of the change in % predicted FVC, using five categories ranging from severe decline of ≥ 20% or death or lung transplant, to moderate improvement of ≥ 10% FVC in MS Table B5.5.3, page 100. In the text associated with these data the MS reports combined categories for moderate or severe decline (≥ 10% decline % predicted FVC) in line with the published trial publication.[2] These data suggest that more participants showed a moderate or severe decline in % predicted FVC at week 72 in the placebo groups (pooled analyses 30.5%) than the pirfenidone treated groups (pooled analyses 21.5%). The text of the MS also reports data from the combination of the two categories of improvement (mild or moderate improvement; >0% improvement % predicted FVC). These data were not presented in the published trial report but show more participants with an improvement in % predicted FVC at week 72 in the pirfenidone group compared to the placebo group (pooled analyses 24.9% versus 17.8% respectively). The ERG has been unable to cross check these data.

The ERG note that in a third category (mild decline of ≤ 10% but ≥ 0%) presented in MS Table B5.5.3 there are some data which do not appear to favour pirfenidone (pooled analyses pirfenidone 53.6% versus placebo 51.6%). Although these differences are small and may be an artefact of the how data were applied to these categories, the MS does not discuss this.

Statistical analyses of these data across all five categories suggested a statistically significant difference between groups, however, caution is required as it is unclear whether these analyses were planned apriori and were powered to detect a difference.

The MS also report data from a subgroup analysis of the 6MWT, which has not been reproduced here as the MS explicitly states that it is post hoc (see MS Table B5.5.9, page 111).

3.3.10 Mixed Treatment Comparison results

No mixed treatment comparison was undertaken by the MS as no relevant data were identified on searches. The ERG undertook searches of Medline, Embase and MEIP and did not identify any studies of relevance for a mixed treatment comparison.

3.3.11 Summary of adverse events

The MS reports adverse events for the CAPACITY trials[2] combined and then adverse events from the SP3[4] trial separately.

[redacted] Few adverse event data are presented in the MS for SP2 although some data are reported in the published paper.[3] Limited data on adverse events are presented from the non-RCTs.

CAPACITY trials:[2] Treatment-emergent adverse events

The summary of adverse events concentrates on treatment-emergent adverse events (TEAEs) which were defined as AEs that occurred after the first dose and within 28 days after the last dose of study treatment. Common TEAEs are reported in MS Table B 5.9.2 (pp153-5). The MS does not state how a common event was defined. The MS reports summary data for each organ class, with a breakdown providing additional detail. The numbers provided in the detailed breakdown can sum to a value greater than that for the overall organ class and therefore, the ERG assumes that the participants are only counted once in the overall organ class data although they may have experienced more than one of the TEAEs contributing to that class. The ERG summarises the top-level data for each class of AE in Table 21 below.

Table 21: Treatment emergent adverse events

Common TEAEs	Number of patients, n (%)	
	Pirfenidone 2403mg/day (N-345)	Placebo (N= 347)
Patients with any common TEAE	336 (97.4%)	326 (93.9%)
Cardiac disorders	20 (5.8%)	18 (5.2%)
Ear and labyrinth disorders	10 (2.9%)	8 (2.3%)
Eye disorders	7 (2.0%)	11 (3.2%)
Gastrointestinal disorders	254 (73.6%)	173 (49.9%)
General disorders and administrative site conditions	147 (42.6%)	112 (32.3%)
Infections and infestations	232 (67.2%)	231 (66.6%)
Investigations	43 (12.5%)	20 (5.8%)
Metabolism and nutrition disorders	65 (18.8%)	22 (6.3%)
Musculoskeletal and connective tissue disorders	92 (26.7%)	84 (24.2%)
Nervous system disorders	107 (31.0%)	79 (22.8%)
Psychiatric disorders	64 (18.6%)	52 (15.0%)
Respiratory, thoracic and mediastinal disorders	196 (56.8%)	207 (59.7%)
Skin and subcutaneous tissue disorders	152 (44.1%)	62 (17.9%)
Vascular disorders	30 (8.7%)	19 (5.5%)

The MS provides a narrative summary of the common adverse reactions of photosensitivity, anorexia, and decreased appetite, the rationale for focussing on these events is not provided.

Photosensitivity reactions are well known adverse effects of pirfenidone treatment and fall into the class 'Skin and subcutaneous tissue disorders' in Table 21 above. In the CAPACITY trials[2] 12.2% of patients in the pirfenidone 2403mg/day group reported a photosensitivity reaction in comparison to 1.7% of the placebo group (Table 22). All photosensitivity reactions were considered to be related to study treatment with the majority of patients experiencing a single event. Most events resolved and about half of the patients were treated with a corticosteroid (more often topical than systemic). The dose of study treatment was modified because of a photosensitivity reaction in more patients treated with pirfenidone 2403 mg/day than with placebo. Nearly half of photosensitivity reactions in the pirfenidone 2403mg/day group were reported between weeks 0 and 18 of study treatment and incidence was greatest in April, May, June and July.

Table 22: Photosensitivity reactions reported in the CAPACITY studies

Adverse event	pirfenidone 2403 mg/day	placebo group
Photosensitivity reaction	42/345 (12.2%)	6/347 (1.7%)
Grade 3 photosensitivity (erythema with desquamation)	3 patients	1 patient
Grade 4 photosensitivity	0	0
Hospitalisation for photosensitivity reaction	0	0

CAPACITY trials:[2] Hepatic function

The MS reports on hepatic function in a section of text and a table (MS Table B5.9.1, 0152) with some liver related test outcomes also reported in the Table of Common TEAEs (MS Table B5.9.2, pp153-5, under 'Investigations') and in the later table (MS Table B5.9.3, pp153-5, Treatment-emergent SAEs, under 'Investigations'). It is difficult to reconcile the data in the adverse events tables with those in the hepatic function table which the ERG has brought together in Table 23.

Table 23: Occurrence of liver-related common TEAEs, treatment-emergent SAEs and Specific liver outcomes in the CAPACITY trials

	Number of patients, n (%)	
	Pirfenidone 2403mg/day (N-345)	Placebo (N= 347)
Common TEAEs: Investigations - GGT increased	17 (4.9%)	8 (2.3%)
Treatment-emergent SAEs: Investigations - Liver function test abnormal	2 (0.6%)	0
Laboratory test result or outcome		
AST or ALT		
> 3 x ULN	14 (4.1)	2 (0.6)
> 5 x ULN	3 (0.9)	2 (0.6)
≥ 10 x ULN	1 (0.3)	1 (0.3)
≥ 20 x ULN	0	1 (0.3)
Total bilirubin >2 x ULN	0	0
Dose modification due to ALT/AST elevation	12 (3.5)	1 (0.3)
Discontinuation due to ALT/AST elevation	2 (0.6)	1 (0.3)
Liver-related SAEs	3 (0.9)	1 (0.3)
Liver-related deaths	0	0
Hy's Law	0	0

ALT - alanine aminotransferase; AST - aspartate aminotransferase; ULN - upper limit of normal.
Liver-related SAEs - ALT/AST increased; hepatitis; LFT abnormal
Hy's Law - ALT or AST > 3 x ULN and total serum bilirubin > 2 x ULN based on the same blood sample;
GGT: gamma-glutamyl transpeptidase

More participants in the pirfenidone 2403 mg/d group compared with the placebo group had a >3 x upper limit of normal (ULN) elevation in serum transaminases but the occurrence of greater elevations (≥10 x ULN) was <1% and comparable between the two treatment groups. The SPC for pirfenidone indicates that dose adjustment should be made for patients with confirmed elevations in ALT, AST or bilirubin during treatment. Dose modification for ALT or AST elevation was required in 3.5% of patients receiving pirfenidone 2403 mg/d, with 0.6% of patients discontinuing treatment.

The MS reports that three patients receiving pirfenidone 2403 mg/day and one placebo patient were identified with liver-related SAEs. These data differ from those reported in the MS Table B5.9.3, p156, for the treatment-emergent SAE of liver function test abnormal (and reproduced in

Table 23 above). This may be because SAEs additional to those that were defined as treatment-emergent were included, or because additional liver-related SAEs are included. None of the patients had clinical sequelae, required hospitalisation or died.

CAPACITY trials:[2] Serious Adverse Events

The summary of serious adverse events reports on the treatment-emergent SAEs that occurred in more than two patients and at a greater incidence in either pirfenidone group (1197mg/ day or 2403 mg/day) than in the placebo group. These treatment-emergent SAEs are reported in MS Table B 5.9.3, p156. The MS does not state how a serious event was defined. The numbers provided in the detailed breakdown frequently sum to a value less than that for the overall organ class and therefore, the ERG assumes that not all of the TE SAEs contributing to that class have been listed. In addition, the way the MS presents the TE SAEs leads to some potentially unexpected results. For example, two SAEs (colitis and gastro-oesophageal reflux disease) are detailed under 'gastrointestinal disorders'. The ERG presumes these two appear because they occurred at a greater incidence in the pirfenidone group than the placebo group. However, it would appear that overall gastrointestinal SAEs were less common in the pirfenidone group than the placebo group [n=8 (2.3%) versus n=13 (3.7%) respectively]. Clinical advice to the ERG indicated that the opposite would be expected. The ERG therefore presumes that other events contributing to this overall value were ones which occurred at a greater incidence in the placebo group. The ERG summarises the top-level data for each class of SAE in Table 24 below.

Overall, the proportions of patients who experienced SAEs were comparable across all treatment groups and patient subsets (MS reports range 32.3% to 35.7% which does not appear to correspond to data presented below). The numbers of patients experiencing individual TE SAEs were small; however, no striking imbalances were noted (Table B5.9.3 EPAR[23]).

Table 24: Treatment emergent SAEs that occurred in > 2 patients and at a greater incidence in either pirfenidone group than in the placebo group in the randomised subset

Treatment-Emergent SAEs	Number of patients, n (%)	
System Organ Class Preferred Term	Pirfenidone 2403mg/day (N=345)	Placebo (N= 347)
Patients with any TE SAE	113 (32.8)	109 (31.4)
Cardiac disorders	21 (6.1)	17 (4.9)
Gastrointestinal disorders	8 (2.3)	13 (3.7)
General disorders and administrative site conditions	8 (2.3)	5 (1.4)
Infections and infestations	27 (7.8)	32 (9.2)
Injury, poisoning, and procedural complications	10 (2.9)	2 (0.6)
Investigations	2 (0.6)	3 (0.9)

Musculoskeletal and connective tissue disorders	7 (2.0)	4 (1.2)
Nervous system disorders	8 (2.3)	10 (2.9)
Psychiatric disorders	2 (0.6)	2 (0.6)
Renal and urinary disorders	8 (2.3)	5 (1.4)
Respiratory, thoracic and mediastinal disorders	40 (11.6)	46 (13.3)
Vascular disorders	6 (1.7)	3 (0.9)

SP3 and SP2 trials

The MS reports adverse events from the SP3 trial[4] that occurred with a frequency ≥ 5% during the study (52 weeks). The ERG has tabulated these data with the adverse event data reported in the trial publication for SP2[3] which encompassed those events observed with a frequency of ≥10% at six months (Table 25).

Table 25: Adverse events reported for SP3 and SP2

Adverse event	**SP3[4] AEs with frequency ≥ 5% during 52 week follow up**			**SP2[3] AEs with frequency ≥ 10% during 6 month follow up**		
	Number of patients, n (%)		**p-value**	**Number of patients, n (%)**		**p-value**
	Pirfenidone 1200mg/day (N= 109)	Placebo (N= 107)		Pirfenidone 1200mg/day (N= 73)	Placebo (N= 36)	
Any adverse event	109 (100.0)	106 (99.1)	0.50	72 (98.6)	32 (88.9)	0.0400
Photosensitivity	56 (51.4)	24 (22.4)	<0.01	32 (43.8)	0 (0.0)	0.000
Eczema asteatotic	0 (0.0)	0 (0.0)				
Anorexia	18 (16.5)	3 (2.8)	<0.01	23 (31.5)	2 (5.6)	0.0030
Abdominal (SP3) or stomach (SP2) discomfort	3 (2.8)	0 (0.0)	0.25	22 (30.1)	3 (8.3)	0.0143
Nausea				16 (21.9)	2 (5.6)	0.0314
Heartburn				12 (16.4)	1 (2.8)	0.0566
Dizziness	8 (7.3)	1 (0.9)	0.04			
Drowsiness				17 (23.3)	6 (16.7)	0.4672
Fatigue				16 (21.9)	1 (2.8)	0.0102
Nasopharyngitis	54 (49.5)	70 (65.4)	0.02			
Upper respiratory tract infection	1 (0.9)	9 (8.4)	<0.01	12 (16.4)	3 (8.3)	0.3767
Fever				6 (8.2)	4 (11.1)	0.7271
Elevation of GOT				4 (5.5)	6 (16.7)	0.0785
ɣ-GTP elevation	25 (22.9)	10 (9.3)	<0.01	20 (27.4)	3 (8.3)	0.0249
WBC decrease	4 (3.7)	0 (0.0)	0.12			
Urinary occult blood positive				6 (8.2)	4 (11.1)	0.7271
Elevation of CRP				15 (20.5)	10 (27.8)	0.4694

CRP - C-reactive protein; GOT - glutamic oxaloacetic transaminase, ɣ-GTP - ɣ-glytamyl-transpeptidase; WBC: white blood cell. P-value using Fisher's exact test.

Adverse events that were statistically significantly more common in the pirfenidone groups of both the SP3[4] and SP2[3] trials were photosensitivity, anorexia and ɣ-GTP elevation. In SP3[4] dizziness was also statistically significantly more common in the pirfenidone group. In SP2[3] the

occurrence of any adverse event, stomach discomfort, nausea, and fatigue were additional adverse events that were statistically significantly more common in the pirfenidone group. Two adverse events, nasopharyngitis and upper respiratory tract infection were significantly more common in the placebo group than in the pirfenidone group in SP3,[4] whereas in SP2[3] no event was reported that was more common in the placebo group than in the pirfenidone group.

For both SP3[4] and SP2[3] it is stated that most adverse events disappeared when the dose of pirfenidone was decreased or medication was temporarily withheld.

SP2 trial:[3] Hepatic function

The MS reports on one hepatic function adverse event from the SP2 study. The patient developed marked elevations of AST, and ALT as well as hyperbilirubinemia and elevated alkaline phosphatase after 56 days of therapy with pirfenidone 1800 mg/day. Pirfenidone was discontinued and by day 72 liver function abnormalities had markedly improved. This adverse event was considered likely to be related to pirfenidone and the patient met the criteria for Hy's law (ALT or AST > 3 x ULN and total serum bilirubin > 2 x ULN based on the same blood sample).

[redacted]

Table 26: Outcomes from random effects meta-analyses of adverse event outcomes for the CAPACITY trials;[2] and SP3[4]

	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

Discontinuations

The proportion of discontinuing participants is reported in several places in the MS: in text describing patient characteristics (MS pages 65 to 67), in flow diagrams (MS pages 76-78), in Table B5.5.1 (MS p88), and in text describing adverse events (MS pages 149 and 157). In general reports of discontinuation were consistent between the different sections. The exception was that the description provided in the text (MS pages 65 to 67) of those who discontinued treatment prematurely from the CAPACITY trials[2] did not specify that the values were for discontinuations due to adverse events and not for overall discontinuations.

The proportion of overall discontinuations from pirfenidone treatment arms (2403mg/day or 1800mg/day) ranged from 23.6% to 37% and those in the placebo arms from 22.4% to 29.8%. For discontinuations due to adverse events the proportion discontinuing from pirfenidone treatment arms (2403mg/day or 1800mg/day) ranged from 13.9% to 17.0% and those in the placebo arms from 5.6% to 10.4%. These discontinuations reported due to adverse events for the pirfenidone groups are in line with the opinion of the ERGs clinical expert who indicated that discontinuations due to adverse events would be expected in at least 10-15% of patients.

The MS does not report details of discontinuations due to non-response to pirfenidone.

3.4 Summary

Overall the systematic review has been undertaken reasonably and there is a low risk that bias will have been introduced by the methods of study selection.

The MS offers limited discussion of the meaning of the various results presented. There are some uncertainties which are addressed within the summary points addressed below.

Efficacy

- Differences between groups were small and varied between the included studies, these were generally favourable to pirfenidone but in some cases these were not statistically significant.
- There appears to be no clear reason why the CAPACITY 1[2] trial did not find statistically significant results when the CAPACITY 2[2] trial did.

- The clinical significance of differences seen between groups is unclear.
    - For FVC, the pooled CAPACITY trials showed a difference of 2.5% between the rate of decline in the pirfenidone group compared to the placebo group. [redacted]
    - The MS does not make a clear distinction between what would be considered a clinically significant effect on change in % predicted FVC in an individual patient and what constitutes a clinically significant effect within a cohort.
    - Clinical advice to the ERG suggests that the traditional threshold for a clinically significant effect of decline in % predicted FVC in an individual (≥10%) accounts for measurement variation in the FVC. A clinically significant effect in a cohort is likely to be lower as there is no net effect from measurement variation when the FVC is averaged out. Although requiring further validation, recent research supports this view, where a decline of 5% to 10% FVC was seen to have prognostic significance[13;14] and the minimally important clinical difference for an IPF cohort was computed as 2-6% of predicted normal values.[15] These studies were sponsored by Intermune.
    - The MS makes the point (MS page 163) that the proportion of patients with a 10% or more decrement [in % predicted FVC] is more directly clinically meaningful than the assessment of differences in treatment group means. In the categorical analysis of FVC change, based on the pooled CAPACITY[2] trials, the proportion of participants with moderate or severe decline (≥ 10%) is lower with pirfenidone (21.5%) than it is with placebo (30.5%). The ERG are unsure how robust this analysis is and whether this absolute difference in the proportion of participants in this category (9%) is statistically significant. In the overall discussion of the clinical effectiveness data the MS discuss the relative difference between these two groups as being 30% (absolute difference/placebo change: 9/30.5) which may be misleading to read without context of the analysis undertaken.
    - On the 6MWT there appeared to be a statistically significant differences in the pooled CAPACITY trial analysis, with a difference between groups of 24 metres, which is on the margin of being clinically significant according to the ERGs expert advice, however, there is the potential for inadvertent bias when the 6MWT is administered.

- PFS was shown to be improved with pirfenidone compared with placebo. This composite outcome includes time to confirmed ≥ 10% decline in % predicted FVC and it is unclear to the ERG whether this produces an issue with multiple counting because decline in % predicted FVC is also a separately measured outcome. The MS (page 163) discuss a possible weakness relating to the lack of adjustment for multiple statistical testing having the potential to over interpret the results but do not discuss the issue of 'double counting' within the outcomes reported.
- Little is known about any impact of pirfenidone on acute exacerbations, dyspnoea, respiratory hospitalisations and quality of life.

- Pirfenidone does appear to slow the rate of decline in lung function from IPF compared to placebo and the interpretation of the MS in this regard appears reasonable.

Safety

- Some adverse events were more frequently reported in those treated with pirfenidone but overall there appear to be no significant safety issues.
- The rates of discontinuations due to adverse events with pirfenidone compared with placebo were in line with the rate of discontinuations expected by the ERGs clinical experts. The MS does not report details of discontinuations due to non-response to treatment.

Other Issues

- The participants included in the RCTs may not be wholly representative of the population seen in secondary care in England and Wales. Although no one definition of 'mild-to-moderate' IPF based on the FVC exists, baseline FVC would suggest that the participants may be more 'mild' than 'moderate' cases of IPF (% predicted FVC/VCs in the region of 73-81% across all four RCTs) which is less severe than would be seen in most cases presenting in secondary care in the UK. However, caution is required as there are limitations with using the FVC to measure disease severity in the real world setting. Many people also present with comorbidities which are not always captured by FVC.
- The MS (page 163) state that there may be an issue with generalisability as patients had relatively few comorbidities in the trials. Clinical advice to the ERG suggests many patients will have comorbidities such as COPD /emphysema, hypertension, coronary

artery disease, and diabetes mellitus and these often contribute to mortality. Deaths from comorbidities would therefore not have been seen, especially during the early follow-up of the included studies.

4 ECONOMIC EVALUATION

4.1 Overview of manufacturer's economic evaluation

The manufacturer's submission to NICE includes:

i) a review of published economic evaluations of pirfenidone for the treatment of IPF,
ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of pirfenidone is compared with BSC for adult patients with mild to moderate IPF.

Manufacturer's review of published economic evaluations

A systematic search of the literature was conducted by the manufacturer to identify economic evaluations of pirfenidone for the treatment of IPF. See section 3.1.1 of this report for the ERG critique of the search strategy. The review did not identify any studies that compared pirfenidone to any other treatments.

CEA Methods

The cost effectiveness analysis (CEA) uses a micro-simulation model to estimate the cost-effectiveness of pirfenidone compared with BSC in adult patients with mild to moderate IPF. The model adopted a lifetime horizon in order to capture the differential effect of the interventions on patients' survival, with a 24-week cycle length. Although triple therapy with prednisolone, azathioprine and N-acetylcysteine was scoped as a relevant comparator, the MS does not present a CEA of pirfenidone compared to triple therapy (see Section 4.2.3).

The economic evaluation uses pooled data from the CAPACITY trials[2] and assumes that the trial patient population is representative of the UK population that is likely to receive pirfenidone. The MS also presents subgroup analysis for patients with a % predicted FVC of <80% in the CAPACITY trials, as these were reported to experience a greater treatment effect in the clinical trials (see however Section 3.3.9).

A Markov-type structure was used to model the progression of patients through six health states: *Alive and hospitalised*, *Alive not hospitalised*, *Dead due to IPF-related causes and*

hospitalised, *Dead due to IPF-related causes not hospitalised*, *Dead due to other causes and hospitalised*, and *Dead due to other causes not hospitalised*. Surrogate outcomes (FVC and 6MWD) are used to estimate the risk of IPF-related mortality, the risk of all-cause hospitalisation, and the SGRQ scores. These estimates were based upon regression analyses using the CAPACITY trial placebo arm data. Data from a placebo-controlled trial to study the effect of interferon gamma-1b on survival in patients with IPF (GIPF-007)[16] was also used in the mortality regression analysis.

Baseline levels and 24-week changes in FVC and 6MWD of patients from both treatment and BSC arms of the CAPACITY trials[2] were used to capture treatment effect. Adverse events are not explicitly included, but serious adverse events are expected to be captured by the number of hospitalisations in each arm.

In the model, SGRQ scores are estimated per patient in each cycle according to their FVC and 6MWD scores. These SGRQ scores are then mapped into EQ-5D utilities using the algorithm described by Starkie et al.[17] Although SGRQ scores were measured during the CAPACITY trials, these were not mapped into utilities. SGRQ scores were estimated by linear regression using FVC and 6MWD as independent variables.

The following cost categories were included by the manufacturer: treatment costs, oxygen and monitoring costs, hospitalisation costs and end of life costs. Treatment costs were derived from the number of pirfenidone pills prescribed per day and the discontinuation rates observed in the CAPACITY trials.[2] Resource use related to oxygen and monitoring costs was based on expert opinion, and their unit costs were obtained from the NHS Reference Costs 2009/10.[18] Hospitalisation costs were derived from the number and length of hospitalisations observed in the CAPACITY trials.[2] Average costs per bed day were derived from Personal Social Services Research Unit (PSSRU) 2010.[19] End of life costs were derived from estimates of a National Audit Office (NAO) report.[20] The cost of IPF-related deaths was derived from the total costs of cancer patients and the cost for other-causes of deaths from the total cost of heart and respiratory failure patients in their last year of life.

Univariate sensitivity analyses and the scenario analyses were performed by the manufacturer and the rationale for these are presented on pages 232 and 233 of the MS. A probabilistic sensitivity analysis (PSA) was undertaken with the parameters included/excluded described on

page 234. Model validation against mortality in the CAPACITY trials[2] is described on page 245 of the MS.

Besides univariate sensitivity analyses, scenario analyses are presented for structural assumptions, such as alternative estimations of IPF-related or other-causes related mortality risks, considering FVC as the only surrogate for estimation of IPF-related mortality, hospitalisation and SGRQ scores (excluding 6MWD), and excluding particular types of costs (oxygen and monitoring, hospitalisation, and end of life costs).

CEA Results

Results from the economic model are presented in section B6.7.3 (page 237 of the MS) as incremental cost per quality-adjusted life years (QALY) gained for pirfenidone compared with BSC. Total and incremental costs, life years gained (LYG) and QALYs are also reported. For the base case an incremental cost per QALY gained [redacted] is reported on page 237 of the MS (see Table 27 below). The ICER was particularly sensitive to the discount rate of costs and the number of pills of pirfenidone per day. An incremental cost [redacted] per QALY gained is presented on page 248 of the MS for the FVC<80% subgroup of patients.

Table 27: Base case cost effectiveness results

	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	Incremental cost per QALY (£)
PFD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BSC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

PFD: pirfenidone; BSC: best supportive care

The MS does not summarise the results of the PSA, however the cost-effectiveness acceptability curve (CEAC) presented on page 243 of the MS shows a [redacted] probability of pirfenidone being cost-effective, relative to BSC, at a willingness to pay threshold range of £20,000 to £30,000 per QALY gained.

4.2 Critical appraisal of the manufacturer's submitted economic evaluation

Manufacturer's review of published economic evaluations

A systematic search of the literature was conducted by the manufacturer to identify economic evaluations of pirfenidone for the treatment of IPF. The inclusion and exclusion criteria for the systematic review are listed in section 9.10.6 of the MS in Appendix 10. The inclusion criteria state that cost effectiveness or cost studies of adults with suspected or diagnosed IPF would be

included. Studies considered to be methodologically unsound, lacking adequate detail or devoid of any costing analysis were excluded.

Three studies were identified from screening 205 titles and abstracts. Of these all three studies were excluded. Two were costing studies and the other study was a cost effectiveness analysis for IPF but not for the use of pirfenidone treatment.

The ERG checked the search strategy used for the cost effectiveness searches and considered them reasonably comprehensive, fit for purpose and reproducible (Section 3.1.1 of this report). An additional search of NHSEED has been run by the ERG and has not found any cost effectiveness studies for pirfenidone.

Manufacturer's submitted economic evaluation

The ERG have considered the methods applied in the economic evaluation in the context of the critical appraisal questions listed in Table 28 below, drawn from common checklists for economic evaluation methods (e.g. Drummond et al[21]).

Table 28: Critical appraisal checklist of economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	Yes	On page 14 the MS states, 'a cost effectiveness model was developed to estimate the total costs and QALYs of pirfenidone compared to BSC in a hypothetical cohort of patients with mild to moderate IPF'.
Is there a clear description of alternatives?	Yes	BSC
Has the correct patient group / population of interest been clearly stated?	Yes	Base case population consists of patients in the CAPACITY trials. (Discussed in section 4.2.2)
Is the correct comparator used?	Yes	The comparator for the baseline was placebo / BSC. Triple therapy was not considered in the model. (Discussed in section 4.2.3)
Is the study type reasonable?	Yes	Cost utility analysis.
Is the perspective of the analysis clearly stated?	Yes	The NHS and Personal Social Services perspective (p15).
Is the perspective employed appropriate?	Yes	According to the NICE reference case[22]
Is effectiveness of the intervention established?	?	Patient-level data from the CAPACITY trials are used for the change in clinical symptoms such as FVC and 6MWD. It is unclear whether the intervention is clinically superior to BSC (ERG report Section 3.3 and 3.4). The model uses surrogate outcomes, rather than final outcomes. (Discussed in section 4.2.4)

Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	Yes	(Discussed in section 4.2.1)
Are the costs and consequences consistent with the perspective employed?	Yes	(Discussed in sections 4.2.6/ 4.2.7 for costs and 4.2.5 for outcomes)
Is differential timing considered?	Yes	Costs and health benefits discounted at 3.5% per year.
Is incremental analysis performed?	Yes	Given in MS Table B6.7.3 for the base case results and Table B6.7.4 for the sensitivity analyses.
Is sensitivity analysis undertaken and presented clearly?	Yes	One way sensitivity analysis is presented in MS Table B6.7.4. PSA is given for the base-case in section 6.7.8.

NICE reference case

The NICE reference case requirements have also been considered for critical appraisal of the submitted economic evaluation in Table 29.

Table 29: NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	
Comparator: Alternative therapies routinely used in the UK NHS	Yes?	BSC, does not include triple therapy. (Discussed in section 4.2.3)
Perspective on costs: NHS and PSS	Yes	
Perspective on outcomes: All health effects on individuals	Yes	(Discussed in section 4.2.5)
Type of economic evaluation: Cost effectiveness analysis	Yes	
Synthesis of evidence on outcomes: Based on a systematic review	Yes	(Discussed in section 4.2.4)
Measure of health benefits: QALYs	Yes	(Discussed in section 4.2.5)
Description of health states for QALY calculations: Use of a standardised and validated generic instrument	Yes	(Discussed in section 4.2.5)
Method of preference elicitation for health state values: Choice based method (e.g. TTO, SG, not rating scale)	Yes	(Discussed in section 4.2.5)
Source of preference data: Representative sample of the public	Yes	(Discussed in section 4.2.5)
Discount rate: 3.5% pa for costs and health effects	Yes	
? = uncertain; N/A=not applicable PSS = personal social services; TTO = time trade off; SG = standard gamble		

4.2.1 Modelling approach / Model Structure

A micro-simulation Markov model was chosen to estimate pirfenidone cost-effectiveness compared to BSC, as the manufacturer considered it to allow for more accurate survival predictions than a survival analysis based model as well as to accurately reflect the heterogeneity and complexity of IPF (MS page 176 and 177). Although individual patient

simulation is a valid modelling approach, the ERG highlights that these concerns could have been addressed with a cohort approach and the reason for not using this more common approach is not made clear.

The model was developed in Excel with most calculations being encoded in VBA. A schematic of the model is given on page 174 of the MS (Figure B6.2.1); however, this schematic does not present the health states listed on section 6.2.4 (p.179 of the MS), but it rather portrays the estimation of outcomes (mortality, hospitalisation costs and HRQoL) for each individual. The ERG presents a diagram with the health states and transitions in the MS model (Figure 1).

Figure 1: Schematic of the manufacturer's Markov model

Given the absence of previously published IPF-specific models, the manufacturer consulted an advisory board of experts in IPF and considered model structures for previous chronic respiratory disease economic evaluations (p.176 of the MS). The ERG considers that the model seems to incorporate the most relevant factors affecting life expectancy and quality of life in patients with a progressive chronic respiratory disease, i.e. mortality, hospitalisation, and HRQoL are estimated according to patients' pulmonary and physical function outcomes (FVC and 6MWD). The six health states modelled - *Alive and hospitalised*, *Alive not hospitalised*, *Dead due to IPF-related causes and hospitalised*, *Dead due to IPF-related causes not hospitalised*, *Dead due to other causes and hospitalised*, and *Dead due to other causes not hospitalised* - seem appropriate to capture cost differences between treatment arms.

At the start of the simulation, patients are assigned baseline characteristics for FVC and 6MWD and these attributes change in each subsequent 24 week cycle. The baseline characteristics of

647 of the 692 patients in the CAPACITY trials[2] were used for sampling as they had all the baseline characteristics required (age, gender, FVC and 6MWD, according to manufacturer's response to the ERG clarifications request). The reason for having 45 patients whose baseline characteristics were not used for sampling is not clearly stated by the manufacturer.

In each 24 week cycle, patients' FVC and 6MWD attributes are updated. The effectiveness of BSC was based on the simulated 24-week changes in FVC and 6MWD. The model independently sampled coupled 24-week changes in FVC and 6MWD for each BSC patient using the trial data for the placebo arms of the CAPACITY trials.[2] For each patient receiving pirfenidone, 24-week changes in FVC and 6MWD are sampled from the pirfenidone arms of the CAPACITY trials.[2] The model sampled from 932 and 929 recorded measures of 24-week coupled changes in FVC and 6MWD for placebo and pirfenidone patients in the CAPACITY trials respectively.

In each simulation, the model samples a cohort of 692 patients, simulates disease progression and mortality per individual, and calculates total costs, life years, and QALYs for the cohort. A thousand simulations are run and incremental cost-effectiveness results are calculated from the simulations' average outcomes. The MS did not explain the rationale for the bootstrapping method used and it is not clear whether it follows the conventional CEA methodology.[23]

The structural assumptions included in the list provided in section 6.3.8 (p.194 of the MS) are that FVC and 6MWD are significant predictors of IPF-related mortality, hospitalisation and SGRQ score; adjustment factors are applicable to the risk of IPF-related mortality which is independent of FVC and 6MWD in both arms, and these adjustment factors are assumed to remain constant throughout the model time horizon. Given the uncertainty surrounding these assumptions, the ERG notes these should be subject to sensitivity analysis (see Section 4.2.9 for assessment of uncertainty).

A lifetime horizon was chosen to capture relevant differences in survival benefits and costs between treatments (MS p.181). A 24-week cycle length was found appropriate by the manufacturer given that that was the data collection frequency in one of the trials from which data for the BSC arm was collected – GIPF-007 trial.[16] In the CAPACITY trials,[2] data collection was performed every 12 weeks, which may be an indicator that a shorter cycle length could be more appropriate.

4.2.2 Patient Group

For the economic evaluation, data from the CAPACITY trials[2] were used which included only patients with mild to moderate IPF reflecting the licenced indication for pirfenidone. The inclusion criteria for the studies were: diagnosis of IPF within 48 months of randomisation, age 40 – 80 years, % FVC ≥ 50%, DLco ≥ 35% of predicted value at screening, 6MWD ≥ 150m.

The base case for the economic evaluation comprises the total pooled patient population recruited into the two CAPACITY trials.[2] The MS (p172) states that using these trials *enables cost-effectiveness to be assessed consistent with the analysis and presentation of primary and key secondary endpoints in the study populations.* Furthermore, the MS justifies pooling the patients from the two trials because the trials were essentially identical in design and pooling the studies increased the sample size and hence provides more power in the statistical analyses. Although the MS also meta-analyses the outcomes from the CAPACITY trials, together with those from SP3, [redacted] these results are not used in the economic model.

The patient population in the model may not be fully reflective of the target population in current clinical practice or the scope of the appraisal, as these patients may have milder IPF than those typically seen.

The MS also presents subgroup cost-effectiveness analysis for patients with a predicted FVC of <80% in the CAPACITY trials, as these patients were reported to have experienced a greater treatment effect in the clinical trials. However no clinical effectiveness data were presented in the MS and the ERG has been unable to check this. It is unclear whether these data are statistically significant or whether the analysis was adequately powered.

4.2.3 Interventions and comparators

The comparator used for the economic analysis was placebo or BSC. The ERG notes, according to clinical advice, that BSC (i.e. not active treatment) is routinely used in UK NHS and that the comparator is appropriate for the economic model.

Triple therapy was not considered in the cost-effectiveness analysis. No RCTs have compared pirfenidone with triple therapy. In terms of RCT evidence for an indirect comparison, no RCTs were identified. The MS states that *all available evidence suggests triple therapy is more costly*

and less effective than placebo (p181). In MS section 5.7.3, the MS describes the PANTHER study, an ongoing study of triple therapy versus N-acetylcysteine alone versus placebo. The interim results from this study led to the termination of the triple therapy arm because there were higher mortality, hospitalisations and serious adverse events than in the placebo arm. However, it should be noted that mortality rates were also seen to be low in the placebo arm. A previous RCT of triple therapy compared to double therapy showed an advantage of triple therapy, which was possible most likely in those with mild IPF. The patients in the CAPACITY trials were mostly mild IPF when identified by FVC.

As noted in section 3.3.10, the ERG found no evidence for any other completed RCTs of triple therapy. On this basis, the manufacturer suggests an economic model would find triple therapy more costly and less effective than placebo. The ERG notes that triple therapy is in the NICE scope for this submission. The MS have not attempted to use any other type of evidence that may be available for triple therapy, nor discussed how they could have included triple therapy within the model even though the evidence may not have been robust. The MS does not discuss the limitations of not fulfilling the scope, nor discuss the relevance of the triple therapy comparator to current clinical practice. Clinical advice to the ERG suggested that triple therapy is used only in selected patients. The ERG note that the model itself is not structured to use data from any alternative comparator, such as triple therapy.

4.2.4 Clinical Effectiveness

The key clinical parameters in the model are change in FVC, and change in 6MWD, and treatment discontinuation rates. A list of all variables applied in the economic model is shown in Table B6.3.1 (p190) of MS. The key clinical events in the model are probability of IPF related mortality and non-IPF related mortality, and probability of hospitalisation.

Death due to non IPF-related mortality in each 24-week cycle was calculated based on all-cause mortality rates stratified by age and gender from the general UK population.[24] The ERG considers that the approach adopted for non IPF related mortality is consistent with standard modelling methodology.

The MS considers that FVC and 6MWD were significant predictors of IPF-related mortality, hospitalisation and SGRQ score. The MS justifies the use of FVC and 6MWD based upon the medical literature and expert opinion that there is a strong relationship between mortality and

FVC / 6MWD (see MS Section 6.3.4, p189). Furthermore they complete statistical analyses to confirm that FVC and 6MWD are significant predictors of the outcomes (see MS Appendix 14). The results of the regression analyses are shown in Appendix 14 (AIC data) but the ERG is unable to check these analyses as they do not have access to the patient-level data. The summary of the trial data for FVC and 6MWD are shown in sections 3.3.1 and 3.3.4 respectively of this report.

Several outcomes in the model are predicted as functions of FVC and 6MWD for each 24 week cycle, i.e. IPF-related mortality, all cause hospitalisation and SGRQ score. These functions are estimated through the use of logistic regression on the CAPACITY trial[2] data. The placebo arms of the CAPACITY trials were used for the regressions of IPF related mortality, all cause hospitalisation and SGRQ.

The probability of IPF-related mortality in a given 24-week cycle was estimated by applying a logistic regression which used FVC and 6MWD as independent variables (MS Appendix 13, AIC data). There were very few IPF related mortality events from the CAPACITY trials in the placebo arms. The manufacturer also included all available data from GIPF-007, an RCT comparing interferon-gamma with placebo in patients with mild to moderate IPF,[16] to provide more IPF-related mortality events, which the MS states makes for more reliable regressions from an epidemiological perspective (MS Appendix 22, CIC data). The ERG considers this to be an appropriate method to derive these parameters. The GIPF-007 trial is in a similar patient group, i.e. mild to moderate IPF, and the ERG considers that it is appropriate to pool these data with those from the CAPACITY trials.

The final logistic regressions chosen for the probability of IPF-related mortality and hospitalisation were:

Probability of IPF-related mortality = 1.177169 – 0.0555011*FVC – 0.0046714*6MWD

Probability of hospitalisation = 5.795926 – 1.115435*Log(FVC) – 0.5898508*Log(6MWD)

The MS states that the final regressions confirmed that FVC and 6MWD were significant predictors for the probability of IPF-related mortality and hospitalisation (see MS Appendix 14, AIC data). The results of the regression analyses are shown in Appendix 14 (AIC data) but the ERG is unable to check these analyses as they do not have access to the patient-level data.

Costs of IPF exacerbations and treatment-related adverse events are not explicitly considered. The costs of adverse events are considered by estimating the number of hospitalisations using FVC and 6MWD. The manufacturer assumes that acute exacerbations and severe adverse events would be captured by hospital stay. The MS states that if they were to additionally account for the adverse events considered in MS Section 5.9 (MS p148) and discussed in Section 6.2.2 (MS p174), this could potentially double count adverse events resulting in hospitalisations. As described in MS Section 6.2.2 (MS p174) adverse events such as nausea, rash, fatigue, diarrhoea, dyspepsia, and photosensitivity reactions would result in negligible costs, and any with severe cost consequences would be captured by hospital stay. The ERG considers that this is an appropriate approach to model adverse events and that the cost of these events would be included within the costs of hospitalisation, as stated by the manufacturer.

4.2.5 Patient outcomes

HRQoL estimates are applied to the patients in the model in each cycle, according to their FVC and 6MWD values. The HRQoL estimates are derived through a two stage process: first SGRQ values are estimated based upon their FVC and 6MWD outcomes and then these SGRQ values are mapped to EQ-5D values.

The manufacturer conducted a systematic review to identify HRQoL studies for IPF patients (MS p199). The search identified 20 relevant studies which are summarised in the MS. The MS does not give a critique of these studies, nor a rationale for not using them in the economic model. The MS describes the similarities of the SGRQ scores in HRQoL studies to that shown in the CAPACITY trials. The ERG notes that none of the HRQoL studies identified provided utility data or were for a large population of patients and suggests that none of these studies would have provided a better source of data than the CAPACITY trial used by the manufacturer.

MS Section 5.3.5 (MS p69) details the HRQoL data collected in the clinical trials (these are described in section 3.3.8 of this report). SGRQ and WHO-QOL instruments were used and measured every 12 weeks in the CAPACITY trials. The manufacturer estimated the SGRQ predicted by linear regression which used FVC and 6MWD as independent variables. The results of the regression analyses are shown in Appendix 13 (AIC data) but the ERG is unable to check these analyses as they do not have access to the patient-level data.

The final regression chosen to predict SGRQ score was:

$$SGRQ = 40.29755 + 1203.251*FVC^{-1} – 0.0436967*6MWD$$

The mean FVC, 6MWD and SGRQ scores for the CAPACITY trials[2] for the placebo arms are shown in Table 30. For comparison, the ERG has estimated mean SGRQ using the regression equation above.

[redacted]

Table 30: Comparison of derived SGRQ and mean trial SGRQ from the CAPACITY trials

	FVC		6MWD		SGRQ		Estimated SGRQ*	
	Baseline	72 weeks	Baseline	72 weeks	Baseline	72 weeks	Baseline	72 weeks
CAPACITY-2	76.2	63.9	410	333.2	[redacted]	[redacted]	[redacted]	[redacted]
CAPACITY-1	73.1	63.6	399.1	322.2	[redacted]	[redacted]	[redacted]	[redacted]

* Estimated by the ERG using regression equation for SGRQ and mean trial data for FVC and 6MWD

The manufacturer identified two studies which mapped either the SGRQ or WHO-QOL onto the EQ-5D (see MS Appendix 17, CIC data). Of the two studies identified the MS stated that the most appropriate study was a recent mapping study in COPD patients.[17] The algorithm from this study was used to map SGRQ scores from the CAPACITY trials[2] to EQ5D scores used in the model (see Section 6.4.6):

$$EQ\text{-}5D\ utility = 0.9617 - 0.0013*SGRQ - 0.0001*SGRQ^2 + 0.0231*Male$$

The mapping study[17] was performed in COPD patients from the TORCH (Towards a Revolution in COPD Health) trial. The trial collected EQ-5D and SGRQ values between weeks 24 to 3 years. Ordinary least squares (OLS), generalized linear models (GLMs) and two-part models were evaluated, and the best mapping equation as measured by goodness of fit was a simple OLS model. The ERG notes that the mapping from SGRQ to EQ-5D is not for IPF patients, but rather for COPD patients, and the MS does not discuss the generalisability of the mapping for patients from different patient groups.

For information, the ERG has derived mean utility scores using the regression equation above and the ERG derived SGRQ estimates for the trial data (Table 31).

Table 31: Derived utility values derived by the ERG using regression equations from mean trial data for FVC and 6MWD from the CAPACITY trials

	Estimated utility	
	Baseline	72 weeks
Male	0.79	0.73
Female	0.76	0.70

The MS states that the method taken to derive HRQoL was aligned with the reference case when EQ-5D is not available, i.e. measurement of HRQoL was from *the patient and valuation through the mapping study* (MS p218). The ERG considers the approach taken to estimate HRQoL to be appropriate and reasonable.

4.2.6 Resource use

The resources used in the model are shown in section B6.5.2 (p224) in the MS. The resource categories were: drug acquisition, oxygen and monitoring costs, hospitalisation costs and end of life costs.

The pirfenidone dosing schedule is stated in the MS 1.10 (p27). The recommended daily dose of pirfenidone for patients with IPF is three 267 mg capsules three times a day providing a total of 2,403 mg/day. This dosage is consistent with that used in the CAPACITY trials.[2] The MS recommends that the dose should be increased up to the recommended daily dose of nine capsules per day over a 14-day period as follows:
Days 1 to 7: one capsule, three times a day (801 mg/day)
Days 8 to 14: two capsules, three times a day (1,602 mg/day)
Day 15 onward: three capsules, three times a day (2,403 mg/day)

The model uses the average number of pills prescribed per day according to patient level data from the CAPACITY trial (MS Table B6.5.6, p228). The number of pills per day varied between 7.84 in cycle 1 and 7.89 in cycle 4+. MS Appendix 19 (CIC data) presents the results of the calculation of the number of pills per day in each cycle. It states that these estimates were based on the mean interval-specific values from the CAPACITY trials. The ERG is unable to check these analyses as they do not have access to the patient-level data.

The model estimates the drug acquisition cost, based upon the discontinuation rates observed in the trials. The MS states that there is no assumed clinical continuation rule. Patients that discontinue treatment do not incur the associated treatment cost. Cumulative rates of therapy

discontinuation are estimated for pirfenidone and placebo patients using patient level data from the CAPACITY trials (MS p183).[2] The MS also presents treatment discontinuation rates for the trials in Figures B5.3.1 – B5.3.2 (see also Section 3.3.11 of this report) and these differ from those used in the model and presented in MS Table B6.2.3 (MS p183). The ERG is unclear of the reasons for the differences between these data. The average cycle-specific cost of pirfenidone is shown in MS Table B6.2.3 (MS page 183, CIC data). This cost is estimated based on discontinuation rates from the trial, rather than numbers of model output related to individual patients. It appears from this table that the pirfenidone cost for cycles 4+ is constant, with a cumulative discontinuation of 29.45% of patients, i.e. pirfenidone cost of about 70% of maximum cost.

The MS does not make it clear how the cost estimation relates to the discontinuation rate. The ERG has examined the model and believes after the fourth cycle it is assumed that 70% of alive patients are on treatment. The ERG suggests that a better approach would have been to link the treatment costs with the patient-level data in the model.

The manufacturer conducted a systematic search to identify resource use for patients with IPF. No studies provided any relevant information regarding resource use and costing for IPF in the UK. The MS states that experts working in IPF were approached to seek advice on the resources required for monitoring and oxygen. Resource use for the first and subsequent model cycles are presented in Table B6.5.2 (MS p 222). The frequency of resource use was generally similar in both the pirfenidone and BSC groups apart from the need to conduct more frequent liver function tests for pirfenidone.

Hospitalisations in the model were considered to account for serious adverse events associated with treatment, and the natural course of the disease. The number of hospitalisations per patient was based upon a post hoc analysis on data from the CAPACITY trials. The post hoc analysis found that the mean number of days spent in hospital was around 50% less for the pirfenidone group (see MS Table B5.5.13, p114 and B6.5.3,p223). The average length of stay in hospital varied between 8.48 and 16.27 days for those receiving pirfenidone and BSC respectively (Table B6.5.3 p 223). As stated in section 3.3.7 of this report, the ERG was unclear for the reasons for the differences between the two groups and these are not discussed in the MS.

4.2.7 Costs

The main costs in the model are treatment costs, oxygen and monitoring costs, hospitalisation costs and end of life costs. MS states that NHS reference costs (2009/10)[18] have been used wherever possible Where these are not available or relevant, estimates from the literature have been used. The ERG confirms that this approach is appropriate and consistent with NICE modelling guidelines.[22] The costs used in the model are shown in the MS Table B6.5.1 - B6.5.5 (pp 222-229).

Drug acquisition costs for pirfenidone are [redacted] based on three 267 mg capsules three times a day for a total of 2,403 mg/day. The drug costs have not been published in the British National Formulary or MIMS at the time of writing.

Oxygen and monitoring unit costs for IPF were taken from the NHS Reference Costs 2009-10,[18] a previous HTA report for psoriasis,[25] and Regional Drugs and Therapeutic Centre.[26] The cost of oxygen is based on a 28-day cost of £403 for a portable cylinder of oxygen and £163 for home cylinder oxygen 8 hours per day. Ambulatory oxygen was assumed to be used in 12.5% of mild-moderate IPF patients 4-hours per day. Long term and nocturnal oxygen (using a home cylinder) was assumed to be used in 15% of IPF patients 8 hours per day. These assumptions were based on expert opinion using a flow rate of 2 litres/ minute.

Total oxygen and monitoring costs were calculated for pirfenidone and BSC respectively as: £1987.65 and £1981.18 for the first 24-weeks; and £1373.99 and 1368.98 in subsequent 24-weeks.

Average costs per bed day (£157.94) were derived from Personal Social Services Research Unit (PSSRU) 2010.[19] These costs are general inpatient bed day costs, rather than specific to IPF. The ERG suggest that bed day costs should be from NHS reference costs (2009/10).[18] The Healthcare Resource Group (HRG) excess bed day cost for *Respiratory Failure without Intubation* varies between £183 to £217 according to the level of associated complications (code DZ27D / DZ27E).

The cost of hospitalisation was calculated as the average cost per bed day multiplied by the average number of bed days observed in the CAPACITY trials (Table B6.5.4, p223). The cost

per bed day used by the manufacturer is £157.94, and the average cost per hospitalisation for pirfenidone and BSC patients is £1338.83 and £2570.21 respectively.

The model also includes end of life costs. The MS justifies the inclusion of these costs on the basis that NHS costs greatly increase in the last year of life. These costs were for the cost of patients in their last year of life with different diseases.[20] The manufacturer assumes that the total cost of patients with cancer in their last year of life can be used as proxy for end of life costs of a non IPF related death, and that end of life costs for patients that die from IPF-related causes was equal to the end of life cost of patients with heart and respiratory failure. The manufacturer inflated these figures in the model to 2011 prices using the CPI index for health from the Office of National Statistics.[27] Total costs of the last year of life for patients with IPF related and non IPF related death are estimated as £21,086.44 and £15,992.04 respectively.

The ERG suggests that a more usual approach would be to include only those costs which are specific to IPF, rather than including non IPF end of life costs. This is stated in the NICE modelling methods (NICE section 5.5.6, p41):[22] '*Costs that are considered to be unrelated to the condition or technology of interest should be excluded*'.

The ERG is uncertain whether the total cost of patients with heart and respiratory failure in their last year provides an accurate proxy for end of life costs of IPF-related causes. However, the manufacturer has provided sensitivity analyses and scenario analyses for end of life costs (Table B6.7.4 and B6.7.6) which have shown that changes to these costs or exclusion from the model have minimal impact on the model results.

4.2.8 Consistency / Model validation

Internal consistency

The economic model was developed in Excel with most calculations coded in VBA. The electronic file is fully executable, the model is generally well presented and user friendly, and the VBA code seems well structured. As the manufacturer submitted a simulation model that is programmed to randomly select values for several parameters for every type of analysis, deterministic analyses are not performed and therefore the model cost-effectiveness results vary for each simulation. The ERG was able to reproduce similar outputs; however, cost-effectiveness results vary as well for each run of 1,000 simulations and no estimate of their variability is reported by the manufacturer. The ERG suggests that the manufacturer should have demonstrated that an appropriate number of iterations per run had been determined to

ensure that the model performance is sufficiently accurate. Also, the manufacturer should have reported a 95% confidence interval of the ICER in order to show how accurately this is being estimated.

The base case analysis is conducted by clicking the button 'Run' in the 'Results' spreadsheet, which also presents the analysis results (including graphs). Parameter values are input in the 'Parameter list' spreadsheet and can be used to run univariate sensitivity analysis. Scenario analyses can be performed by unticking parameter-related boxes in several spreadsheets, such as the one in 'Monitoring and oxygen costs' spreadsheet. Subgroup analysis can be conducted for patients with FVC<80% by choosing the option 'Baseline FVC< 80' for both the 'baseline population' and the 'sampling data population' buttons available in the 'Results' spreadsheet. Results of subgroup, univariate and scenario analyses are displayed in the 'Results' spreadsheet. The PSA is run and its results are presented numerically and graphically in the 'PSA' spreadsheet.

The manufacturer states that two independent health economists were involved in the design and build of the model, conducted its quality assurance, and that clinical experts were consulted for the plausibility of the results (MS p.245). The MS does not report any techniques used for model internal validation. The ERG has not performed a full detailed cell-by-cell examination of the spreadsheets or of the VBA code, but key calculations have been randomly checked such as those used for the estimation of IPF-related mortality, risk of hospitalisation, SGRQ scores, and EQ-5D utilities. Results vary in the expected direction when parameter values are changed. No input errors have been found. Results of the scenario analysis on the BSC adjustment factor could not be reproduced by unticking the respective box in cell M7 in 'Results' spreadsheet. However, the same results were obtained by changing cell L7 formula directly in 'Results' spreadsheet. The submitted model does not reproduce the cost-effectiveness acceptability curve (CEAC) in the 'PSA' spreadsheet. This has been corrected by the ERG.

External consistency

The techniques for external validation are reported on page 245 of the MS. As the manufacturer did not find previous models in the literature review, the model's survival estimates were compared with those from the CAPACITY trials. The results presented show that the model estimates exactly the same IPF-related mortality as that seen in the trials (using the respective

adjustment factors). However, the model seemed to be overestimating the number of non-IPF related deaths and all-cause deaths. The clinical advisory group considered this to be a consequence of the clinical trial setting where unfit patients might not be entered as opposed to clinical practice (p.187 of the MS).

MS Table B.6.7.1 (MS p235) shows additional model results compared to clinical data for SGRQ scores at baseline and 72-weeks, and change from baseline SGRQ scores. The model seems to be underestimating the changes from baseline SGRQ scores for both arms. The justification that FVC and 6MWD may not fully describe SGRQ scores is given and the manufacturer states that the relative difference and absolute values are fairly similar so model results are not expected to be affected largely. The manufacturer's conclusions seem valid given the data presented.

The ERG considers that the validity of the model structure in terms of adequately reflecting disease progression should have been checked. The ERG requested clarification regarding the simulated changes of FVC and 6MWD over time as there were substantial variations that did not seem to reflect the natural progressive decline of the disease. The manufacturer responded that these were a product of averaging changes in FVC and 6MWD for the 692 patients' cohort, but noted that this would not reflect the average change in FVC over time (see NICE evaluation report for full details). The ERG verified that sudden unrealistic variations of FVC and 6MWD are simulated for individual patients in some model iterations as well. For model validation purposes, the manufacturer could have reported the model output for the average change in FVC and 6MWD over time and compared it with the changes observed in the CAPACITY trials or include realistic constraints on clinical characteristics in the models.

4.2.9 Assessment of Uncertainty

One-way sensitivity analyses

Several one-way sensitivity analyses were undertaken by the manufacturer and their results are reported in Table B.6.7.4 (page 238 of the MS). The variables subjected to sensitivity analysis were discount rates, pirfenidone discontinuation rate, number of pills per day, total hospitalisation cost for both treatment arms, total monitoring and oxygen cost for both treatment arms, IPF- and non IPF-related end-of-life cost. The values and ranges used for sensitivity analysis are clearly stated and their rationale is provided on page 233 of the MS. The

manufacturer took a common pragmatic approach using a range of +/- 25% of the mean value for most parameters, such as pirfenidone discontinuation rate and costs (MS p.233 and 238).

The manufacturer stated that the coefficients of the regressions used to estimate IPF-related mortality, hospitalisation, and SGRQ scores, and to map utility values from SGRQ scores were not subject to sensitivity analysis as they had been extensively analysed in Appendices 13 and 14 of the MS (p.233 of the MS). However, given the uncertainty of these estimates, the ERG suggests these should be subject to sensitivity analyses in order to analyse the impact of their variation on the ICER (see section 4.3 for additional ERG analyses).

The manufacturer concluded that the results are sensitive to the discount rates for costs and outcomes as well as the number of pills per day of pirfenidone, as the ICER varied from [redacted] (0% discount rate for costs) to [redacted] (three pills per day). This sensitivity analysis run with a third of the recommended daily dose assumed the same treatment effect as with full dose. Thus, the impact of varying the number of pills of pirfenidone per day reflects the ICER sensitivity to pirfenidone's cost and shows that an ICER [redacted] is estimated when nine pills per day (the recommended daily dose over a 14-day titration period according to the SPC[28]) are considered instead of the average [redacted] pills per day prescribed in the CAPACITY trials.[2] Table 32 below shows the results reported in the MS for the most influential parameters (MS p.238), apart from discount rates.

Table 32: Univariate analysis results of the most influential parameters

Parameter	Value	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)	Difference from base case (£/QALY)
Base case	-	[redacted]	[redacted]	[redacted]	[redacted]
PFD discontinuation rate	+ 25%	[redacted]	[redacted]	[redacted]	[redacted]
	- 25%	[redacted]	[redacted]	[redacted]	[redacted]
PFD Pills per day	3	[redacted]	[redacted]	[redacted]	[redacted]
	6	[redacted]	[redacted]	[redacted]	[redacted]
	9	[redacted]	[redacted]	[redacted]	[redacted]

PFD - pirfenidone

Scenario Analysis

The scenario analyses undertaken by the manufacturer are described in section 6.6.1 on page 232 of the MS and their results are reported on table B.6.7.6 (p.244 of the MS). The first scenario stated by the manufacturer on page 232 of the MS regards the extrapolation of survival

curves. Reference is made to Appendix 12 (AIC data) of the MS where the manufacturer presents the results obtained by fitting parametric survival models to clinical trials' survival data. However, the manufacturer does not report the impact of extrapolation with survival curves on the ICER.

Results for the following scenarios were presented (MS p.232 and 244) and are shown in Table 33 below:

- BSC IPF-related mortality adjustment factor ($\lambda_1 = 1$)
- Pirfenidone IPF-related mortality adjustment factor ($\lambda_2 = 1$)
- FVC as only predictor for IPF-related mortality, hospitalisation and SGRQ
- Exclude oxygen and monitoring costs
- Exclude hospitalisation costs
- Exclude end of life costs

The impact of alternative survival estimates on the ICER was explored by the manufacturer. The same base case method was used, i.e. a regression to predict IPF-related mortality, but by running the analyses excluding each of the adjustment factors individually (λ_1 or $\lambda_2 = 1$). The manufacturer also conducted the CEA using only FVC (the primary outcome of the CAPACITY trials) as a predictor of IPF-related mortality, hospitalisation and HRQoL scores, and explored the impact of several groups of costs on the CEA results.

Table 33: Results of the scenario analysis performed by the manufacturer

Parameter	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)	Difference from base case (£/QALY)
Base Case	[redacted]	[redacted]	[redacted]	[redacted]
BSC adjustment factor (λ_1=1)	[redacted]	[redacted]	[redacted]	[redacted]
PFD adjustment factor (λ_2=1)	[redacted]	[redacted]	[redacted]	[redacted]
Independent covariates in the logistic regressions for IPF-related mortality, hospitalisation and SGRQ[a]	[redacted]	[redacted]	[redacted]	[redacted]
Exclude oxygen and monitoring costs	[redacted]	[redacted]	[redacted]	[redacted]
Exclude hospitalisation costs	[redacted]	[redacted]	[redacted]	[redacted]
Exclude end of life costs	[redacted]	[redacted]	[redacted]	[redacted]

[a] This scenario is using FVC alone to predict outcomes (excluding 6MWD). PFD - pirfenidone

The ICER was most sensitive to not adjusting the BSC survival curve (λ_1=1).

[redacted]. The ICER varied from [redacted] (when BSC adjustment factor =1) to [redacted] (with the exclusion of oxygen and monitoring costs).

The manufacturer did not explore the impact of all the assumptions listed in section 6.3.8 of the MS (p.194), such as those related to treatment costs, HRQoL data (alternative scenarios using SGRQ scores from the CAPACITY trials could have been used), effectiveness estimates remaining the same throughout the model time horizon and being derived from surrogate outcomes using data from GIPF-007 interferon-gamma trial for only one of the treatment arms (BSC). Please see section 4.3 for further scenario analyses conducted by the ERG.

Probabilistic Sensitivity Analysis

The probabilistic sensitivity analysis can be run (in approximately 3 minutes) by clicking the respective button on the model's 'PSA' spreadsheet. This spreadsheet presents results for 1,000 iterations (incremental QALYs and ICER per iteration) as well as the average incremental cost, life years and QALYs gained, the ICER of pirfenidone compared to BSC, and the probability of it being cost-effective over a range of cost-effectiveness thresholds (£0 – £2,000,000 per QALY). PSA results are presented in Table B6.7.5 of the MS (p. 242), and a cost-effectiveness scatterplot and a CEAC are shown on page 243 of the MS. Credible intervals are not reported. The manufacturer did not discuss or conclude on any findings of the PSA. The probability of pirfenidone being cost-effective can be found in the CEAC which shows a probability of [redacted] at a willingness-to-pay range of £20,000 – 30,000 per QALY gained.

Not all variables were included in the PSA. In the PSA conducted by the manufacturer, only costs were added to the variables randomly drawn in the base case analysis (baseline characteristics and 24-week change of FVC and 6MWD). The variables included, their confidence intervals, and the distributions assigned are reported in Table B6.3.1 (page 190) of the MS. The manufacturer reports that all costs have been assigned a lognormal distribution and, where variability estimates were not available, standard errors were assumed to be 20% of the mean value. The ERG suggests that the CAPACITY trial data should have been used to determine the most adequate parameter distribution and respective variability.

Although confidence interval are reported in Table B6.3.1 (MS p.190) for most of the regression coefficients used to predict outcomes, these and the adjustment factors for IPF-related mortality

were excluded from the PSA, as the manufacturer states they had been extensively explored in scenario analysis and in the development of the equations (MS p. 234). Given the uncertainty surrounding these estimates, the ERG considers their inclusion in the PSA essential.

4.2.10 Comment on validity of results with reference to methodology used

The structure adopted for the economic model is reasonable, and consistent with the clinical pathway for IPF. The model has been coded in VBA which has made it less accessible and more difficult for the ERG to critique. The ERG has not found any errors in the coding of the model structure.

The parameters used for the model are generally appropriate. The population used in the model are those from the relevant trials and may not be representative of those treated in secondary care in the UK. The methods of analysis are generally appropriate and conform to NICE methodological guidelines.[22]

4.3 Additional work undertaken by the ERG

The ERG conducted the following additional analyses:

a) Variation of the regression coefficients used to estimate treatment effect
b) Variation of the hazard ratio for IPF related mortality
c) Variation of patients' HRQoL
d) Assuming the same length of stay in hospital for both treatment arms

a) Variation of the regression coefficients used to estimate treatment effect

The regression coefficients used in the model to estimate patients' survival, hospitalisation, and HRQoL were not subjected to sensitivity analysis in the MS. The ERG explored the impact that varying these regression coefficients has on the ICER. The model was run for the upper and lower confidence interval estimates for each of the regression coefficients. Table 34 shows the results of these analyses.

Table 34: Results of the variation of the regression coefficients

Parameter	Value	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)	Difference from base case (£/QALY)
Base case[a]	-	[redacted]	[redacted]	[redacted]	[redacted]
*Probability of IPF-related mortality coefficients [P = 1.177169 - 0.0555011*FVC - 0.0046714*6MWD]*					
95% CI LL	β_0=0.221182 β_1=-0.733194 β_2=-0.0062867	[redacted]	[redacted]	[redacted]	[redacted]

95% CI UL	β_0=2.13315 β_1=-0.0376828 β_2=-0.0030561[b]	[redacted]	[redacted]	[redacted]	[redacted]
*Probability of hospitalisation coefficients [P = 5.795926 - 1.115435*Log(FVC) - 0.5898508*Log(6MWD)]*					
95% CI LL	β_0= 3.90788 β_1=-1.595345 β_2=-0.757769	[redacted]	[redacted]	[redacted]	[redacted]
95% CI UL	β_0= 7.683972 β_1=-0.6355242 β_2=-0.4239247	[redacted]	[redacted]	[redacted]	[redacted]
*SGRQ score coefficients [SGRQ = 40.29755 + 1203.251*FVC^{-1} - 0.0436967*6MWD]*					
95% CI LL	β_0= 37.02277 β_1=1042.073 β_2=-0.0478999	[redacted]	[redacted]	[redacted]	[redacted]
95% CI UL	β_0= 43.57233 β_1=1364.43 β_2=-0.0394934	[redacted]	[redacted]	[redacted]	[redacted]
*EQ-5D mapping coefficients [EQ-5D utility = 0.9617 - 0.0013*SGRQ - 0.0001*SGRQ^2 + 0.0231*Male]*					
95% CI LL[c]	β_0= 0.5847 β_1=-0.0008 β_2=-0.00006 β_3=0.0140	[redacted]	[redacted]	[redacted]	[redacted]
95% CI UL[c]	β_0= 1.3387 β_1=-0.0018 β_2=-0.00014 β_3=0.0322	[redacted]	[redacted]	[redacted]	[redacted]

PFD – pirfenidone; UL – upper limit; LL – lower limit; [a] estimates used in the MS base case analysis are shown in each equation; [b] estimate derived by the ERG to correct the value erroneously reported in the MS (same estimate as the UL of β_1); [c] derived from assuming SE=mean*0.2

The ICER varied from [redacted] to [redacted]: a wider range than reported in the MS sensitivity analyses (see section 4.2.9 of this report). The ICER was particularly sensitive to variations in the coefficients used for the estimation of IPF-related mortality and for mapping the SGRQ scores into EQ-5D values.

The ERG also ran the model using the upper limit of all coefficients for the four regression equations, in which the ICER dropped to [redacted] gained.

b) Variation of the hazard ratio for IPF-related mortality

The ERG also analysed the impact of using different values for the IPF-related mortality HR at 72-weeks [mean= 0.53 (95% CI, 0.288 to 1.028)] by using the upper and lower limits of the HR 95% CI. The ERG obtained the IPF-related mortality hazard ratio using different estimates for the adjustment factor for pirfenidone treatment (λ_2), keeping the base case adjustment factor for

BSC constant ($\lambda_1 = 1.56$). The sensitivity of the ICER to the different survival estimates is shown in Table 35 below, and these reflect the uncertainty around the IPF-related mortality hazard ratio.

Table 35: Sensitivity analyses on the IPF mortality hazard ratio

Parameter	Value	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)	Difference from base case (£/QALY)
Base case	$\lambda_2 = 0.88$	[redacted]	[redacted]	[redacted]	[redacted]
Hazard ratio = 0.29	$\lambda_2 = 0.475$	[redacted]	[redacted]	[redacted]	[redacted]
Hazard ratio = 1.03	$\lambda_2 = 1.74$	[redacted]	[redacted]	[redacted]	[redacted]

c) Variation of patients' HRQoL

Two different scenarios were conducted to represent overestimation or underestimation of patients' HRQoL. Overestimation was obtained using the lower confidence limits of the coefficients used to estimate SGRQ score and the upper confidence limits of those used in the equation for mapping into EQ-5D. The opposite was done for the underestimation scenario. Results of these two scenarios are presented in Table 36, showing the high sensitivity of the ICER to the estimation of patients' HRQoL.

Table 36: Scenario results for variation in estimation of HRQoL

Parameter	Value	Incremental cost (£)	Incremental QALYs	ICER (£/QALY)	Difference from base case (£/QALY)
Base case[a]	-	[redacted]	[redacted]	[redacted]	[redacted]
Overestimation of patients' HRQoL					
SGRQ coefficients 95% CI LL	β_0= 37.02277 β_1=1042.073 β_2=-0.0478999	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D coefficients 95% CI UL	β_0= 1.3387 β_1=-0.0018 β_2=-0.00014 β_3=0.0322				
Underestimation of patients' HRQoL					
SGRQ coefficients 95% CI UL	β_0= 43.57233 β_1=1364.43 β_2=-0.0394934	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D coefficients 95% CI LL	β_0= 0.5847 β_1=-0.0008 β_2=-0.00006 β_3=0.0140				

[a] SGRQ = 40.29755 + 1203.251*FVC^{-1} - 0.0436967*6MWD
EQ-5D utility = 0.9617 - 0.0013*SGRQ - 0.0001*$SGRQ^2$ + 0.0231*Male

d) Assuming the same length of stay in hospital for both treatment arms

The model was run using the mean length of stay in hospital for the BSC arm (16.27 days) in both treatment arms. The ICER became [redacted], which does not substantially differ from the base case ICER.

4.4 Summary of uncertainties and issues from the critique of cost-effectiveness

- The comparator used for the economic analysis was placebo or BSC. Triple therapy was not considered although it was in the NICE scope for this submission. The MS have not attempted to use any other type of evidence that may be available for triple therapy, nor discussed how they could have included triple therapy within the model even though the evidence may not have been robust. The MS does not discuss the limitations of not fulfilling the scope, nor discuss the relevance of the triple therapy comparator to current clinical practice.
- The economic model has been coded as an individual patient simulation in VBA which has made it less accessible and more difficult to interpret and critique. It is uncertain whether the bootstrapping of the baseline characteristics and the individual patient simulation were adequately combined in order to accurately perform the CEA.

- The MS has not included all model parameters in either the univariate or probabilistic sensitivity analyses and so the full uncertainty around the model results has not been shown. In particular key parameters associated with overall survival, hospitalisations, and HRQoL have been omitted.
- There is some uncertainty around the discontinuation rates reported in the MS, where the rates reported in different sections differ. The reason for these differences is unclear. Furthermore, it is unclear what the long term discontinuation rates for patients on pirfenidone treatment would be.
- The average length of stay in hospital is significantly lower in the pirfenidone group than in BSC group. The reasons for the differences between the two groups are unclear and are not discussed in the MS.

5 END OF LIFE

The MS do not apply the NICE end of life criteria in the submission.

6 DISCUSSION

6.1 Summary of clinical effectiveness issues

The MS includes evidence on the efficacy of pirfenidone relative to placebo from four RCTs, and also includes data on adverse events from two non-RCTs. Overall the MS contains an unbiased estimate of the efficacy of pirfenidone at approximately 72 weeks follow-up. Results generally favoured pirfenidone, although in many cases differences observed were small and not statistically significant. Adverse events appear to be mostly mild to moderate. The MS provides a limited interpretation of the clinical evidence.

The MS does not provide an estimate of the clinical effectiveness of pirfenidone in relation to triple therapy (a scoped comparator intervention) in this population owing to limitations in the evidence base.

The population within the included RCTs may not be generalisable to those presenting to secondary care in England and Wales. Based on baseline FVC scores participants in the trials were likely to be of less severe IPF and few participants had the types of comorbidities expected to be seen in clinical practice.

A wide range of outcomes were reported across the included RCTs, and results between and within RCTs on these outcomes was seen to be varied. Overall it is unclear how meaningful changes on these surrogate outcomes are to those with IPF and the MS offers limited discussion of the issue of clinical significance.

[redacted] The methods chosen for the meta-analysis meant that the ERG could not check the data presented.

6.2 Summary of cost effectiveness issues

The MS includes evidence on the cost effectiveness of pirfenidone compared to BSC for IPF. The model structure and methods adopted for the economic evaluation are reasonable and are generally appropriate. The model structure and model parameter inputs are consistent with the clinical disease pathways and the available clinical trial evidence. The model results suggest that pirfenidone is not a cost effective option for a willingness-to-pay threshold of £20,000 per QALY.

The economic analysis did not include triple therapy, one of the comparators in the NICE scope for this submission. The MS does not discuss the limitations of not fulfilling the scope, nor discuss the relevance of the triple therapy comparator to current clinical practice.

It is uncertain whether the bootstrapping methodology and the individual patient simulation were adequately combined to accurately perform the CEA. The MS did not explain the rationale for the bootstrapping method used and it is not clear whether it follows the conventional CEA methodology.[23]

The MS has not included all model parameters in either the univariate or probabilistic sensitivity analyses and so the full uncertainty around the model results has not been shown.

There is some uncertainty around the discontinuation rates reported in the MS as these differ in different sections, and the reasons for the difference in hospital length of stay mean estimates between the two groups are unclear and are not discussed in the MS.

7 REFERENCES

1. National Institues of Health. News Release: Commonly used three-drug regimen for idiopathic pulmonary fibrosis found harmful. (Interim results of PANTHER study). 2011.

2. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011;**377**:1760-69.

3. Azuma A, Nukiwa T, Tsuboi E, et al. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005;**171**:1040-47.

4. Taniguchi H, Ebina M, Kondoh Y, et al. Pirfenidone in idiopathic pulmonary fibrosis. *European Respiratory Journal* 2010;**35**:821-29.

5. Raghu G, Johnson WC, Lockhart D, et al. Treatment of idiopathic pulmonary fibrosis with a new antifibrotic agent, pirfenidone: results of a prospective, open-label Phase II study. *Am J Respir Crit Care Med* 1999;**159**:1061-69.

6. Nagai S, Hamada K, Shigematsu M, et al. Open-label compassionate use one year-treatment with pirfenidone to patients with chronic pulmonary fibrosis. *Intern Med* 2002;**41**:1118-23.

7. Costabel U, Albera C, Cohen A, and et al. The long-term safety of pirfenidone in patients with idiopathic pulmonary fibrosis (IPF): Interim data from the RECAP extension study. ERS. 2011.

8. PIPF-002 Pirfenidone for the Treatment of Patients With Pulmonary Fibrosis/Idiopathic Pulmonary Fibrosis (PF/IPF). 2011.

9. Noble PW, Albera C, Bradford WZ, Costabel U, Kardatzke D, King TE *et al*. Effect of pirfenidone on lung function and progression-free survival (PFS) in patients with idiopathic pulmonary fibrosis (IPF): A meta-analysis of three phase 3 studies. *Am J Respir Crit Care Med* 2010;**181**:A1257.

10. King TE, Albera C, du Bois RM, Bradford WZ, Costabel U, Noble PW *et al*. The effect of treatment with pirfenidone on longitudinal change in lung volume in patients with idiopathic puomonary fibrosis (IPF): A meta-analysis of outcomes in four randomized controlled clinical trials. *Am J Respir Crit Care Med* 2011;**183**:A5302.

11. Demedts M, Behr J, Buhl R, et al. High-dose acetylcysteine in idiopathic pulmonary fibrosis. *NEJM* 2005;**353**:2229-42.

12. Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York Publishing Services Ltd.: CRD; 2009. Third edition

13. 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.

14. Wells AU, Richards TJ, Martinez FJ. Baseline values and short serial change. A "road map" for a poor early outcome in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;**184**:395-97.

15. du Bois RM, Weycker D, Albera C, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;**184**:1382-89.

16. King TE, Jr., Albera C, Bradford WZ, Costabel U, Hormel P, Lancaster L *et al*. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. *Lancet* 2009;**374**:222-8.

17. Starkie HJ, Briggs AH, Chambers MG, Jones P. Predicting EQ-5D values using the SGRQ. *Value.Health* 2011;**14**:354-60.

18. Department of Health. NHS reference costs 2009-2010. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122803 . 2010. Department of Health, England, UK. 7-6-2010.

19. Curtis, L. Unit Costs of Health and Social Care 2009. http://www.pssru.ac.uk/pdf/uc/uc2010/uc2010.pdf . 2010. Personal Social Services Research Unit.

20. Daly A, Hatziandreu E, and Archontakis F. *The potential cost savings of greater use of home- and hospicebased end of life care in England*. National Audit Office; 2008.

21. Drummond MF, O'Brien B, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes (3rd edition). *Oxford University Press* 2005.

22. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. *NICE, London* 2008.

23. Gray A, Clarke P, Wolstenholme J, Wordsworth S. Applied Methods of Cost-Effectiveness Analysis in Healthcare. Oxford University Press, 2011.

24. Office for National Statistics. Interim life tables, England and Wales. Period expectation of life based on years 2008-10. www.ons.gov.uk . 2011.

25. Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Vergel YB *et al*. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technol.Assess.* 2006;**10**:1-iv.

26. Regional Drug and Therapeutics Centre. Drug Update, Domiciliary Oxygen Therapy. www.nyrdtc.nhs.uk . 2004.

27. Office for National Statistics. CPI detailed annual average indices: 1996 to 2010. www.ons.gov.uk . 2011.

28. InterMune. Esbriet summary of product characteristics. www.ema.europa.eu . 2011.

8 APPENDICES

Appendix 1: Mortality data from included RCTs

The manufacturer reports the numbers of deaths for the different studies in a number of places in the MS (data on overall survival, in a series of meta-analyses, and in the section on adverse events). These data are presented using different follow-up periods and by different definitions. In some cases the definitions or time point for these data are not presented.

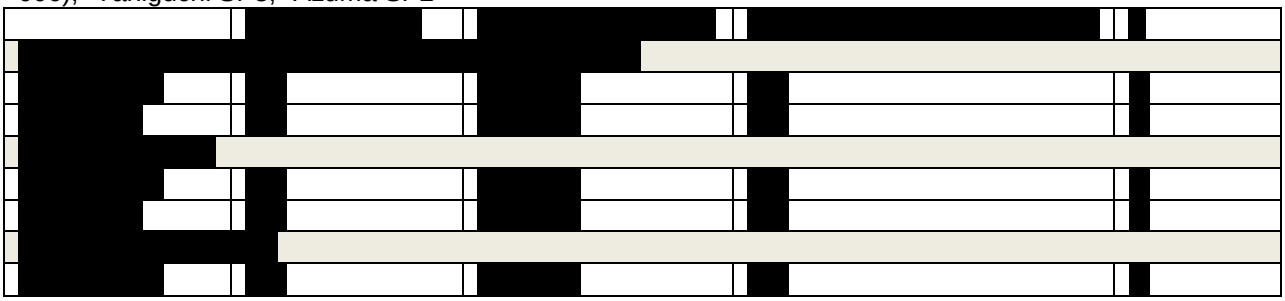
The ERG report reproduces the data on mortality used in the two CAPACITY trials[2] for the estimation of overall survival. Data for the complete set of mortality presented in the MS are reproduced here for completeness.

Table 1: All-cause mortality including deaths in the follow-up period; all-cause mortality; treatment emergent deaths

All-cause mortality including deaths in the follow-up period[a]		
	Number deaths pirfenidone	Number of deaths placebo
CAPACITY 2 (PIPF-004);[2]	[redacted]	[redacted]
CAPACITY 1 (PIPF-006);[2]	[redacted]	[redacted]
Taniguchi,[4] SP3	[redacted]	[redacted]
Azuma,[3] SP2	[redacted]	[redacted]
All-cause mortality		
CAPACITY 2 (PIPF-004);[2]	[redacted]	[redacted]
CAPACITY 1 (PIPF-006);[2]	[redacted]	[redacted]
Taniguchi,[4] SP3	[redacted]	[redacted]
Azuma,[3] SP2	[redacted]	[redacted]
Treatment emergent deaths		
CAPACITY 2 (PIPF-004);[2]	[redacted]	[redacted]
CAPACITY 1 (PIPF-006);[2]	[redacted]	[redacted]
Taniguchi,[4] SP3,	[redacted]	[redacted]
Azuma,[3] SP2	[redacted]	[redacted]

[a]ERG assume this relates to beyond study end-points but not stated in MS and definition is unclear. SP2 and SP3 n's appear to be different from those previously reported in the MS, assume include some or all initially randomised in the studies.

Table 2: Outcomes from meta-analysis of mortality for CAPACITY 2 (PIPF-004);[2] CAPACITY 1 (PIPF-006);[2] Taniguchi SP3;[4] Azuma SP2[3]



[redacted]

CAPACITY trials:[2] Deaths

The MS states that the lower incidence of all-cause mortality that was associated with pirfenidone is driven by a reduction in the incidence of IPF-related deaths. There was no individual cause of death that occurred in a clearly greater proportion of patients treated with pirfenidone 2403 mg/d relative to placebo.

Table 3: On treatment and IPF-related deaths from the pooled CAPACITY trials

	Pirfenidone 2403mg/day (N-345)	Placebo (N= 347)	
On-treatment deaths	19/345 (5.5%)	29/347 (8.4%)	35% lower p=0.141
IPF-related deaths	12/345 (3.5%)	25/347 (7.2%)	52% lower p=0.030
adjusted incidence of all-cause mortality (i.e. number of deaths per 100 patient-exposure years)	3.9	6.0	
adjusted incidence of IPF-related mortality	2.5	5.1	

p-value - log-rank test comparing pirfenidone 2403mg/day with placebo.