NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282) [ID837]

Appraisal Committee Meeting – 4 August 2016

The following documents are made available to the consultees and commentators:

1. **Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)**

2. **Consultee and commentator comments on the Appraisal Consultation Document** from:
   - Roche
   - British Thoracic Society
     The Royal College of Physicians endorsed the British Thoracic Society comments
   - Royal College of Nursing
   - Boehringer Ingelheim
     The Department of Health indicated that they had no comments

3. **Critique of the additional information submitted by Roche prepared by the Evidence Review Group**

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single Technology Appraisal

Pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282)

Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)

Definitions:

Consultees – Organisations that accept an invitation to participate in the appraisal including the companies, national professional organisations, national patient organisations, the Department of Health and the Welsh Government and relevant NHS organisations in England. Consultees can make a submission and participate in the consultation on the appraisal consultation document (ACD; if produced). All non-company consultees can nominate clinical experts and/or patient experts to verbally present their personal views to the Appraisal Committee. Company consultees can also nominate clinical experts. Representatives from NHS England and clinical commissioning groups invited to participate in the appraisal may also attend the Appraisal Committee as NHS commissioning experts. All consultees have the opportunity to consider an appeal against the final recommendations, or report any factual errors, within the final appraisal determination (FAD).

Clinical and patient experts and NHS commissioning experts – The Chair of the Appraisal Committee and the NICE project team select clinical experts and patient experts from nominations by consultees and commentators. They attend the Appraisal Committee meeting as individuals to answer questions to help clarify issues about the submitted evidence and to provide their views and experiences of the technology and/or condition. Before they attend the meeting, all experts must either submit a written statement (using a template) or indicate they agree with the submission made by their nominating organisation.

Commentators – Commentators can participate in the consultation on the ACD (if produced), but NICE does not ask them to make any submission for the appraisal. Non-company commentator organisations can nominate clinical experts and patient experts to verbally present their personal views to the Appraisal Committee. Commentator organisations representing relevant comparator technology companies can also nominate clinical experts. These organisations receive the FAD and have opportunity to report any factual errors. These organisations include comparator technology companies, Healthcare Improvement Scotland any relevant National Collaborating Centre (a group commissioned by NICE to develop clinical guidelines), other related research groups where appropriate (for example, the Medical Research Council and National Cancer Research Institute); other groups such as the NHS Confederation, the NHS Commercial Medicines Unit, the Scottish Medicines Consortium, the Medicines and Healthcare Products Regulatory Agency, the Department of Health, Social Services and Public Safety for Northern Ireland).

Public – Members of the public have the opportunity to comment on the ACD when it is posted on the Institute's web site 5 days after it is sent to consultees and commentators. These comments are usually presented to the appraisal committee in full, but NICE reserves the right to summarise and edit comments received during consultations, or not to publish them at all, where in the reasonable opinion of NICE, the comments are voluminous, publication would be unlawful or publication would be otherwise inappropriate.

Please note: Comments received in the course of consultations carried out by NICE are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the submissions that NICE has received, and are not endorsed by NICE, its officers or advisory committees.

Comments received from consultees

Consultee	Comment [sic]	Response
Roche	Thank you for the opportunity to comment on the Appraisal Consultation Document (ACD) for the appraisal of pirfenidone (Esbriet®) in the treatment of idiopathic pulmonary fibrosis (IPF). We are pleased to see that NICE recognised the continued role for pirfenidone in the management of patients with IPF, for those with a forced vital capacity (FVC) between 50-80%-predicted. We do, however, remain concerned that the Committee has failed to recognise the importance of treating patients with earlier stages of IPF (FVC ≥ 80%). Idiopathic pulmonary fibrosis is a chronic, progressive, and fatal lung disease that is characterised by irreversible loss of lung function. Early treatment to delay progression should, therefore, be an important goal for the management of the condition. Clinical opinion strongly advocates for earlier access to treatments, but this is not reflected in the prevailing guidance from NICE, which is ultimately to the determinant of patients. We are also concerned with several aspects of the approach taken to the assessment of the evidence by the ERG, primarily related to what appears to have been the starting point of their assessment: i.e. the existence of patient subgroups (defined by FVC %-predicted), despite no clinical evidence or opinion to support such an approach. We believe assessment of the clinical data should begin with consideration of the totality of the data. Lack of any evidence to suggest a difference in treatment effect based on FVC at baseline, combined with the clinical view that any threshold level is arbitrary and to the detriment of patients with preserved FVC %-predicted, is supportive of an approach to consider the entirety of the data, and is in line with NICE's Guide to Methods. In this respect, the review seems to have been artificially limited by the historical recommendations of TA282. This was developed prior to the availability of results from the ASCEND trial, which provides a substantial body of evidence for the effectiveness of pirfenidone across groups of patients defined by levels of FVC %-predicted. Given the recommendations of the ACD, and the prior agreement with the Department of Health regarding the proposed patient access scheme (PAS), analyses were re-run accounting for the PAS defined in TA282. These analyses were based on the ERG-preferred settings of the model, with the exception of a corrected CODA sample, and use of the most plausible parameterisation of the overall survival (OS) curve (Weibull: explanations provided below). Revised ICERs for the ITT population (assuming a lifetime treatment effect) were £18,167 with the stopping rule for pirfenidone, and £25,986 without the stopping rule. In order to achieve an ICER of approximately £30,000, the duration of treatment effect needed would be four and eight years (including and excluding the stopping rule, respectively). Both durations are clinically plausible, as they lie within the range of evidence from the pirfenidone clinical trial programme and comparable registry data.	Thank you for your comment. After considering the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee agreed that the recommendations in NICE's previous technology appraisal guidance on pirfenidone remained appropriate for people with idiopathic pulmonary fibrosis with an FVC between 50% and 80% predicted. Please see the Final Appraisal

Consultee	Comment [sic]	Response
	During the Appraisal Committee meeting, there was discussion that the majority of evidence from the ASCEND and CAPACITY trials is from patients with FVC < 90%-predicted. Analyses were, therefore, performed to demonstrate the cost-effectiveness of pirfenidone within this patient population. Within this subgroup, the ICERs ranged between an optimistic £16,676 (weighted OS curve, lifetime duration of treatment effect and stopping rule for pirfenidone) and a pessimistic £34,267 (Gompertz OS curve, eight-year duration of treatment effect and excluding the stopping rule). The following document provides further detail on our concerns with the assessment and interpretation of the evidence supporting the review of pirfenidone, and suggested approaches to allow the Committee to make a more considered recommendation.	Determination (FAD) document for further details. Please see the committee's considerations about subgroups are outlined in the FAD (see section 4.4 and 4.5).
Roche	**<u>Inappropriate focus on subgroup analyses, without evidence to support their existence</u>** The final scope for this appraisal stated that – subject to the evidence available – assessment would be made of patient subgroups by disease severity, defined by FVC and/or diffusing capacity for carbon monoxide. As discussed at the Appraisal Committee meeting, there are no accepted thresholds of FVC %-predicted used to define the disease severity of a patient with IPF in UK clinical practice, although there is a general acceptance that an FVC <50% predicted and DLco <35% predicted defines severe disease. Other staging systems include the GAP index, which includes age and gender as predictors of mortality, along with %-predicted DLco (Ley 2012). The composite physiology index adds forced expiratory volume in 1 second (FEV1) to FVC and DLco predicted values (Wells 2003). Consequently, including all characteristics of potential importance (as indicated by these alternate staging systems) is inherently associated with numerous theoretical and methodological caveats. Based on an assessment of the evidence available at the time of the first appraisal for pirfenidone in 2012-2013, NICE issued guidance which restricted use of the treatment to patients with an FVC < 80% predicted (NICE 2013a). As described in paragraph 4.16 of the current ACD, the Committee are still confident in the data supporting the initial review: "*In the original appraisal of pirfenidone, the committee had concluded that pirfenidone was clinically effective for moderate disease. In the current appraisal, the committee decided it had seen no evidence to alter that conclusion*". In line with the stated objective of this 2016 appraisal (paragraph 4.1), this review provides the opportunity for NICE to assess the totality of evidence to extend the Guidance for pirfenidone to patients facing a significant unmet need. Since publication of TA282 in April 2013, the ASCEND study has reported results; this study included approximately 100 patients with a FVC of ≥ 80% and < 90% predicted, a group of patients with no active treatment for management of their IPF. Importantly, when pooled with earlier data from the CAPACITY trials, ASCEND roughly doubled the number of patients available for assessment within this group, allowing for a robust analysis of any important differences in treatment effect between this group and the larger cohort. The cohort of patients with baseline FVC up to 90% predicted was discussed at	Comment noted. After considering the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee concluded that the subgroup of people with an FVC between 80% and 90% predicted was the relevant population for decision-making. Please see the committee's considerations about subgroups are outlined in the FAD (see

Consultee	Comment [sic]	Response
	the Appraisal Committee meeting, although this is not specifically captured in the ACD. Analyses of the clinical- and cost-effectiveness of pirfenidone in this group of patients – which account for approximately 90% of the pooled pirfenidone clinical trial population – are provided below. Using the most statistically appropriate methods, no difference in the efficacy of pirfenidone by patient subgroup has been identified [see following section], and we believe the data from the trial populations should be assessed in their entirety. This view is supported by information presented to NICE by a variety of stakeholders during the evidence submission stage, the Appraisal Committee meeting on 5th May, along with that heard during the appraisal of nintedanib (TA379), where it was described that there is no clear definition of what constitutes 'mild' or 'moderate' IPF, particularly when characterised by FVC %-predicted alone (NICE 2016a, NICE 2016b). The NICE Guide to Methods includes clear direction on the identification and assessment of subgroups. Whilst the Methods Guide may have been designed in the context of first / new appraisals for a treatment, the clear starting point is for an assessment across the licensed population: subgroups are to be identified within that population, based on criteria described in the Guide: *"5.10.1 …. The characteristics of patients in the subgroup should be clearly defined and should preferably be identified on the basis of an expectation of differential clinical or cost effectiveness because of known, biologically plausible mechanisms, social characteristics or other clearly justified factors.* *5.10.2 … There should be a clear justification and, if appropriate, biological plausibility for the definition of the patient subgroup and the expectation of a differential effect. Post hoc data 'dredging' in search of subgroup effects is to be avoided and will be viewed sceptically.* *5.10.6 The standard subgroup analyses performed in RCTs or systematic reviews seek to determine whether there are differences in relative treatment effects between subgroups (through the analysis of interactions between the effectiveness of the technology and patient characteristics). The possibility of differences emerging by chance, particularly when multiple subgroups are reported, is high and should be taken into account. Pre-specification of a particular subgroup in the study or review protocol, with a clear rationale for anticipating a difference in efficacy and a prediction of the direction of the effect, will increase the credibility of a subgroup analysis.* *5.10.7 In considering subgroup analyses, the Appraisal Committee will take specific note of the biological or clinical plausibility of a subgroup effect in addition to the strength of the evidence in favour of such an effect (for example, if it has a clear, pre-specified rationale and is consistent across studies)… "* We are concerned that the assessment of the evidence by the ERG – and seemingly supported by the Committee – has been conducted from a starting point of the existence of patient subgroups. This may be an artefact of the recommendations	section 4.4 and 4.5).

Consultee	Comment [sic]	Response
	made in TA282. We believe that, if due consideration was given to the rationale for those initial recommendations, the additional evidence provided through ASCEND, along with the prevailing clinical opinion, a fresh perspective should have been taken to this review. We attempted to provide this within our evidence submission through presentation of results related to the ITT population, but the current assessment has focussed on unproven differences in subgroup treatment effects as a means to justify subgroup analysis: this is in contrast to the approach set out in NICE's Guide to Methods. Indeed, the ACD states "*The committee understood that the results of the treatment-by-subgroup interaction tests were not significant in either of the subgroup analyses. However, it heard from the ERG that a non-significant interaction test does not conclusively mean that there is no difference in treatment effect between subgroups. The ERG explained that the interaction test may not have been powered to detect a difference between the subgroups. The committee concluded that it did not see robust evidence that pirfenidone is clinically effective in people with mild idiopathic pulmonary fibrosis (that is, an FVC above 80% predicted)*" (paragraph 4.16). Consistent with the NICE Guide to Methods, there is no biologically plausible mechanism whereby either the efficacy of pirfenidone and/or the capacity to benefit would differ between patients with differing FVC levels. IPF is a complex disease that is not yet fully understood, and research is hampered by the lack of a model which fully represents the human disease (White 2016). Notwithstanding this, the complex pathogenic cascade leading to the development of fibrosis in IPF involves numerous mediators and signalling pathways, and it is likely that an effective IPF therapy would need to target more than one mechanism/pathway. Both pirfenidone and nintedanib have pleiotropic effects, and it is likely their clinical efficacy derives from their broad-based mechanisms of action, which would be inconsistent with a subgroup effect based on lung function parameters alone. As described in the ATS/ERS guidelines: "*While the traditional approach to IPF staging has been useful, it is arbitrary and is not based on epidemiological or biological data. It remains unclear if these stages are truly relevant to the management of IPF. Critically, these traditional stages are not known to reflect distinct biological or clinical phenotypes and the true therapeutic and prognostic relevance of these stages remains undetermined*" (Raghu 2015) The intention of paragraph 5.10.6 of the Methods Guide (in the context of 5.10.2) is to seek robust evidence that a subgroup exists, implying a starting point of there being no difference in treatment outcomes across the populations included in the clinical trials. Based on this, the approach taken to the interpretation of the evidence is both surprising and confusing. The quotation taken from 4.16 of the ACD is correct and consistent with paragraph 5.10.6 of the Methods Guide, in that the subgroup interaction test may have been underpowered to detect a difference in treatment effect. However, this underpowering does not prove that there *is* a difference, and we would expect the ERG and NICE to challenge on this basis, were a company to try and argue such a case. The statement in the ACD also does not account for the consistency in treatment effect across FVC subgroup seen across trial outcomes: Table 9 of the Pre-Meeting Briefing document [represented below in Table 1] presents no treatment differences in OS or progression free survival (PFS) outcomes by FVC subgroup, and similar findings for other outcome measures, as presented in Figure 17 of the initial evidence submission (represented as Figure 1 below, NICE 2016a). This figure demonstrates no statistically significant difference in treatment effect across subgroups (defined by FVC %-predicted and GAP stage at baseline) for FVC, 6-minute walking distance (6MWD) and the University of California San Diego Shortness of Breath Questionnaire (UCSD SOBQ)).	

Consultee	Comment [sic]	Response
	Table 1: Representation of Table 9 of Pre-Meeting Briefing: Treatment effect of pirfenidone (overall survival and progression-free survival to week 52), according to baseline disease severity	

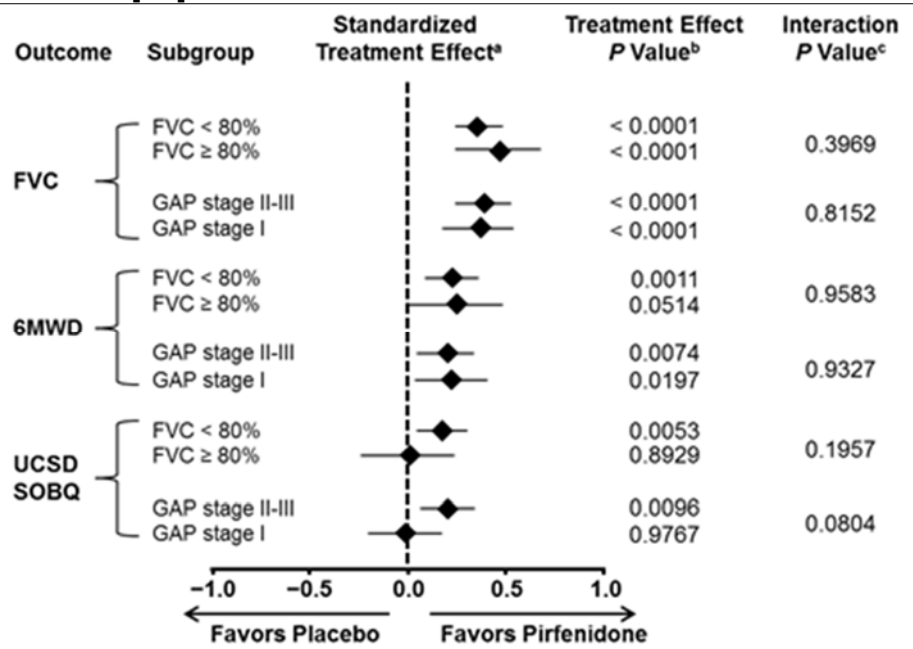
Trial	Percent predicted FVC	Hazard ratio[a] (95% CI)	Treatment effect[b], p value
Overall survival			
ASCEND	≤80%	0.63 (0.29 to 1.34)	0.22
	>80%	<0.01 (0.00 to not evaluable)	0.12
CAPACITY 1	≤80%	0.60 (0.17 to 2.04)	0.41
	>80%	0.77 (0.11 to 5.59)	0.80
CAPACITY 2	≤80%	0.25 (0.08 to 0.76)	0.01
	>80%	Not evaluable	Not evaluable
Progression-free survival			
ASCEND	≤80%	0.56 (0.41 to 0.76)	<0.05
	>80%	0.64 (0.30 to 1.40)	0.26
CAPACITY 1	≤80%	0.84 (0.53 to 1.32)	0.44
	>80%	0.63 (0.29 to 1.41)	0.26
CAPACITY 2	≤80%	0.60 (0.40 to 0.92)	0.02
	>80%	0.40 (0.18 to 0.89)	0.02

[a] hazard ratios below 1 indicate that patients having pirfenidone had a lower risk of the event than patients having placebo

[b] p value indicates significance of the difference between pirfenidone and placebo within the subgroup; p values for treatment-by-subgroup interaction test not reported

CI, confidence interval; FVC, forced vital capacity

Figure 1: Representation of Figure 17 of evidence submission: Treatment effect of pirfenidone by baseline disease severity from pooled data of ASCEND, CAPACITY 1 & 2

Consultee	Comment [sic]	Response
	 We are particularly confused with the statement "*The committee concluded that it did not see robust evidence that pirfenidone is clinically effective in people with mild idiopathic pulmonary fibrosis (that is, an FVC above 80% predicted*" (ACD, 4.16): when considering the change in FVC %-predicted from baseline, the treatment effect of pirfenidone in patients with FVC ≥ 80%-predicted is not only statistically significant, it is also of a similar size to the effect observed in patients with FVC 50-80%-predicted [Table 8 of Pre-Meeting and Table 7-10 of response to clarification question A29]. The *European Respiratory Journal* have also recently accepted for publication a manuscript presenting the efficacy of pirfenidone when used [redacted]. The results presented in the *ERJ* manuscript are in line with those presented in response to clarification question A29. Furthermore, as the Committee did not consider nintedanib to be a relevant comparator in this review (paragraph 4.6), all relevant comparisons to allow the Committee to make a decision on this cohort of patients were presented within the ITT analysis (i.e. pirfenidone vs. BSC).	

Consultee	Comment [sic]	Response
Roche	**Misinterpretation of pre-specified vs. *post-hoc* statistical analyses** In paragraph 4.16 of the ACD, it is stated: "*The committee was aware that the analysis of covariance (ANCOVA) in 2 **post-hoc** subgroups … suggested that pirfenidone was associated with a statistically significant benefit compared with placebo in both subgroups. However, the committee noted that, in the company's **pre-specified** analysis across 3 subgroups …, there was a nonsignificant tendency for better outcomes in the placebo group than the pirfenidone group among people with a baseline FVC above 80% predicted. The committee was aware of the company's opinion that the analysis with 3 subgroups was not as robust as the ANCOVA method, but the committee agreed that it was not appropriate to disregard a **pre-specified** analysis. In addition, during the committee meeting, the company could not fully explain the methods of the ANCOVA analysis. The committee understood that the results of the treatment-by-subgroup interaction tests were not significant in either of the subgroup analyses*" (emphasis added). Firstly, it is important to note that neither approach to the statistical assessment of subgroup differences identified a statistically significant interaction between treatment and patient subgroup: that is, there is no statistical evidence supporting a difference in treatment effect by baseline FVC %-predicted. This is not consistent with the definition of patient subgroups, set out in the Methods Guide (paragraphs 5.10.2 and 5.10.7). Secondly, it is incorrect to refer to the ANCOVA method as *post-hoc*: the analysis of standardised ranks was the **only** pre-specified analysis in the Integrated Summary of Efficacy (ISE) Statistical Analysis Plan (SAP) for the EMA for **assessment of the primary efficacy analysis and subgroups** [sent to NICE as part of our factual accuracy response to the ERG report]: • "*Data for the change from Baseline outcomes are not expected to be normally distributed. Therefore, data will be analyzed using a rank analysis of covariance (ANCOVA) model with a* *standardized rank change from Baseline as the outcome and standardized rank Baseline value as a covariate*" (Section 6.2, page 12). • "*The primary efficacy analysis of the pooled data is the rank ANCOVA model for the change from Baseline to Month 12 in %FVC between the 2403 mg/day pirfenidone and placebo groups*." (Section 6.4, page 13). • "*Interactions of subpopulations with treatment will be tested for in the analysis of change from Baseline to Month 12 in %FVC. Each factor will be tested individually. The factor and interaction with treatment will be added to the rank ANCOVA model for the Month 12 assessment.*"(Section 6.8, page 18) Assessment based on absolute change in FVC **was not** a specified analysis in the ISE SAP for the EMA. As FVC was not anticipated to be normally distributed, and as some patients were not expected to complete the full 12 month assessment (e.g. due to death), an analysis based on absolute change from baseline for completers of the 12 months follow-up would not be an appropriate method for either the primary endpoint or subgroup interaction tests. This is because such analyses are: (i) less robust to deviations from the normality assumption, compared to analysis of ranks (via a standardised treatment effect), and;	Comment noted. Please see the committee's considerations about the evidence outlined in the FAD (see section 4.8 to 4.13).

Consultee	Comment [sic]	Response
	(ii) likely to miss the effect on FVC, which were not captured in the event of mortality or study discontinuation. We believe the Committee's focus on assessment of absolute change in FVC may stem from Figure 16 on page 113 of our evidence submission. This related to the assessment of CAPACITY 1 and CAPACITY 2 studies (not ASCEND, nor the pooled analysis of all three trials). Figure 16 was submitted in error, as this was based on the (incorrect) analysis of absolute change in FVC. It was also inaccurate and potentially misleading for the title on page 114 of our submission to refer to the analysis of pooled subgroups which followed as post-hoc: as described above, the ISE SAP for EMA was planned to assess the primary efficacy outcomes (and subgroups) via analysis of standardised ranks. We apologise for these potentially misleading erroneous inclusions. It should also be noted [redacted]. However, regardless of the most appropriate method of statistical assessment, results of both analyses were non-significant, in line with there being no difference in the efficacy of pirfenidone by level of baseline FVC %-predicted.	
Roche	**Additional analysis in subgroup where most data are available (FVC< 90%-predicted)** We do not agree that considering subgroups of patients with IPF on the basis of FVC %-predicted alone is a valid approach, based on the argumentation described above. We do, however, acknowledge the discussion held at the Appraisal Committee meeting, regarding the majority of evidence from the ASCEND and CAPACITY being in patients with FVC < 90%-predicted. This is largely due to the enrolment criteria used in the ASCEND trial (FVC 50-90%-predicted, as opposed to FVC ≥ 50%-predicted in the CAPACITY trials). As described in the introduction to this response, analysis is presented for patients with IPF and baseline FVC < 90%-predicted. These assessments are based on data from all patients across the ASCEND and CAPACITY trials, excluding 101 patients within CAPACITY 1 and CAPACITY 2. Similar to the discussion above, there is no evidence of a treatment-interaction when the FVC < 90% subgroup is explored, and there is no clinical or biological rationale why treatment effect should differ in this subgroup. These analyses are solely presented on the basis that this group of patients represents the majority of those included in the pirfenidone clinical trials. Results of economic analysis using this subgroup are presented within the revised economic analyses in the sections below.	Comment noted. Please see the committee's considerations about subgroups are outlined in the FAD (see section 4.4 and 4.5).
Roche	**Appropriateness of a survival based model** Section 4.9 of the ACD states the Committee *"would have preferred to see a model that captured the progressive nature of idiopathic fibrosis, and linked clinical outcomes with each other and with time on treatment."* Whilst this might be the Committee's preference there are clear and valid reasons why such a model structure is not ideal for an assessment of the	Comment noted. After considering the comments received in response to the

Consultee	Comment [sic]	Response
	cost-effectiveness of pirfenidone. When determining a suitable model structure to use for a given decision problem, both the characteristics of the disease and the availability of data should be taken into account. As discussed previously, when considering disease characteristics, it is clear that modelling all characteristics of potential importance would ultimately surpass both available data and available clinical knowledge regarding the interactions between the various factors considered important to prognosis and quality of life (including FVC, DLCo, 6MWD and acute exacerbations). *(i) Choice of model was informed by availability of evidence for OS* The most reliable data available to evaluate the benefits of treatment with pirfenidone come in the form of a hard endpoint: OS. These data show that pirfenidone is currently the only treatment in IPF to demonstrate a significant OS benefit, with maximum follow-up available for over 8 years as part of the RECAP study. The demonstration of this significant OS benefit for pirfenidone is in fact what triggered this re-review of TA282. Overall survival data are also available from registries of patients with IPF, which include up to 14 years of follow-up for patients treated with best supportive care (BSC). Propensity scoring models were used to adjust for imbalances in patient characteristics. *(ii) Availability of OS data reduces need for model calibration, complexity and assumptions* The strength of the model structure submitted as part of this re-review lies in its simplicity and lack of necessary assumptions: as OS data are modelled independently, a more accurate prediction can be made in line with the clinical trial evidence, without the complication of attempting to link OS to imperfect predictors: • Within the existing partitioned survival structure, there is no requirement to back calculate OS which can result in inaccuracies in matching observed information from the clinical trials. It is noted that a major limitation of the nintedanib model submitted as part of TA379 was lack of accurate projection of outcomes over time – the assumed mortality rate for pirfenidone (approximately 6.1% based on digitisation of Figure 41 within the nintedanib company submission document) was almost double that actually observed at 1 year (3.6% based on the pooled pirfenidone Phase III trials) (Boehringer Ingelheim 2015). Projected OS for pirfenidone at 5 years was < 50%, compared to observed data indicating 70% survival. The model also failed to accurately project health state split, with consistent bias in terms of a lower proportion of patients modelled in severe health states vs. observed data, and over-prediction in mild health states in both arms. • Similar issues were observed with the microsimulation model submitted by InterMune as part of earlier review of pirfenidone (TA282), which required the use of calibration factors to provide a more accurate fit to actual observed data (NICE 2013a). • These issues underline the benefits of relying on more complete data, with fewer assumptions and less complexity, to produce projections. Follow up for change in FVC %-predicted over time is restricted to approximately 1 year in the nintedanib and pirfenidone clinical trial programmes.	ACD, the committee acknowledged the limitations in the data and concluded that the model could be used for its decision-making. Please see the committee's considerations in the FAD (see section 4.14).

<table>
<tr><th>Consultee</th><th>Comment [sic]</th><th>Response</th></tr>
<tr><td></td><td>
The model used to support the nintedanib appraisal defined health states for FVC in 10%-predicted categories. As patients could ‘skip’ FVC categories, this required assessment of patients who experience falls in FVC of ≥ 20%. We do not believe data are available to reliably inform such a model structure. Less than 2% of patients treated with pirfenidone experienced a fall in FVC %-predicted of ≥ 20% in the 12 month trial periods (Table 2). Similarly, there were limited data to inform such progressions in the nintedanib trial programme: less than 5% of patients experienced a fall in FVC of ≥ 20%-predicted. Further splitting these progressions patients into additional categories, such as those defined in the Boehringer Ingelheim company submission, would inevitably lead to major assumptions being made on a very small number of patients (Boehringer Ingelheim 2015).

Therefore, we do not believe that the impact of multiple FVC progressions (using a model structure similar to that presented in TA379, which seems to represent the ERG’s preferred approach) simply <u>cannot</u> be accurately projected with the available data. Whilst we agree it would be desirable to include such progressions within the model to allow a fuller assessment of the progressive nature of IPF, there is no reliable evidence to justify adding this additional complexity to the model.

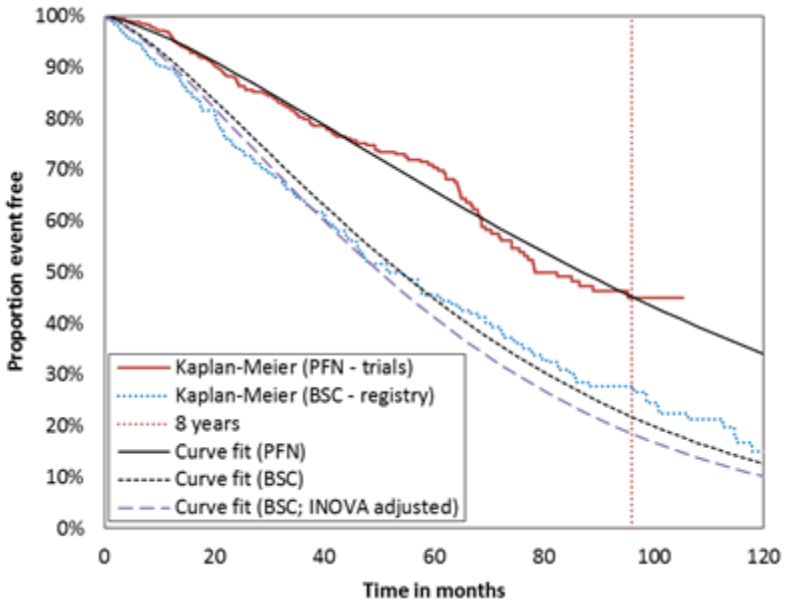
Table 2: Patient numbers for change in FVC %-predicted from baseline to 12 months
<table>
<tr><th>Change in FVC %-predicted</th><th>PBO</th><th>PFN</th><th>All</th></tr>
<tr><td>Increase of ≥ 30%</td><td>0 (0%)</td><td>0 (0%)</td><td>0 (0%)</td></tr>
<tr><td>Increase of 20-30%</td><td>1 (0.2%)</td><td>0 (0%)</td><td>1 (0.1%)</td></tr>
<tr><td>Increase of 10-20%</td><td>6 (1%)</td><td>9 (1.4%)</td><td>15 (1.2%)</td></tr>
<tr><td>Increase of 0-10%</td><td>120 (19.2%)</td><td>187 (30%)</td><td>307 (24.6%)</td></tr>
<tr><td>Decrease of 0-10%</td><td>339 (54.3%)</td><td>343 (55.1%)</td><td>682 (54.7%)</td></tr>
<tr><td>Decrease of 10-20%</td><td>133 (21.3%)</td><td>74 (11.9%)</td><td>207 (16.6%)</td></tr>
<tr><td>Decrease of 20-30%</td><td>21 (3.4%)</td><td>9 (1.4%)</td><td>30 (2.4%)</td></tr>
<tr><td>Decrease of ≥30%</td><td>4 (0.6%)</td><td>1 (0.2%)</td><td>5 (0.4%)</td></tr>
<tr><td colspan="4">Key: PBO, placebo; PFN, pirfenidone.</td></tr>
</table>

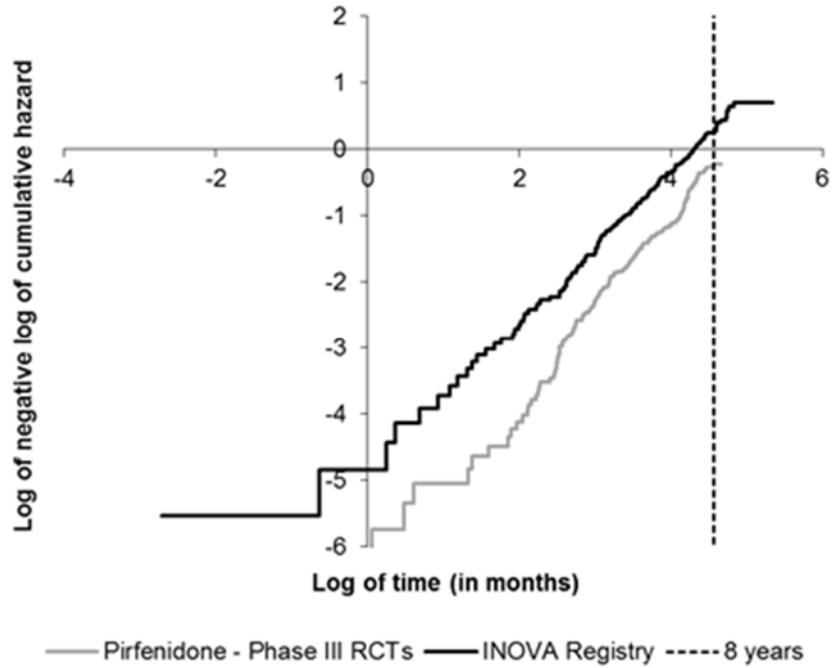
(iii) Additional complexity does not improve fit to cost or utility data
In addition to data limitations, the additional complexity of attempting to capture progression in terms of FVC-alone is not clinically warranted. The utility regression analysis presented with our submission indicated that inclusion of FVC %-predicted, FEV1, FEV1/FVC ratio, age and gender all harmed model fit: the final equation was informed purely by the St George’s Respiratory Questionnaire score (as reported on page 225 within Section 5.4 in the company submission, NICE 2016a).

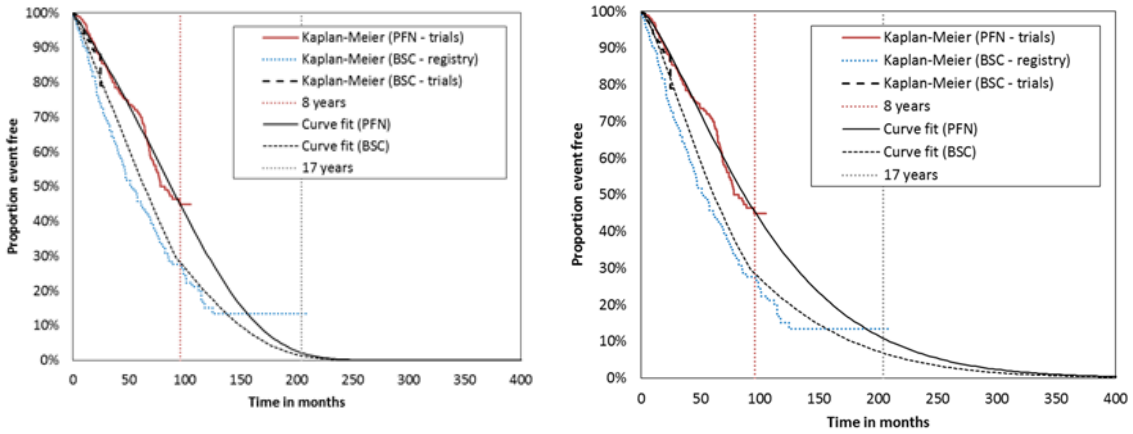
We note also that within the cost analysis conducted for the nintedanib appraisal, only hospitalisation costs showed change
</td><td></td></tr>
</table>

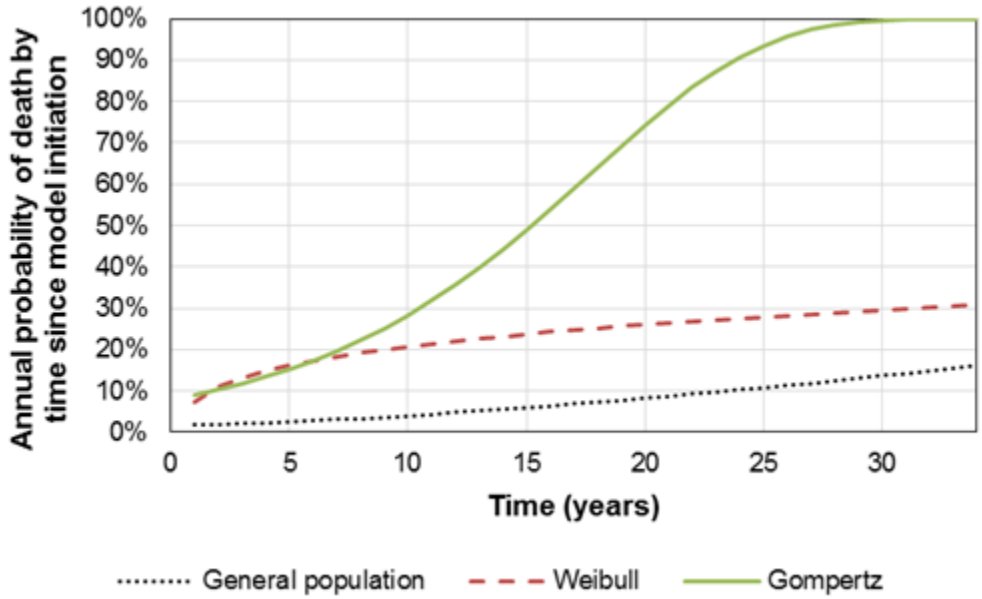
Consultee	Comment [sic]	Response
	with absolute FVC level, however, these costs are likely to be highly confounded with both acute exacerbation rates and adverse event rates making interpretation difficult. *(iv) Relationship between FVC and mortality is non-linear* It should be acknowledged that the relationship between FVC and mortality is non-linear; both rapidity and absolute levels of change may impact patient prognosis. FVC is also subject to measurement error, meaning that there is a large volume of noise in any individual patient outcomes. A patient with an FVC of 70%-predicted could well have a better prognosis than one with an FVC of 80%-predicted, if their disease trajectory is one of less rapid decline, or if there are comorbidities, such as emphysema (previously discussed in NICE committee meetings for both pirfenidone and nintedanib). Additionally, although the mechanism of action of pirfenidone has not been fully established, data suggest that pirfenidone exerts both anti-fibrotic and anti-inflammatory properties in a variety of *in vitro* systems and animal models of pulmonary fibrosis (bleomycin- and transplant-induced fibrosis). Fibrosis often occurs as a result of sustained injury to the epithelium, which causes the overproduction of cytokines and growth factors. This is a core feature shared by pathologic fibrosis among multiple organs tissues such as lung, kidney, and liver. As well as its impact on lung fibrosis, the anti-fibrotic properties of pirfenidone have also been demonstrated in other fibrotic diseases such as diabetic nephropathy and liver fibrosis. This suggests that pirfenidone could have multiple systemic anti-fibrotic effects, and could potentially contribute to an explanation for the non-linearity between FVC change and mortality benefit. *(v) Lack of data to support robust inclusion of acute exacerbations, without clear impact on model results* Modelling of acute exacerbations is even more complex, as FVC decline and acute exacerbations are difficult to distinguish, as discussed in recent NICE Appraisal Committee meetings for pirfenidone and nintedanib. Inclusion of acute exacerbations would also not add to the information available, and would again require the use of multiple unnecessary assumptions. The impacts of acute exacerbations on the patient and healthcare system are: • Mortality – already captured within the OS projections used within the model. • FVC decline – already captured within the PFS measure included in the model. • Hospitalisation cost – already captured within the model using data on hospitalisation days and length of stay from the clinical trials. • Decrement to quality of life during the acute phase – already captured within the model separately in addition to the above. We note the following statements within the ACD in Section 4.10 *"The committee agreed that these events were likely to be linked, so it was not appropriate to model them independently"* and *"The committee agreed that the model may have underestimated the impact of exacerbations."* We would strongly disagree with both of these statements: • Modelling events independently is a common methodology employed when more data are available for solid outcomes	

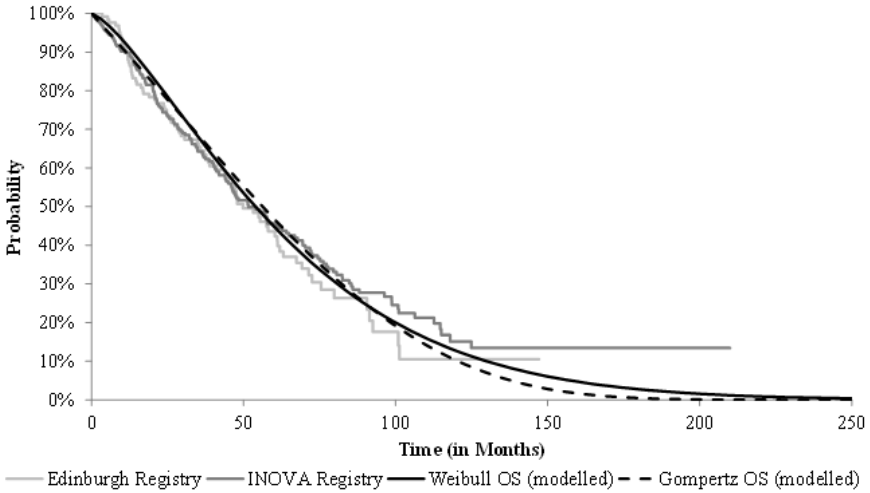
Consultee	Comment [sic]	Response
	(such as OS) than for potential surrogates which might be used to estimate them (such as change in FVC). The partitioned survival approach submitted has previously been accepted by NICE within multiple appraisals for exactly this reason. • As shown above the impact of exacerbations is most likely adequately captured through the use of solid outcome data within the model, however, if this was in fact underestimated we note that the direction of bias is against pirfenidone (i.e. ICERs should be reduced) as pirfenidone reduces the rate of acute exacerbations vs. best supportive care.	
Roche	**<u>Evidence that a long-term treatment effect with pirfenidone is plausible</u>** Long-term follow-up data for pirfenidone are available through the three pivotal Phase III placebo-controlled randomised controlled trials CAPACITY 1, CAPACITY 2 and ASCEND, along with the RECAP extension study. Overall survival data for pirfenidone in these trials extends up until approximately 8 years, as demonstrated for the population of patients with FVC ≥ 50%-predicted. Figure 2 shows the comparison of the Kaplan-Meier plots with curve fits using the Weibull distribution, and assuming a lifetime duration of treatment effect. Included are the standard curve fits previously included in the submission, as well as a curve fit using a hazard ratio for pirfenidone vs. BSC based on the INOVA registry, after trimming data using a propensity score model to adjust for remaining imbalances. This was performed to address the comparability of the entire registry to patients within the context of available trials, and was described in Section 4.10 of the evidence submission. The figure shows that post-adjustment, short-term outcomes are more similar to the projected outcomes for BSC patients using the placebo arm of the clinical trials, and long-term outcomes are predicted to be slightly lower than the registry itself based upon the survival curve fit type assumed for pirfenidone (Weibull). As previously discussed in Section 5.3 (page 210) of the evidence submission, registry data were used to validate long-term survival outcomes for BSC patients, as follow up for BSC patients within the context of the trials is limited to approximately two years. Of the 624 patients randomised to the BSC arms of the trials, only 54 (that is, 8.7% of all BSC patients) died, resulting in high levels of censoring and uncertain outcomes for BSC patients if trial data alone are used. As discussed at the Appraisal Committee meeting, the registries themselves were not used to provide estimates of OS in the model base case, but rather served as reference material for long-term outcomes had trial data for BSC patients continued beyond 2 years. Within the ERG report, the ERG present optimistic (lifetime) and pessimistic (two years) assumptions regarding the duration of treatment effect for pirfenidone. The ERG based this on the availability of RCT data up until approximately two years. However, we do not believe the ERG have appropriately considered the availability of long-term OS data for patients treated with pirfenidone, which demonstrate an ongoing treatment effect up until the end of available pirfenidone data, as presented in Figure 2.	Comment noted. After considering the comments received in response to the ACD, the committee considered the potential long-term mortality benefit with pirfenidone by extrapolating from relatively short trials over a patient's lifetime and the assumption that the mortality benefit of pirfenidone compared with best supportive care remains constant over a person's lifetime. Please see the committee's considerations outlined in sections 4.15 and

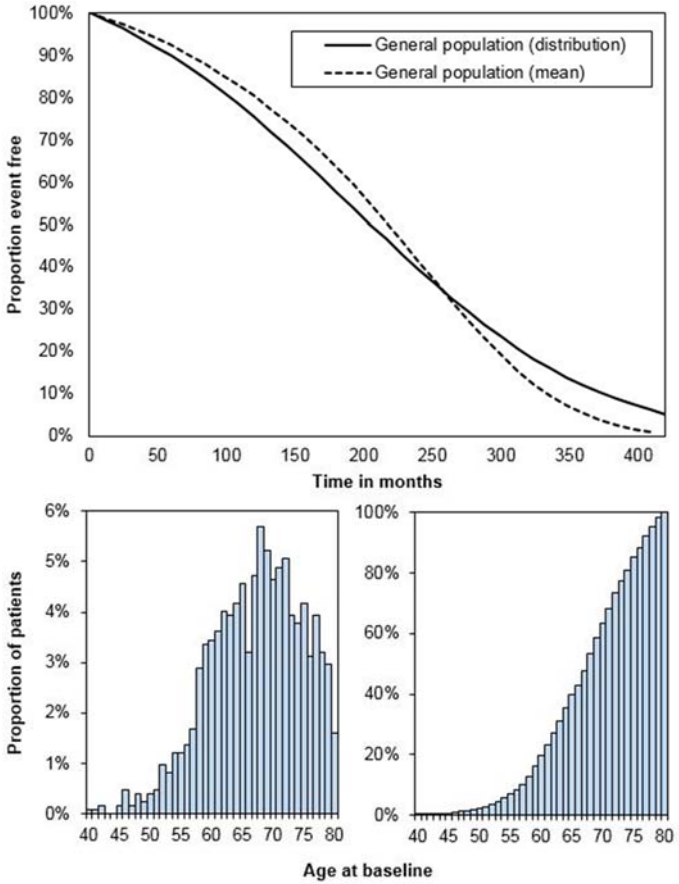
Consultee	Comment [sic]	Response
	Figure 2: Comparison of Kaplan-Meier plot of overall survival for pirfenidone patients with hazard ratio for pirfenidone vs BSC based on INOVA registry (post trimming data using propensity score model to adjust for remaining imbalances)  Key: BSC, best supportive care; PFN, pirfenidone. **Figure 3: Log-cumulative hazard plots for OS within the ASCEND/CAPACITY/RECAP trials for pirfenidone, and for the INOVA registry for BSC**	4.16 of the FAD.

Consultee	Comment [sic]	Response
	 When looking solely at within trial data from CAPACITY and ASCEND, as previously stated there was no significant interaction observed between treatment effect and time. This applies within the pooled dataset and individually within CAPACITY 2 and ASCEND. Furthermore, no convergence of curves is observed based upon the log-cumulative hazard plots for these trials (Figure 3). There is no clinical rationale why the treatment effect of pirfenidone would diminish in the long-term. Unlike treatments used in the treatment of cancers, which this Committee may frequently assess, within IPF there is no mechanism by which resistance to treatment may develop over time, and this is supported by expert clinical advice sought as part of this appraisal. Additionally there is no evidence of the treatment effect of pirfenidone diminishing (if one does not account for CAPACITY 1 where, as stated by the clinical expert at the Appraisal Committee meeting, the placebo arm performs differently to the placebo arms observed in all other trials in IPF). We also note that diminishment of treatment effect was not explored within the nintedanib appraisal, we assume due to the similar lack of a mechanism by which this would occur. When all relevant datasets are considered at the end of follow up data, it can be seen that a pessimistic assumption	

Consultee	Comment [sic]	Response
	regarding the duration of treatment effect may be considered as approximately eight years. An assessment of a combination of Figure 2 and Figure 3 demonstrates evidence of a treatment effect of at least eight years, and this is supported by the totality of the evidence provided through the NMA, RECAP and INOVA registry data. Based on these data, curve fits are presented in Figure 4, assuming a duration of treatment effect of eight years. Weibull (our preferred option) and Gompertz (ERG's preferred option) curve fits are shown. Economic analyses are also presented using this reduced treatment effect in order to provide a more plausible pessimistic scenario for the Committee's assessment. **Figure 4: Kaplan-Meier plot of overall survival for pirfenidone patients (left to right; Weibull and Gompertz)**  	
Roche	**<u>Evidence that the survival curve selected by the Committee on the advice of the ERG to model long-term outcomes with pirfenidone is not supported by the available data</u>** As highlighted as part of our factual accuracy response to the ERG report, along with comments during the Appraisal Committee meeting, we strongly disagree with the use of the Gompertz distribution to model long-term outcomes, and do not believe this is supported by the available evidence. Furthermore, the ERG's justification for choice of distribution appears to be based on inaccurate analyses. Our rationale for this view, along with further support for the use of the Weibull distribution, is set out below: *(i) In-appropriateness of ERG analysis for use of the Gompertz curve*	Comment noted. After considering the comments received in response to the ACD, the committee considered the potential long-term mortality

Consultee	Comment [sic]	Response
	As highlighted by the manufacturer representatives during the Appraisal Committee meeting, we do not believe the evidence presented by the ERG in relation to the comparison of hazards for different curve fits and Office for National Statistics (ONS) general population data is appropriate. The figure presented by the ERG within their report (Figure 38) has been reproduced using the distribution of age from within the trials in Figure 5 below. **Figure 5: Revision of Figure 38 from ERG report: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and distribution-adjusted life tables in the UK in the ITT population (Revision of plot originally produced by the ERG)**  This revised plot demonstrates that there is no issue relating to the Weibull distribution producing hazards which cross those of the general population within the model time horizon. The ERG commented that from Figure 38 of their report, it can be seen "*that the use of the Weibull distribution in the model leads in some occasions to lower probabilities of death in people with IPF initiating pirfenidone compared with the probability of death from the general population*." We would consider this argument to be inappropriate as the earliest time at which the lines for the Weibull curve and the ERG-produced (i.e. non-distribution adjusted) general population cross is at approximately age 90 (i.e. 23 years or 276	benefit with pirfenidone by extrapolating from relatively short trials over a patient's lifetime and the assumption that the mortality benefit of pirfenidone compared with best supportive care remains constant over a person's lifetime. Please see the committee's considerations outlined in sections 4.15 and 4.16 of the FAD.

Consultee	Comment [sic]	Response
	months into the model time horizon). It would be expected that by this time, only a small proportion of patients would still be alive within the model. This assumption is verified by considering the figure presented by the ERG within their report for long-term OS extrapolations, re-presented in Figure 6. **Figure 6: Re-presentation of Figure 37 of ERG report: Plot of the KM for OS from registries and modelled survival for BSC using the Weibull and Gompertz distribution (Plot drawn by the ERG)**  Key: BSC, best supportive care; ERG, Evidence Review Group; KM, Kaplan-Meier; OS, overall survival. Furthermore, the ERG used general population data based on the mean age at treatment initiation, which does not consider the distribution of age at baseline. The difference in these curves are presented in **Error! Reference source not found.**, along with the distribution of age at baseline in distributional and cumulative form. *(ii) Lack of clinical plausibility of the Gompertz curve* The Gompertz curve fit appears to disregard the fact that IPF registries indicate a proportion of patients surviving for a substantial length of time post diagnosis. For example approximately 13% of patients are still alive after 17 years in the INOVA registry. Within the ERG report, it is stated that adjustment of registry data likely biases OS outcomes in favour of pirfenidone (Section	

Consultee	Comment [sic]	Response
	5.2.2.5 on page 175), suggesting that registry data used within the model represent a more severe group of patients than those in the CAPACITY and ASCEND trials. This statement is simply incompatible with the selection of a survival curve which predicts lower mortality than the registry. The Committee themselves spoke to a patient who has so far survived 10 years from diagnosis at the meeting. The Weibull curve takes into account the existence of this tail in its long-term extrapolation.  *(iii) Conclusion on most appropriate curve to use for survival analysis* We propose the Weibull distribution, with a duration of treatment effect of approximately eight years, provides a statistically	

Consultee	Comment [sic]	Response

good fit to the observed data (43.6 times more probable than the Gompertz curve to be the best fitting curve based on AIC scores), good visual fit to observed data from both the trials and registry as well as realistic long-term extrapolation.

(iv) Alternative analysis for consideration

Whilst on balance we consider the Weibull curve be the best representation of long-term survival, based upon statistical goodness of fit and clinical plausibility, in order to provide the Committee with a more statistically robust estimation of long-term survival, an average curve based on weightings identified through AIC scores is presented to address the uncertainty in an individual curve fit choice. This method utilises the theory behind AIC to provide a weighted average curve based on the probability of the given curve providing the best fit to observed data. The weights are presented in Table 3.

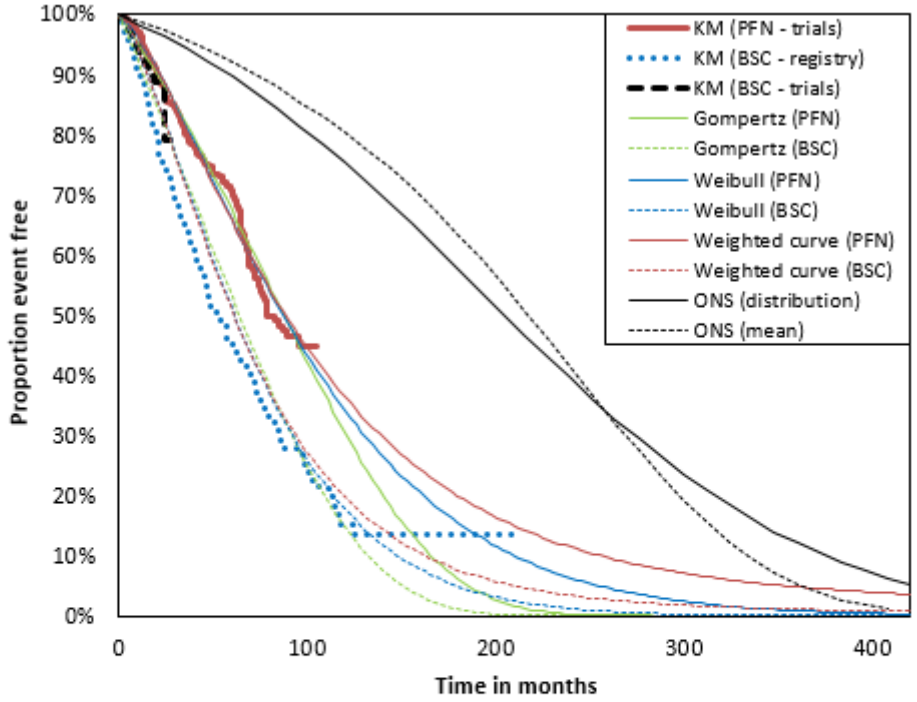
Table 3: Derivation of weighted average curve

Distribution	OS AIC	Probability of best fit	Weight
Curve fit	AIC_i	$P_i = e^{(Best\ AIC - AIC_i)/2}$	$\frac{P_i}{\sum_{j=1}^{n} P_j}$
Exponential	865.47	0%	0%
Weibull	844.15	100%	43%
Log-Normal	853.23	1%	0%
Gamma	845.78	44%	19%
Log-Logistic	844.54	82%	36%
Gompertz	851.70	2%	1%

Key: AIC, Akaike Information Criterion; OS, overall survival.

The next best statistical fits after Weibull were the Log-logistic and Gamma distributions, and consequently the weighted average curve produces survival estimates above that of the Weibull and Gompertz distributions alone, as shown in Figure 7. Therefore, it may be derived that the Weibull distribution provides a reasonable estimate between the lower bound provided by the Gompertz distribution and an upper bound provided by the weighted average curve. As can be seen, for both the Weibull and Weighted average curves, estimates remain below general population mortality when the age distribution of patients in the clinical trials is taken into account (Figure 7).

Figure 7: Comparison of overall survival estimates used within the model based on Weibull, Gompertz and Akaike Information Criterion weighted curves

Consultee	Comment [sic]	Response
	 **Proportion of patients still alive** Gompertz (PFN) 100.0% 42.9% 2.5% 0.0% 0.0% Weibull (PFN) 100.0% 43.8% 11.4% 2.3% 0.4% Weighted (PFN) 100.0% 45.0% 16.2% 7.0% 3.9% Gompertz (BSC) 100.0% 25.6% 0.3% 0.0% 0.0% Weibull (BSC) 100.0% 26.4% 3.0% 0.2% 0.0% Weighted (BSC) 100.0% 27.6% 5.6% 1.8% 0.8% Key: BSC, best supportive care; ERG, Evidence Review Group; KM, Kaplan-Meier; OS, overall survival; PFN, pirfenidone.	
Roche	**Results of revised economic analyses & presentation of new patient subgroup** Based upon the discussions above, revised analyses are presented for the Committee's consideration. These explore the	Comment noted. After considering

Consultee	Comment [sic]	Response
	uncertainty associated with curve fit, and the duration of treatment effect. New analyses are also presented for the subgroup of patients with FVC < 90%-predicted: in line with the discussion at the Appraisal Committee meeting, with represents the group of patients in whom the majority of clinical trial data for pirfenidone are available. *Incorporation of changes from the ERG report* The revised analyses presented below incorporate all changes highlighted in the ERG report, including the use of NMA results for OS at 72 weeks, but excluding the following features: • The choice of curve for OS – Weibull, the weighted average curve and Gompertz curves are presented, in line with Committee request • The ERG-produced CODA sample The CODA sample produced by the ERG contained hazard ratios varying between 0.09 to 7.19 for OS and 0.00 to 30.96 for PFS. We do not consider the extremes demonstrated within these estimates to be clinically plausible. These estimates are drawn from CODA samples using estimates from the predictive distributions, which was the ERG's preference in their report but did not lead to consensus among experts at the 1st Appraisal Committee meeting (5 May 2016). In addition, the ICERs were calculated using means which are influenced by outliers (in contrast to medians). With such a small network of trials, the predictive distributions will strongly depend on the choice of priors and guidance on choice of informative priors is limited ((Dias, 2011; p39-40). Overall, we believe: the between-study heterogeneity is already taken into account by a random effects model; using predictive distributions is excessive, and; our original approach – where posterior medians and 95% credible intervals are used within economic analysis – is in line with previous submissions (NICE 2013a, NICE 2016b). *Analyses conducted* The updated model (available upon request) has been produced using probabilistic ICERs based on the CODA samples from the NMA conducted by Roche as part of the initial evidence submission. Results are presented with incorporation of the pirfenidone PAS used within TA282: that is, a discount of [redacted] on the list price of pirfenidone. As a condition of supplying new analyses, NICE requested that analyses for the patient subgroups presented within the initial submission are also re-run and supplied as part of this response to the ACD [email correspondence received from NICE on 14 June 2016]. Therefore, the following subgroups according to FVC %-predicted were explored within the revised economic analysis outlined above: • ITT: Patients with FVC %-predicted ≥ 50 • Patients with FVC %-predicted ≥ 80 • Patients with 50 ≤ FVC %-predicted < 80	the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee agreed that the recommendations in NICE's previous technology appraisal guidance on pirfenidone remained appropriate for people with idiopathic pulmonary fibrosis with an FVC between 50% and 80% predicted. Please see the committee's considerations outlined in sections 4.5, 4.18 to 4.20 of the FAD.

Consultee	Comment [sic]	Response
	• New analysis: Patients with 50 ≤ FVC %-predicted < 90 The latter subgroup was explored to demonstrate results for the population of patients from which the majority of clinical data are available (since patients with FVC ≥ 90%-predicted were excluded from the ASCEND study). Further details regarding the consideration of this subgroup of patients are presented in the "Additional analysis in subgroup where most data are available (FVC< 90%-predicted)" section. Scenario analyses are supplied for the Committee's consideration around the key aspects of uncertainty identified by the Committee (survival curve fit, treatment effect duration, stopping rule) in line with the ACD: • As described above, analyses are performed for three different parameterisations of the OS curve: Weibull, Gompertz and weighted average distribution (see Table 3) • The duration of treatment effect is explored, with the pessimistic scenario defined as a duration of treatment effect of 8 years (as discussed in the section "**Error! Reference source not found.**") and an optimistic scenario defined as a lifetime duration of treatment effect (as presented by the ERG in their report). • Inclusion/exclusion of the stopping rule for pirfenidone is also considered. The Committee considered these scenarios to form bounds between which the true ICER is likely to lie, if the stopping rule were to be implemented in clinical practice). In addition to the base case results presented in the tables below, plots of ICERs by duration of treatment effect were produced to assist the Committee in their decision making. These plots show at which point the duration of treatment effect is long enough in order for pirfenidone to appear cost-effective at a threshold of £30,000 per QALY gained. ***[The consultee submitted several pages of revised cost-effectiveness results and references in its response to consultation which have not been reproduced here. Please see Committee papers for the full response]***	
British Thoracic Society	**Has all of the relevant evidence been taken into account?** The stopping rule: The Society is concerned that although change in FVC is considered to be a suitable endpoint in clinical trials, there is considerable intra-subject variability, which limits its usefulness in disease monitoring on an individual basis (Raghu G, Collard H *et al Am J Respir Crit Care Med* 2011) Data from the recently published paper by Nathan SD *et al* (*Thorax* 2016; 71:429) showed that FVC change during the first 6 months of pirfenidone treatment was not predictive of change during the second interval. 59 patients had decline in FVC of 10% or more during the first 6 months of treatment. Importantly however, in 16 of these patients, FVC stabilised or improved with continued treatment for 6 months. Fewer patients in the pirfenidone treatment group compared to placebo group experienced a second >10% decline in FVC (2/34 in pirfenidone group compared to 19/68 in the placebo group). More patients in the treatment group had no further decline in FVC. This suggests that in IPF patient with progressive disease	Thank you for your comments. After considering the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee

Consultee	Comment [sic]	Response
	continued treatment with pirfenidone may reduce the risk of subsequent decline or death. It is unknown what an individual's decline in FVC would be without treatment. Hence, >10% decline in FVC may be a treatment response depending upon the starting FVC, i.e. treatment changes the trajectory of FVC decline. The effect of an exacerbation which may result in FVC declining by >10% should also be considered. For nintedanib, the effect of treatment is equivalent regardless of starting FVC. **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?** We are uncertain if the committee considered the pooled mortality data from ASCEND and CAPACITY studies – pirfenidone reduced all-cause mortality and respiratory-related mortality after 1 year of treatment. This is likely to impact the cost effectiveness model. **Are the provisional recommendations sound and a suitable basis for guidance to the NHS?** The Society would support expansion of the current treatment guideline to include patients with FVC>80%. In many cases the starting FVC is unknown so FVC>80% often reflects clinically significant disease. It's important to note that FVC can be artificially elevated by the presence of co-existent emphysema, so this patient cohort is unfairly disadvantaged. In addition there is concern about the accuracy of current lung function measurements and analysis in an older population. In some patients, change in TLco (DLco) is a more sensitive measure of disease progression rather than change in FVC. Data from the BTS ILD registry shows that 40% patients have an FVC >80% , so this represents a substantial number of patients who are not eligible for treatment. In the Netherlands, treatment guidelines are similar, but, for patients with FVC>80%, who have progressive disease (for example starting FVC 120% with documented evidence of progression radiologically and/or physiologically), their suitability for anti-fibrotic therapy is evaluated by an expert panel. This may be an option to consider; a defined yearly envelope of funding would ensure that there is clear attention given to which patients the expert panel "sign-off", rather than just everyone.	concluded that pirfenidone was not cost effective without the stopping rule. Please see the committee's considerations outlined in sections 4.3, 4.6, 4.12, 4.13 and 4.18 of the FAD. The committee was aware that the company and the evidence review group (ERG) included data from ASCEND in their network meta-analyses with data from CAPACITY 1, CAPACITY 2 and SP3. Please see the committee's considerations

Consultee	Comment [sic]	Response
		outlined in sections 4.8 of the FAD.
Department of Health	No comment	Noted.
Royal College of Nursing	**Has the relevant evidence has been taken into account?** The evidence considered seems comprehensive. ii) **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?** We would ask that the summaries of the clinical and cost effectiveness of this appraisal should be aligned to the clinical pathway followed by patients with idiopathic pulmonary fibrosis. The preliminary views on resource impact and implications should be in line with established standard clinical practice. iii) **Are the provisional recommendations sound and a suitable basis for guidance to the NHS?** Nurses working in this area of health have reviewed the recommendations of the Appraisal Committee and do not have any other comments to add. The RCN would welcome guidance to the NHS on the use of this health technology. Iv) **Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of gender, race, disability, age, sexual orientation, religion or belief?** Given the robustness of the evaluation and taking opinions into consideration the draft recommendations are probably not unexpected although it is somewhat disappointing that the upper limit of forced vital capacity - FVC >80% remains because of the confounder of emphysema so these patients whist preserving their FVC actually can do badly.	Thank you for your comment. Comment noted. Comment noted. The committee recognised the limitations of FVC but understood

Consultee	Comment [sic]	Response
	Applying a FVC cut-off at 80% removes a significant number of patients that have IPF with some emphysema, they present with a preserved at best and or inflated FVC. As treatment with Pirfendione only slows progression, it does not seem sensible to wait for advanced disease before starting treatment. The role of FVC as an outcome measure for efficacy (there is reasonable evidence for this) but as a sole measure of severity, there is no evidence. There is much better evidence that the combination of DLCO and FVC provides a better measure of severity than FVC alone (Ley B. Ann Int Med 2012;156:684-691). We agree with the stoppage of the medication if it is not affecting lung function decline. v) **Are there any equality-related issues that need special consideration that are not covered in the appraisal consultation document?** We are not aware of any specific issue at this stage. We would ask that any guidance issued should show that an equality impact analysis has been considered and that the guidance demonstrates an understanding of issues relating to all the protected characteristics where appropriate. Lastly, we already have a drug which meets the >50% <80% criteria, we need to treat the milder patients earlier and raise the upper limit to at least 90%. There is a need for more drugs in the management of IPF, even if their efficacy is similar. Prognosis is poor for the patients, early treatment is paramount	that, in clinical practice, wider patient characteristics would be taken into account when interpreting percent predicted FVC. It concluded that its recommendations did not discriminate against any groups of people protected by the Equality Act. Please see the committee's considerations in section 4.22 of the FAD. After considering the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee agreed that the recommendations in NICE's

Consultee	Comment [sic]	Response
		previous technology appraisal guidance on pirfenidone remained appropriate for people with idiopathic pulmonary fibrosis with an FVC between 50% and 80% predicted. Please see the FAD document for further details.

Comments received from Commentators

Commentator	Comment [sic]	Response
Boehringer Ingelheim	**Has all of the relevant evidence been taken into account?** Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease. **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?** Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease. **Are the recommendations sound and a suitable basis for guidance to the NHS?** Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease. **Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?** ACD sections 4.5, 4.6 and 4.17 focus on the population of patients with percent predicted FVC of >80%, and calls them 'mild' patients. Although this terminology has been used in clinically and in NICE TA282 and TA379, it risks not conveying the seriousness of the condition in these patients. IPF is progressive and poor-prognosis disease with a median survival estimated at around 3y (range: 2-5y) across stages. Early stage (>70% predicted FVC patients) patients have a median survival of around 55 months (less than 5y). This is worse than many early, or even late, stage cancers. And it is progressive. Hence, just as no 'mild cancer' terminology exists, calling such a group of patients 'early stage IPF' rather than 'mild IPF' might be more appropriate. At present, NICE does not recommend any active treatment for these early stage (nee 'mild') IPF patients; in effect, limiting patient and clinician choice of therapy in this group of early stage disease in a progressive poor-prognosis condition to 'best supportive care' (BSC, which is effectively no active treatment; especially as previous treatment options like N-Acetyl Cysteine have been shown to be potentially harmful to IPF patients). This is so even though active treatments, like nintedanib and pirfenidone, have been shown to be effective and is licenced in this population of early stage IPF patients.	Thank you for your comment. After considering the comments received in response to the ACD in conjunction with the new evidence submitted by the company, the committee agreed that the recommendations in NICE's previous technology appraisal guidance on pirfenidone remained appropriate for people with idiopathic pulmonary fibrosis with an FVC between 50% and 80% predicted. Please see the FAD for further details. Where possible, the terminology 'mild' has been removed from the FAD. However, we note that pirfenidone has a marketing authorisation in the UK for treating 'mild to moderate' idiopathic pulmonary fibrosis in adults.

Commentator	Comment [sic]	Response
	We understand that in this assessment the case for pirfenidone has not satisfied the appraisal committee for its clinical effectiveness or cost effectiveness versus BSC in this population. We would, however, urge NICE and the committee to reconsider both the evidence for nintedanib and pirfenidone in this population of early stage IPF patients to provide an option for slowing progression of this poor-prognosis condition even in early stages of IPF instead of letting patients progress to a threshold (80% predicted FVC in this case) and then be able to be started on active treatment. This leaves an unmet need for early stage IPF patients despite two active treatments licensed in this indication. We would also suggest an assessment from an equality standpoint versus comparable poor-prognosis indications (e.g. some cancers).	

Comments received from clinical experts and patient experts

No comments were received from clinical experts and patient experts

Comments received from members of the public

No comments were received from members of the public

1st Floor,
10 Spring Gardens,
London
SW1A 2BU

BY EMAIL

24th June 2016

RE: Idiopathic pulmonary fibrosis – pirfenidone (review of TA282) [ID837]

Dear Meindert,

Thank you for the opportunity to comment on the Appraisal Consultation Document (ACD) for the appraisal of pirfenidone (Esbriet®) in the treatment of idiopathic pulmonary fibrosis (IPF). We are pleased to see that NICE recognised the continued role for pirfenidone in the management of patients with IPF, for those with a forced vital capacity (FVC) between 50-80%-predicted.

We do, however, remain concerned that the Committee has failed to recognise the importance of treating patients with earlier stages of IPF (FVC ≥ 80%). Idiopathic pulmonary fibrosis is a chronic, progressive, and fatal lung disease that is characterised by irreversible loss of lung function. Early treatment to delay progression should, therefore, be an important goal for the management of the condition. Clinical opinion strongly advocates for earlier access to treatments, but this is not reflected in the prevailing guidance from NICE, which is ultimately to the determinant of patients.

We are also concerned with several aspects of the approach taken to the assessment of the evidence by the ERG, primarily related to what appears to have been the starting point of their assessment: i.e. the existence of patient subgroups (defined by FVC %-predicted), despite no clinical evidence or opinion to support such an approach. We believe assessment of the clinical data should begin with consideration of the totality of the data. Lack of any evidence to suggest a

Roche Products Limited
6 Falcon Way
Shire Park
Welwyn Garden City
Hertfordshire
AL7 1TW
Health Economics and
Strategic Pricing
Tel. +44(0)1707 361032
Fax +44(0)1707 384123

difference in treatment effect based on FVC at baseline, combined with the clinical view that any threshold level is arbitrary and to the detriment of patients with preserved FVC %-predicted, is supportive of an approach to consider the entirety of the data, and is in line with NICE's Guide to Methods. In this respect, the review seems to have been artificially limited by the historical recommendations of TA282. This was developed prior to the availability of results from the ASCEND trial, which provides a substantial body of evidence for the effectiveness of pirfenidone across groups of patients defined by levels of FVC %-predicted.

Given the recommendations of the ACD, and the prior agreement with the Department of Health regarding the proposed patient access scheme (PAS), analyses were re-run accounting for the PAS defined in TA282. These analyses were based on the ERG-preferred settings of the model, with the exception of a corrected CODA sample, and use of the most plausible parameterisation of the overall survival (OS) curve (Weibull: explanations provided below). Revised ICERs for the ITT population (assuming a lifetime treatment effect) were £18,167 with the stopping rule for pirfenidone, and £25,986 without the stopping rule. In order to achieve an ICER of approximately £30,000, the duration of treatment effect needed would be four and eight years (including and excluding the stopping rule, respectively). Both durations are clinically plausible, as they lie within the range of evidence from the pirfenidone clinical trial programme and comparable registry data.

During the Appraisal Committee meeting, there was discussion that the majority of evidence from the ASCEND and CAPACITY trials is from patients with FVC < 90%-predicted. Analyses were, therefore, performed to demonstrate the cost-effectiveness of pirfenidone within this patient population. Within this subgroup, the ICERs ranged between an optimistic £16,676 (weighted OS curve, lifetime duration of treatment effect and stopping rule for pirfenidone) and a pessimistic £34,267 (Gompertz OS curve, eight-year duration of treatment effect and excluding the stopping rule).

The following document provides further detail on our concerns with the assessment and interpretation of the evidence supporting the review of pirfenidone, and suggested approaches to allow the Committee to make a more considered recommendation.

Yours sincerely,

Denzyl Cain, Head of Health Economics and Strategic Pricing

Roche Products Limited

Inappropriate focus on subgroup analyses, without evidence to support their existence

The final scope for this appraisal stated that – subject to the evidence available – assessment would be made of patient subgroups by disease severity, defined by FVC and/or diffusing capacity for carbon monoxide.

As discussed at the Appraisal Committee meeting, there are no accepted thresholds of FVC %-predicted used to define the disease severity of a patient with IPF in UK clinical practice, although there is a general acceptance that an FVC <50% predicted and DLco <35% predicted defines severe disease. Other staging systems include the GAP index, which includes age and gender as predictors of mortality, along with %-predicted DLco (Ley 2012). The composite physiology index adds forced expiratory volume in 1 second (FEV1) to FVC and DLco predicted values (Wells 2003). Consequently, including all characteristics of potential importance (as indicated by these alternate staging systems) is inherently associated with numerous theoretical and methodological caveats.

Based on an assessment of the evidence available at the time of the first appraisal for pirfenidone in 2012-2013, NICE issued guidance which restricted use of the treatment to patients with an FVC < 80% predicted (NICE 2013a). As described in paragraph 4.16 of the current ACD, the Committee are still confident in the data supporting the initial review: "*In the original appraisal of pirfenidone, the committee had concluded that pirfenidone was clinically effective for moderate disease. In the current appraisal, the committee decided it had seen no evidence to alter that conclusion*". In line with the stated objective of this 2016 appraisal (paragraph 4.1), this review provides the opportunity for NICE to assess the totality of evidence to extend the Guidance for pirfenidone to patients facing a significant unmet need.

Since publication of TA282 in April 2013, the ASCEND study has reported results; this study included approximately 100 patients with a FVC of ≥ 80% and < 90% predicted, a group of patients with no active treatment for management of their IPF. Importantly, when pooled with earlier data from the CAPACITY trials, ASCEND roughly doubled the number of patients available for assessment within this group, allowing for a robust analysis of any important differences in treatment effect between this group and the larger cohort. The cohort of patients with baseline FVC up to 90% predicted was discussed at the Appraisal Committee meeting, although this is not specifically captured in the ACD. Analyses of the clinical- and cost-effectiveness of pirfenidone in this group of patients – which account for approximately 90% of the pooled pirfenidone clinical trial population – are provided below.

Using the most statistically appropriate methods, no difference in the efficacy of pirfenidone by patient subgroup has been identified [see following section], and we believe the data from the trial populations should be assessed in their entirety.

This view is supported by information presented to NICE by a variety of stakeholders during the evidence submission stage, the Appraisal Committee meeting on 5th May, along with that heard during the appraisal of nintedanib (TA379), where it was described that there is no clear definition of what constitutes ‘mild’ or ‘moderate’ IPF, particularly when characterised by FVC %-predicted alone (NICE 2016a, NICE 2016b).

The NICE Guide to Methods includes clear direction on the identification and assessment of subgroups. Whilst the Methods Guide may have been designed in the context of first / new appraisals for a treatment, the clear starting point is for an assessment across the licensed population: subgroups are to be identified within that population, based on criteria described in the

Guide:

> *"5.10.1 …. The characteristics of patients in the subgroup should be clearly defined and should preferably be identified on the basis of an expectation of differential clinical or cost effectiveness because of known, biologically plausible mechanisms, social characteristics or other clearly justified factors.*
>
> *5.10.2 … There should be a clear justification and, if appropriate, biological plausibility for the definition of the patient subgroup and the expectation of a differential effect. Post hoc data 'dredging' in search of subgroup effects is to be avoided and will be viewed sceptically.*
>
> *5.10.6 The standard subgroup analyses performed in RCTs or systematic reviews seek to determine whether there are differences in relative treatment effects between subgroups (through the analysis of interactions between the effectiveness of the technology and patient characteristics). The possibility of differences emerging by chance, particularly when multiple subgroups are reported, is high and should be taken into account. Pre-specification of a particular subgroup in the study or review protocol, with a clear rationale for anticipating a difference in efficacy and a prediction of the direction of the effect, will increase the credibility of a subgroup analysis.*
>
> *5.10.7 In considering subgroup analyses, the Appraisal Committee will take specific note of the biological or clinical plausibility of a subgroup effect in addition to the strength of the evidence in favour of such an effect (for example, if it has a clear, pre-specified rationale and is consistent across studies)… "*

We are concerned that the assessment of the evidence by the ERG – and seemingly supported by the Committee – has been conducted from a starting point of the existence of patient subgroups. This may be an artefact of the recommendations made in TA282. We believe that, if due consideration was given to the rationale for those initial recommendations, the additional evidence

provided through ASCEND, along with the prevailing clinical opinion, a fresh perspective should have been taken to this review. We attempted to provide this within our evidence submission through presentation of results related to the ITT population, but the current assessment has focussed on unproven differences in subgroup treatment effects as a means to justify subgroup analysis: this is in contrast to the approach set out in NICE's Guide to Methods.

Indeed, the ACD states "*The committee understood that the results of the treatment-by-subgroup interaction tests were not significant in either of the subgroup analyses. However, it heard from the ERG that a non-significant interaction test does not conclusively mean that there is no difference in treatment effect between subgroups. The ERG explained that the interaction test may not have been powered to detect a difference between the subgroups. The committee concluded that it did not see robust evidence that pirfenidone is clinically effective in people with mild idiopathic pulmonary fibrosis (that is, an FVC above 80% predicted)*" (paragraph 4.16).

Consistent with the NICE Guide to Methods, there is no biologically plausible mechanism whereby either the efficacy of pirfenidone and/or the capacity to benefit would differ between patients with differing FVC levels. IPF is a complex disease that is not yet fully understood, and research is hampered by the lack of a model which fully represents the human disease (White 2016). Notwithstanding this, the complex pathogenic cascade leading to the development of fibrosis in IPF involves numerous mediators and signalling pathways, and it is likely that an effective IPF therapy would need to target more than one mechanism/pathway. Both pirfenidone and nintedanib have pleiotropic effects, and it is likely their clinical efficacy derives from their broad-based mechanisms of action, which would be inconsistent with a subgroup effect based on lung function parameters alone. As described in the ATS/ERS guidelines: "*While the traditional approach to IPF staging has been useful, it is arbitrary and is not based on epidemiological or biological data. It remains unclear*

if these stages are truly relevant to the management of IPF. Critically, these traditional stages are not known to reflect distinct biological or clinical phenotypes and the true therapeutic and prognostic relevance of these stages remains undetermined" (Raghu 2015)

The intention of paragraph 5.10.6 of the Methods Guide (in the context of 5.10.2) is to seek robust evidence that a subgroup exists, implying a starting point of there being no difference in treatment outcomes across the populations included in the clinical trials. Based on this, the approach taken to the interpretation of the evidence is both surprising and confusing. The quotation taken from 4.16 of the ACD is correct and consistent with paragraph 5.10.6 of the Methods Guide, in that the subgroup interaction test may have been underpowered to detect a difference in treatment effect. However, this underpowering does not prove that there *is* a difference, and we would expect the ERG and NICE to challenge on this basis, were a company to try and argue such a case. The statement in the ACD also does not account for the consistency in treatment effect across FVC subgroup seen across trial outcomes: Table 9 of the Pre-Meeting Briefing document [represented below in Table 1] presents no treatment differences in OS or progression free survival (PFS) outcomes by FVC subgroup, and similar findings for other outcome measures, as presented in Figure 17 of the initial evidence submission (represented as Figure 1 below, NICE 2016a). This figure demonstrates no statistically significant difference in treatment effect across subgroups (defined by FVC %-predicted and GAP stage at baseline) for FVC, 6-minute walking distance (6MWD) and the University of California San Diego Shortness of Breath Questionnaire (UCSD SOBQ)).

Table 1: Representation of Table 9 of Pre-Meeting Briefing: Treatment effect of pirfenidone (overall survival and progression-free survival to week 52), according to baseline disease severity

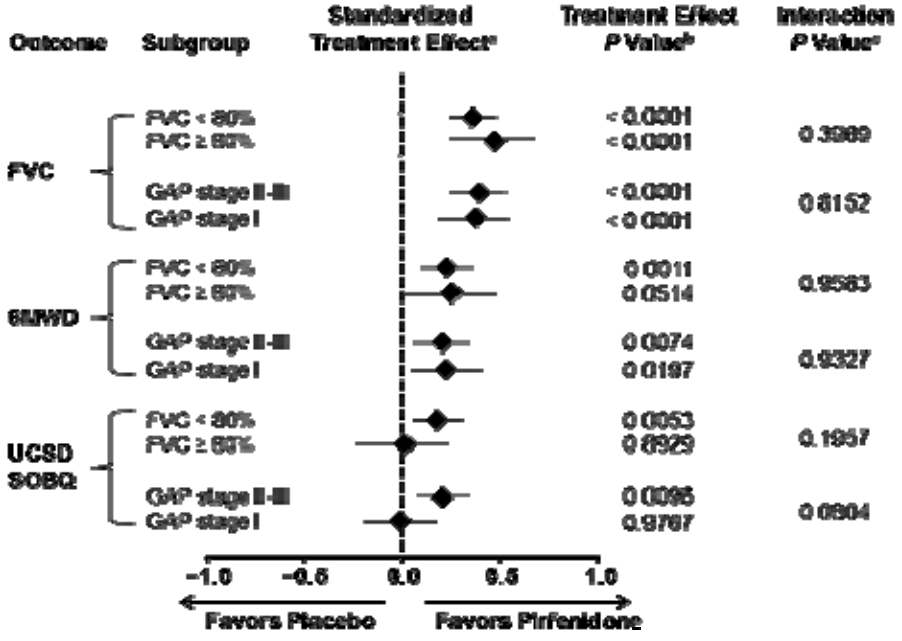
Trial	Percent predicted FVC	Hazard ratio[a] (95% CI)	Treatment effect[b], p value
Overall survival			
ASCEND	≤80%	0.63 (0.29 to 1.34)	0.22
	>80%	<0.01 (0.00 to not evaluable)	0.12
CAPACITY 1	≤80%	0.60 (0.17 to 2.04)	0.41
	>80%	0.77 (0.11 to 5.59)	0.80
CAPACITY 2	≤80%	0.25 (0.08 to 0.76)	0.01
	>80%	Not evaluable	Not evaluable
Progression-free survival			
ASCEND	≤80%	0.56 (0.41 to 0.76)	<0.05
	>80%	0.64 (0.30 to 1.40)	0.26
CAPACITY 1	≤80%	0.84 (0.53 to 1.32)	0.44
	>80%	0.63 (0.29 to 1.41)	0.26
CAPACITY 2	≤80%	0.60 (0.40 to 0.92)	0.02
	>80%	0.40 (0.18 to 0.89)	0.02

[a] hazard ratios below 1 indicate that patients having pirfenidone had a lower risk of the event than patients having placebo

[b] p value indicates significance of the difference between pirfenidone and placebo within the subgroup; p values for treatment-by-subgroup interaction test not reported

CI, confidence interval; FVC, forced vital capacity

Figure 1: Representation of Figure 17 of evidence submission: Treatment effect of pirfenidone by baseline disease severity from pooled data of ASCEND, CAPACITY 1 & 2



We are particularly confused with the statement "*The committee concluded that it did not see robust evidence that pirfenidone is clinically effective in people with mild idiopathic pulmonary fibrosis (that is, an FVC above 80% predicted*" (ACD, 4.16): when considering the change in FVC %-predicted from baseline, the treatment effect of pirfenidone in patients with FVC ≥ 80%-predicted is not only statistically significant, it is also of a similar size to the effect observed in patients with FVC 50-80%-predicted [Table 8 of Pre-Meeting and Table 7-10 of response to clarification question A29].

The *European Respiratory Journal* have also recently accepted for publication a manuscript presenting the efficacy of pirfenidone when used [redacted]. The results presented in the *ERJ* manuscript are in line with those presented in response to clarification question A29.

Furthermore, as the Committee did not consider nintedanib to be a relevant comparator in this review (paragraph 4.6), all relevant comparisons to allow the Committee to make a decision on this cohort of patients were presented within the ITT analysis (i.e. pirfenidone vs. BSC).

Misinterpretation of pre-specified vs. *post-hoc* statistical analyses

In paragraph 4.16 of the ACD, it is stated: "*The committee was aware that the analysis of covariance (ANCOVA) in 2* ***post-hoc*** *subgroups … suggested that pirfenidone was associated with a statistically significant benefit compared with placebo in both subgroups. However, the committee noted that, in the company's* ***pre-specified*** *analysis across 3 subgroups …, there was a nonsignificant tendency for better outcomes in the placebo group than the pirfenidone group among people with a baseline FVC above 80% predicted. The committee was aware of the company's opinion that the analysis with 3 subgroups was not as robust as the ANCOVA method, but the committee agreed that it was not appropriate to disregard a* ***pre-specified*** *analysis. In addition, during the committee meeting, the company could not fully explain the methods of the ANCOVA analysis. The committee understood that the results of the treatment-by-subgroup interaction tests were not significant in either of the subgroup analyses*" (emphasis added).

Firstly, it is important to note that neither approach to the statistical assessment of subgroup differences identified a statistically significant interaction between treatment and patient subgroup: that is, there is no statistical evidence supporting a difference in treatment effect by baseline FVC %-predicted. This is not consistent with the definition of patient subgroups, set out in the Methods Guide (paragraphs 5.10.2 and 5.10.7).

Secondly, it is incorrect to refer to the ANCOVA method as *post-hoc*: the analysis of standardised ranks was the **only** pre-specified analysis in the Integrated Summary of Efficacy (ISE) Statistical Analysis Plan (SAP) for the EMA for **assessment of the primary efficacy analysis and subgroups** [sent to NICE as part of our factual accuracy response to the ERG report]:

- "*Data for the change from Baseline outcomes are not expected to be normally distributed. Therefore, data will be analyzed using a rank analysis of covariance (ANCOVA) model with a*

standardized rank change from Baseline as the outcome and standardized rank Baseline value as a covariate" (Section 6.2, page 12).

- "*The primary efficacy analysis of the pooled data is the rank ANCOVA model for the change from Baseline to Month 12 in %FVC between the 2403 mg/day pirfenidone and placebo groups*." (Section 6.4, page 13).
- "*Interactions of subpopulations with treatment will be tested for in the analysis of change from Baseline to Month 12 in %FVC. Each factor will be tested individually. The factor and interaction with treatment will be added to the rank ANCOVA model for the Month 12 assessment.*"(Section 6.8, page 18)

Assessment based on absolute change in FVC **was not** a specified analysis in the ISE SAP for the EMA. As FVC was not anticipated to be normally distributed, and as some patients were not expected to complete the full 12 month assessment (e.g. due to death), an analysis based on absolute change from baseline for completers of the 12 months follow-up would not be an appropriate method for either the primary endpoint or subgroup interaction tests. This is because such analyses are:

(i) less robust to deviations from the normality assumption, compared to analysis of ranks (via a standardised treatment effect), and;

(ii) likely to miss the effect on FVC, which were not captured in the event of mortality or study discontinuation.

We believe the Committee's focus on assessment of absolute change in FVC may stem from Figure 16 on page 113 of our evidence submission. This related to the assessment of CAPACITY 1 and CAPACITY 2 studies (not ASCEND, nor the pooled analysis of all three trials). Figure 16 was submitted in error, as this was based on the (incorrect) analysis of absolute change in FVC. It

was also inaccurate and potentially misleading for the title on page 114 of our submission to refer to the analysis of pooled subgroups which followed as post-hoc: as described above, the ISE SAP for EMA was planned to assess the primary efficacy outcomes (and subgroups) via analysis of standardised ranks. We apologise for these potentially misleading erroneous inclusions.

It should also be noted that [redacted].

However, regardless of the most appropriate method of statistical assessment, results of both analyses were non-significant, in line with there being no difference in the efficacy of pirfenidone by level of baseline FVC %-predicted.

Additional analysis in subgroup where most data are available (FVC< 90%-predicted)

We do not agree that considering subgroups of patients with IPF on the basis of FVC %-predicted alone is a valid approach, based on the argumentation described above. We do, however, acknowledge the discussion held at the Appraisal Committee meeting, regarding the majority of evidence from the ASCEND and CAPACITY being in patients with FVC < 90%-predicted. This is largely due to the enrolment criteria used in the ASCEND trial (FVC 50-90%-predicted, as opposed to FVC ≥ 50%-predicted in the CAPACITY trials).

As described in the introduction to this response, analysis is presented for patients with IPF and baseline FVC < 90%-predicted. These assessments are based on data from all patients across

the ASCEND and CAPACITY trials, excluding 101 patients within CAPACITY 1 and CAPACITY 2. Similar to the discussion above, there is no evidence of a treatment-interaction when the FVC < 90% subgroup is explored, and there is no clinical or biological rationale why treatment effect should differ in this subgroup. These analyses are solely presented on the basis that this group of patients represents the majority of those included in the pirfenidone clinical trials.

Results of economic analysis using this subgroup are presented within the revised economic analyses in the sections below.

Appropriateness of a survival based model

Section 4.9 of the ACD states the Committee *"would have preferred to see a model that captured the progressive nature of idiopathic fibrosis, and linked clinical outcomes with each other and with time on treatment."* Whilst this might be the Committee's preference there are clear and valid reasons why such a model structure is not ideal for an assessment of the cost-effectiveness of pirfenidone.

When determining a suitable model structure to use for a given decision problem, both the characteristics of the disease and the availability of data should be taken into account. As discussed previously, when considering disease characteristics, it is clear that modelling all characteristics of potential importance would ultimately surpass both available data and available clinical knowledge regarding the interactions between the various factors considered important to prognosis and quality of life (including FVC, DLCo, 6MWD and acute exacerbations).

(i) Choice of model was informed by availability of evidence for OS

The most reliable data available to evaluate the benefits of treatment with pirfenidone come in the form of a hard endpoint: OS. These data show that pirfenidone is currently the only treatment in

IPF to demonstrate a significant OS benefit, with maximum follow-up available for over 8 years as part of the RECAP study. The demonstration of this significant OS benefit for pirfenidone is in fact what triggered this re-review of TA282. Overall survival data are also available from registries of patients with IPF, which include up to 14 years of follow-up for patients treated with best supportive care (BSC). Propensity scoring models were used to adjust for imbalances in patient characteristics.

(ii) Availability of OS data reduces need for model calibration, complexity and assumptions

The strength of the model structure submitted as part of this re-review lies in its simplicity and lack of necessary assumptions: as OS data are modelled independently, a more accurate prediction can be made in line with the clinical trial evidence, without the complication of attempting to link OS to imperfect predictors:

- Within the existing partitioned survival structure, there is no requirement to back calculate OS which can result in inaccuracies in matching observed information from the clinical trials. It is noted that a major limitation of the nintedanib model submitted as part of TA379 was lack of accurate projection of outcomes over time – the assumed mortality rate for pirfenidone (approximately 6.1% based on digitisation of Figure 41 within the nintedanib company submission document) was almost double that actually observed at 1 year (3.6% based on the pooled pirfenidone Phase III trials) (Boehringer Ingelheim 2015). Projected OS for pirfenidone at 5 years was < 50%, compared to observed data indicating 70% survival. The model also failed to accurately project health state split, with consistent bias in terms of a lower proportion of patients modelled in severe health states vs. observed data, and over-prediction in mild health states in both arms.
- Similar issues were observed with the microsimulation model submitted by InterMune as part of earlier review of pirfenidone (TA282), which required the use of calibration factors to provide a

more accurate fit to actual observed data (NICE 2013a).

- These issues underline the benefits of relying on more complete data, with fewer assumptions and less complexity, to produce projections.

Follow up for change in FVC %-predicted over time is restricted to approximately 1 year in the nintedanib and pirfenidone clinical trial programmes.

The model used to support the nintedanib appraisal defined health states for FVC in 10%-predicted categories. As patients could 'skip' FVC categories, this required assessment of patients who experience falls in FVC of ≥ 20%. We do not believe data are available to reliably inform such a model structure. Less than 2% of patients treated with pirfenidone experienced a fall in FVC %-predicted of ≥ 20% in the 12 month trial periods (Table 2). Similarly, there were limited data to inform such progressions in the nintedanib trial programme: less than 5% of patients experienced a fall in FVC of ≥ 20%-predicted. Further splitting these progressions patients into additional categories, such as those defined in the Boehringer Ingelheim company submission, would inevitably lead to major assumptions being made on a very small number of patients (Boehringer Ingelheim 2015).

Therefore, we do not believe that the impact of multiple FVC progressions (using a model structure similar to that presented in TA379, which seems to represent the ERG's preferred approach) simply **cannot** be accurately projected with the available data. Whilst we agree it would be desirable to include such progressions within the model to allow a fuller assessment of the progressive nature of IPF, there is no reliable evidence to justify adding this additional complexity to the model.

Table 2: Patient numbers for change in FVC %-predicted from baseline to 12 months

Change in FVC %-predicted	PBO	PFN	All
Increase of ≥ 30%	0 (0%)	0 (0%)	0 (0%)
Increase of 20-30%	1 (0.2%)	0 (0%)	1 (0.1%)
Increase of 10-20%	6 (1%)	9 (1.4%)	15 (1.2%)
Increase of 0-10%	120 (19.2%)	187 (30%)	307 (24.6%)
Decrease of 0-10%	339 (54.3%)	343 (55.1%)	682 (54.7%)
Decrease of 10-20%	133 (21.3%)	74 (11.9%)	207 (16.6%)
Decrease of 20-30%	21 (3.4%)	9 (1.4%)	30 (2.4%)
Decrease of ≥30%	4 (0.6%)	1 (0.2%)	5 (0.4%)
Key: PBO, placebo; PFN, pirfenidone.			

(iii) Additional complexity does not improve fit to cost or utility data

In addition to data limitations, the additional complexity of attempting to capture progression in terms of FVC-alone is not clinically warranted. The utility regression analysis presented with our submission indicated that inclusion of FVC %-predicted, FEV1, FEV1/FVC ratio, age and gender all harmed model fit: the final equation was informed purely by the St George's Respiratory Questionnaire score (as reported on page 225 within Section 5.4 in the company submission, NICE 2016a).

We note also that within the cost analysis conducted for the nintedanib appraisal, only hospitalisation costs showed change with absolute FVC level, however, these costs are likely to be highly confounded with both acute exacerbation rates and adverse event rates making interpretation difficult.

(iv) Relationship between FVC and mortality is non-linear

It should be acknowledged that the relationship between FVC and mortality is non-linear; both rapidity and absolute levels of change may impact patient prognosis. FVC is also subject to measurement error, meaning that there is a large volume of noise in any individual patient

outcomes. A patient with an FVC of 70%-predicted could well have a better prognosis than one with an FVC of 80%-predicted, if their disease trajectory is one of less rapid decline, or if there are comorbidities, such as emphysema (previously discussed in NICE committee meetings for both pirfenidone and nintedanib).

Additionally, although the mechanism of action of pirfenidone has not been fully established, data suggest that pirfenidone exerts both anti-fibrotic and anti-inflammatory properties in a variety of *in vitro* systems and animal models of pulmonary fibrosis (bleomycin- and transplant-induced fibrosis). Fibrosis often occurs as a result of sustained injury to the epithelium, which causes the overproduction of cytokines and growth factors. This is a core feature shared by pathologic fibrosis among multiple organs tissues such as lung, kidney, and liver. As well as its impact on lung fibrosis, the anti-fibrotic properties of pirfenidone have also been demonstrated in other fibrotic diseases such as diabetic nephropathy and liver fibrosis. This suggests that pirfenidone could have multiple systemic anti-fibrotic effects, and could potentially contribute to an explanation for the non-linearity between FVC change and mortality benefit.

(v) Lack of data to support robust inclusion of acute exacerbations, without clear impact on model results

Modelling of acute exacerbations is even more complex, as FVC decline and acute exacerbations are difficult to distinguish, as discussed in recent NICE Appraisal Committee meetings for pirfenidone and nintedanib. Inclusion of acute exacerbations would also not add to the information available, and would again require the use of multiple unnecessary assumptions. The impacts of acute exacerbations on the patient and healthcare system are:

- Mortality – already captured within the OS projections used within the model.
- FVC decline – already captured within the PFS measure included in the model.

- Hospitalisation cost – already captured within the model using data on hospitalisation days and length of stay from the clinical trials.
- Decrement to quality of life during the acute phase – already captured within the model separately in addition to the above.

We note the following statements within the ACD in Section 4.10 *"The committee agreed that these events were likely to be linked, so it was not appropriate to model them independently"* and *"The committee agreed that the model may have underestimated the impact of exacerbations."* We would strongly disagree with both of these statements:

- Modelling events independently is a common methodology employed when more data are available for solid outcomes (such as OS) than for potential surrogates which might be used to estimate them (such as change in FVC). The partitioned survival approach submitted has previously been accepted by NICE within multiple appraisals for exactly this reason.
- As shown above the impact of exacerbations is most likely adequately captured through the use of solid outcome data within the model, however, if this was in fact underestimated we note that the direction of bias is against pirfenidone (i.e. ICERs should be reduced) as pirfenidone reduces the rate of acute exacerbations vs. best supportive care.

Evidence that a long-term treatment effect with pirfenidone is plausible

Long-term follow-up data for pirfenidone are available through the three pivotal Phase III placebo-controlled randomised controlled trials CAPACITY 1, CAPACITY 2 and ASCEND, along with the RECAP extension study. Overall survival data for pirfenidone in these trials extends up until approximately 8 years, as demonstrated for the population of patients with FVC ≥ 50%-predicted. Figure 2 shows the comparison of the Kaplan-Meier plots with curve fits using the Weibull distribution, and assuming a lifetime duration of treatment effect. Included are the standard curve

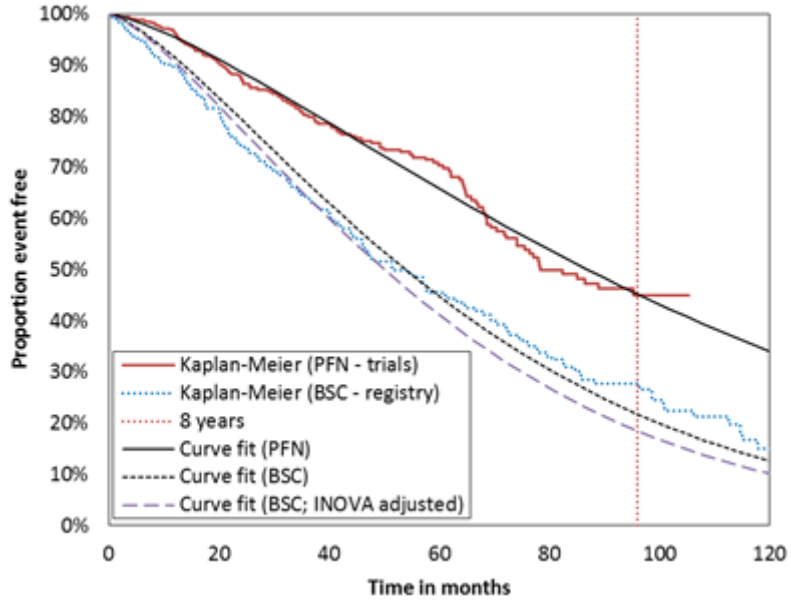
fits previously included in the submission, as well as a curve fit using a hazard ratio for pirfenidone vs. BSC based on the INOVA registry, after trimming data using a propensity score model to adjust for remaining imbalances. This was performed to address the comparability of the entire registry to patients within the context of available trials, and was described in Section 4.10 of the evidence submission. The figure shows that post-adjustment, short-term outcomes are more similar to the projected outcomes for BSC patients using the placebo arm of the clinical trials, and long-term outcomes are predicted to be slightly lower than the registry itself based upon the survival curve fit type assumed for pirfenidone (Weibull).

As previously discussed in Section 5.3 (page 210) of the evidence submission, registry data were used to validate long-term survival outcomes for BSC patients, as follow up for BSC patients within the context of the trials is limited to approximately two years. Of the 624 patients randomised to the BSC arms of the trials, only 54 (that is, 8.7% of all BSC patients) died, resulting in high levels of censoring and uncertain outcomes for BSC patients if trial data alone are used.

As discussed at the Appraisal Committee meeting, the registries themselves were not used to provide estimates of OS in the model base case, but rather served as reference material for long-term outcomes had trial data for BSC patients continued beyond 2 years.

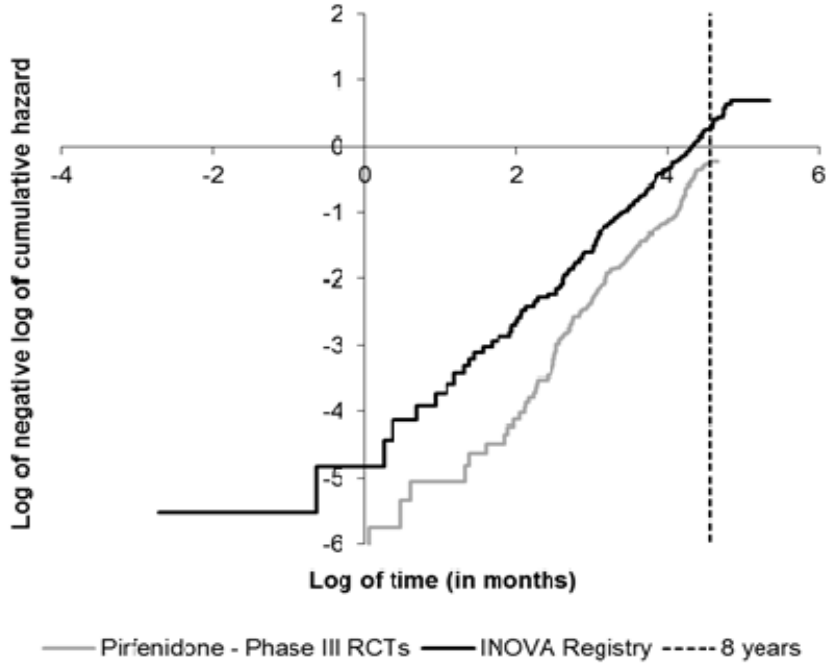
Within the ERG report, the ERG present optimistic (lifetime) and pessimistic (two years) assumptions regarding the duration of treatment effect for pirfenidone. The ERG based this on the availability of RCT data up until approximately two years. However, we do not believe the ERG have appropriately considered the availability of long-term OS data for patients treated with pirfenidone, which demonstrate an ongoing treatment effect up until the end of available pirfenidone data, as presented in Figure 2.

Figure 2: Comparison of Kaplan-Meier plot of overall survival for pirfenidone patients with hazard ratio for pirfenidone vs BSC based on INOVA registry (post trimming data using propensity score model to adjust for remaining imbalances)



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

Figure 3: Log-cumulative hazard plots for OS within the ASCEND/CAPACITY/RECAP trials for pirfenidone, and for the INOVA registry for BSC



When looking solely at within trial data from CAPACITY and ASCEND, as previously stated there was no significant interaction observed between treatment effect and time. This applies within the pooled dataset and individually within CAPACITY 2 and ASCEND. Furthermore, no convergence of curves is observed based upon the log-cumulative hazard plots for these trials (
Figure 3).

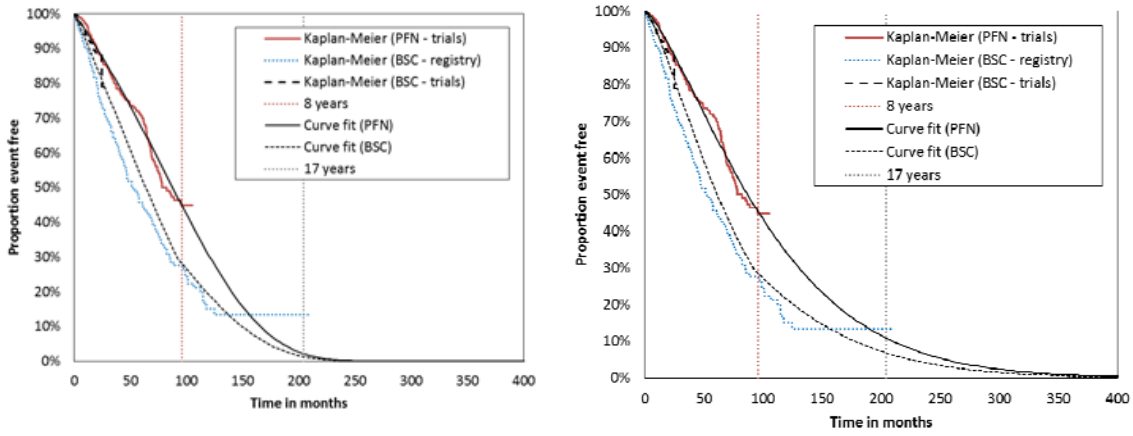
There is no clinical rationale why the treatment effect of pirfenidone would diminish in the long-term. Unlike treatments used in the treatment of cancers, which this Committee may frequently assess, within IPF there is no mechanism by which resistance to treatment may develop over time, and this is supported by expert clinical advice sought as part of this appraisal. Additionally there is no evidence of the treatment effect of pirfenidone diminishing (if one does not account for CAPACITY 1 where, as stated by the clinical expert at the Appraisal Committee meeting, the placebo arm performs differently to the placebo arms observed in all other trials in IPF). We also note that diminishment of treatment effect was not explored within the nintedanib appraisal, we assume due to the similar lack of a mechanism by which this would occur.

When all relevant datasets are considered at the end of follow up data, it can be seen that a pessimistic assumption regarding the duration of treatment effect may be considered as approximately eight years. An assessment of a combination of Figure 2 and
Figure 3 demonstrates evidence of a treatment effect of at least eight years, and this is supported by the totality of the evidence provided through the NMA, RECAP and INOVA registry data.

Based on these data, curve fits are presented in Figure 4, assuming a duration of treatment effect of eight years. Weibull (our preferred option) and Gompertz (ERG's preferred option) curve fits are shown. Economic analyses are also presented using this reduced treatment effect in order to

provide a more plausible pessimistic scenario for the Committee's assessment.

Figure 4: Kaplan-Meier plot of overall survival for pirfenidone patients (left to right; Weibull and Gompertz)



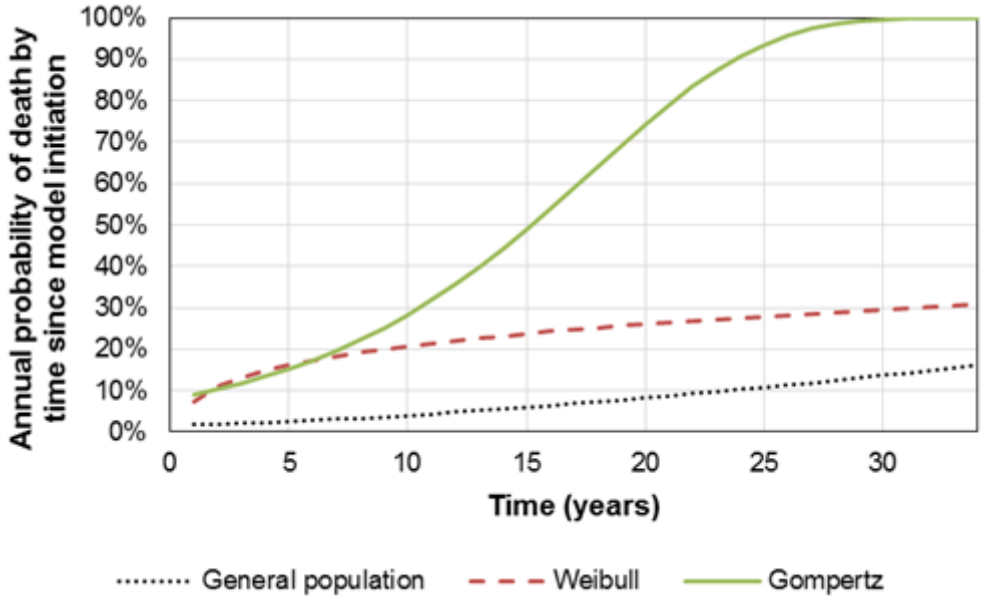
<u>Evidence that the survival curve selected by the Committee on the advice of the ERG to model long-term outcomes with pirfenidone is not supported by the available data</u>

As highlighted as part of our factual accuracy response to the ERG report, along with comments during the Appraisal Committee meeting, we strongly disagree with the use of the Gompertz distribution to model long-term outcomes, and do not believe this is supported by the available evidence. Furthermore, the ERG's justification for choice of distribution appears to be based on inaccurate analyses. Our rationale for this view, along with further support for the use of the Weibull distribution, is set out below:

(i) In-appropriateness of ERG analysis for use of the Gompertz curve

As highlighted by the manufacturer representatives during the Appraisal Committee meeting, we do not believe the evidence presented by the ERG in relation to the comparison of hazards for different curve fits and Office for National Statistics (ONS) general population data is appropriate. The figure presented by the ERG within their report (Figure 38) has been reproduced using the distribution of age from within the trials in Figure 5 below.

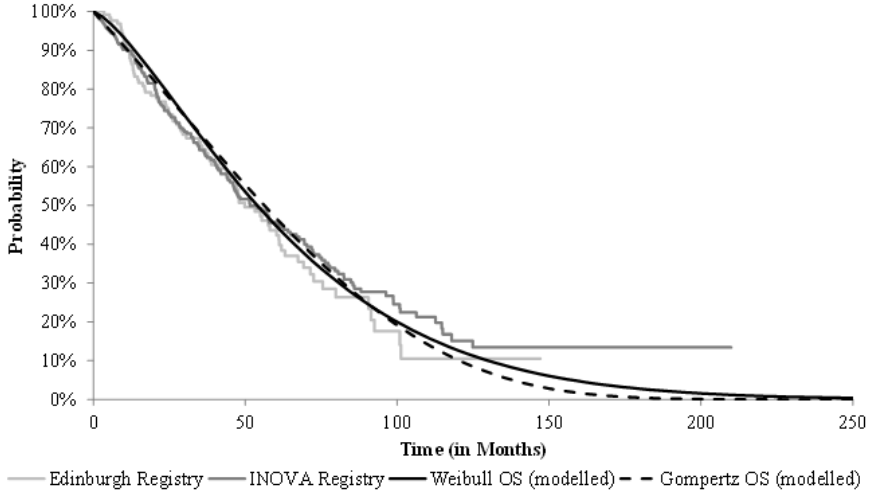
Figure 5: Revision of Figure 38 from ERG report: Plot of the annual hazard of death of modelled OS for BSC using the Weibull and Gompertz distribution and distribution-adjusted life tables in the UK in the ITT population (Revision of plot originally produced by the ERG)



This revised plot demonstrates that there is no issue relating to the Weibull distribution producing hazards which cross those of the general population within the model time horizon. The ERG commented that from Figure 38 of their report, it can be seen "*that the use of the Weibull distribution in the model leads in some occasions to lower probabilities of death in people with IPF initiating pirfenidone compared with the probability of death from the general population*."

We would consider this argument to be inappropriate as the earliest time at which the lines for the Weibull curve and the ERG-produced (i.e. non-distribution adjusted) general population cross is at approximately age 90 (i.e. 23 years or 276 months into the model time horizon). It would be expected that by this time, only a small proportion of patients would still be alive within the model. This assumption is verified by considering the figure presented by the ERG within their report for long-term OS extrapolations, re-presented in Figure 6.

Figure 6: Re-presentation of Figure 37 of ERG report: Plot of the KM for OS from registries and modelled survival for BSC using the Weibull and Gompertz distribution (Plot drawn by the ERG)



Key: BSC, best supportive care; ERG, Evidence Review Group; KM, Kaplan-Meier; OS, overall survival.

Furthermore, the ERG used general population data based on the mean age at treatment initiation, which does not consider the distribution of age at baseline. The difference in these curves are presented in Within the ERG report, it is stated that adjustment of registry data likely biases OS outcomes in favour of pirfenidone (Section 5.2.2.5 on page 175), suggesting that registry data used within the model represent a more severe group of patients than those in the CAPACITY and ASCEND trials. This statement is simply incompatible with the selection of a survival curve which predicts lower mortality than the registry. The Committee themselves spoke to a patient who has so far survived 10 years from diagnosis at the meeting. The Weibull curve takes into account the existence of this tail in its long-term extrapolation.

Figure 7, along with the distribution of age at baseline in distributional and cumulative form.

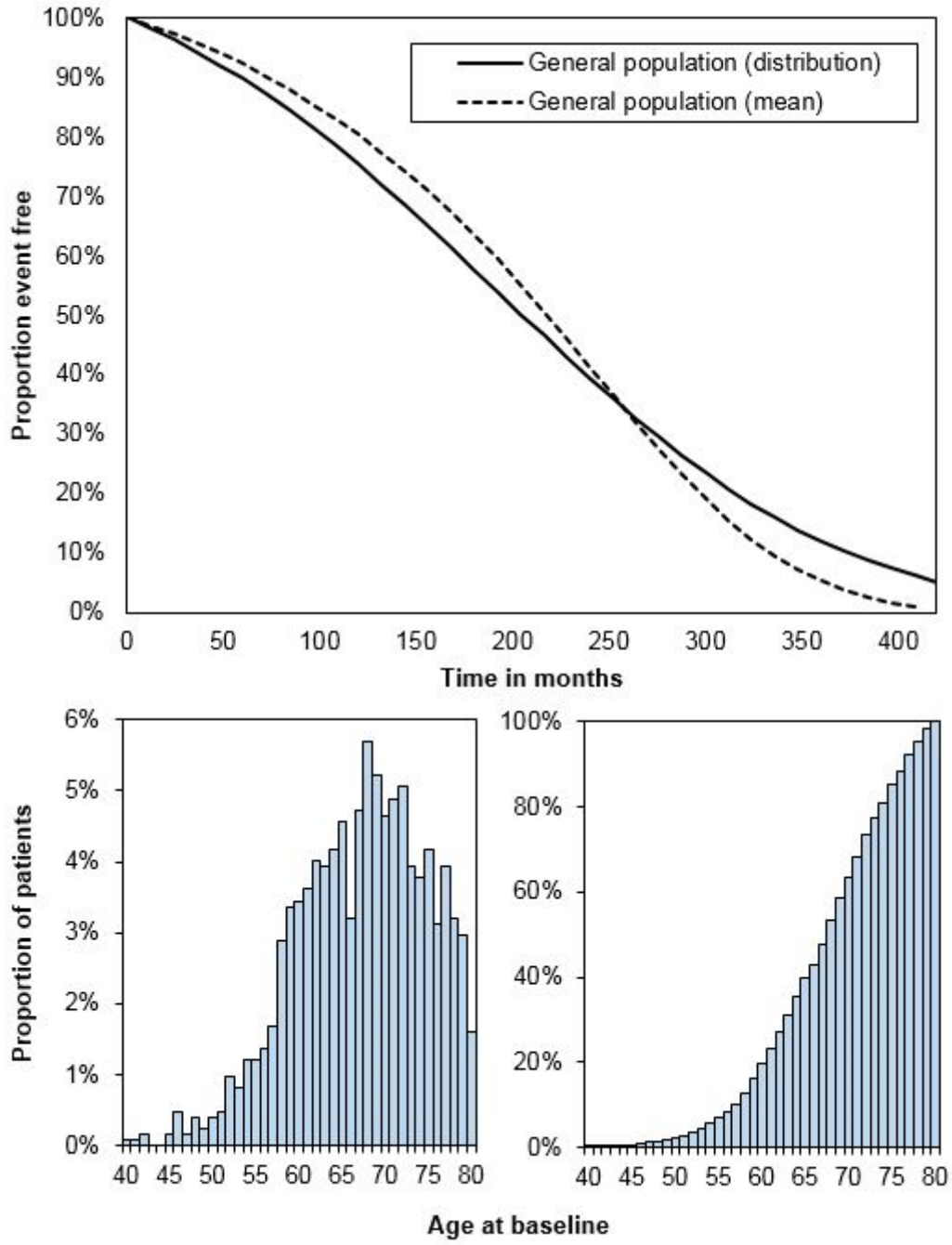
(ii) Lack of clinical plausibility of the Gompertz curve

The Gompertz curve fit appears to disregard the fact that IPF registries indicate a proportion of patients surviving for a substantial length of time post diagnosis. For example approximately 13%

of patients are still alive after 17 years in the INOVA registry.

Within the ERG report, it is stated that adjustment of registry data likely biases OS outcomes in favour of pirfenidone (Section 5.2.2.5 on page 175), suggesting that registry data used within the model represent a more severe group of patients than those in the CAPACITY and ASCEND trials. This statement is simply incompatible with the selection of a survival curve which predicts lower mortality than the registry. The Committee themselves spoke to a patient who has so far survived 10 years from diagnosis at the meeting. The Weibull curve takes into account the existence of this tail in its long-term extrapolation.

Figure 7: Difference in estimation of equivalent general population survival based on use of mean or distribution of age at baseline



(iii) Conclusion on most appropriate curve to use for survival analysis

We propose the Weibull distribution, with a duration of treatment effect of approximately eight years, provides a statistically good fit to the observed data (43.6 times more probable than the

Gompertz curve to be the best fitting curve based on AIC scores), good visual fit to observed data from both the trials and registry as well as realistic long-term extrapolation.

(iv) Alternative analysis for consideration

Whilst on balance we consider the Weibull curve be the best representation of long-term survival, based upon statistical goodness of fit and clinical plausibility, in order to provide the Committee with a more statistically robust estimation of long-term survival, an average curve based on weightings identified through AIC scores is presented to address the uncertainty in an individual curve fit choice. This method utilises the theory behind AIC to provide a weighted average curve based on the probability of the given curve providing the best fit to observed data. The weights are presented in Table 3.

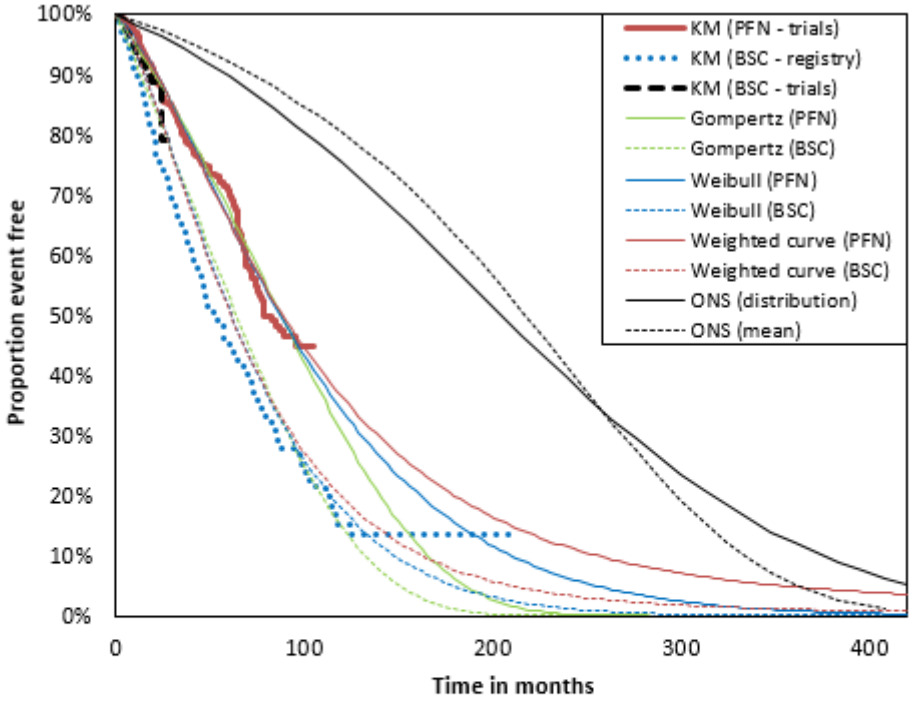
Table 3: Derivation of weighted average curve

Distribution	**OS AIC**	**Probability of best fit**	**Weight**
Curve fit	AIC_i	$P_i = e^{(Best\ AIC - AIC_i)/2}$	$\frac{P_i}{\sum_{j=1}^{n} P_j}$
Exponential	865.47	0%	0%
Weibull	844.15	100%	43%
Log-Normal	853.23	1%	0%
Gamma	845.78	44%	19%
Log-Logistic	844.54	82%	36%
Gompertz	851.70	2%	1%

Key: AIC, Akaike Information Criterion; OS, overall survival.

The next best statistical fits after Weibull were the Log-logistic and Gamma distributions, and consequently the weighted average curve produces survival estimates above that of the Weibull and Gompertz distributions alone, as shown in Figure 8. Therefore, it may be derived that the Weibull distribution provides a reasonable estimate between the lower bound provided by the Gompertz distribution and an upper bound provided by the weighted average curve. As can be seen, for both the Weibull and Weighted average curves, estimates remain below general population mortality when the age distribution of patients in the clinical trials is taken into account (Figure 8).

Figure 8: Comparison of overall survival estimates used within the model based on Weibull, Gompertz and Akaike Information Criterion weighted curves



Proportion of patients still alive					
Gompertz (PFN)	100.0%	42.9%	2.5%	0.0%	0.0%
Weibull (PFN)	100.0%	43.8%	11.4%	2.3%	0.4%
Weighted (PFN)	100.0%	45.0%	16.2%	7.0%	3.9%
Gompertz (BSC)	100.0%	25.6%	0.3%	0.0%	0.0%
Weibull (BSC)	100.0%	26.4%	3.0%	0.2%	0.0%
Weighted (BSC)	100.0%	27.6%	5.6%	1.8%	0.8%

Key: BSC, best supportive care; ERG, Evidence Review Group; KM, Kaplan-Meier; OS, overall survival; PFN, pirfenidone.

Results of revised economic analyses & presentation of new patient subgroup

Based upon the discussions above, revised analyses are presented for the Committee's consideration. These explore the uncertainty associated with curve fit, and the duration of treatment effect. New analyses are also presented for the subgroup of patients with FVC < 90%-predicted: in line with the discussion at the Appraisal Committee meeting, with represents the group of patients in whom the majority of clinical trial data for pirfenidone are available.

Incorporation of changes from the ERG report

The revised analyses presented below incorporate all changes highlighted in the ERG report, including the use of NMA results for OS at 72 weeks, but excluding the following features:

- The choice of curve for OS – Weibull, the weighted average curve and Gompertz curves are presented, in line with Committee request
- The ERG-produced CODA sample

The CODA sample produced by the ERG contained hazard ratios varying between 0.09 to 7.19 for OS and 0.00 to 30.96 for PFS. We do not consider the extremes demonstrated within these estimates to be clinically plausible. These estimates are drawn from CODA samples using estimates from the predictive distributions, which was the ERG's preference in their report but did not lead to consensus among experts at the 1st Appraisal Committee meeting (5 May 2016). In addition, the ICERs were calculated using means which are influenced by outliers (in contrast to medians).

With such a small network of trials, the predictive distributions will strongly depend on the choice of priors and guidance on choice of informative priors is limited (Dias, 2011; p39-40). Overall, we believe: the between-study heterogeneity is already taken into account by a random effects model; using predictive distributions is excessive, and; our original approach – where posterior medians and 95% credible intervals are used within economic analysis – is in line with previous submissions (NICE 2013a, NICE 2016b).

Analyses conducted

The updated model (available upon request) has been produced using probabilistic ICERs based on the CODA samples from the NMA conducted by Roche as part of the initial evidence

submission. Results are presented with incorporation of the pirfenidone PAS used within TA282: that is, a discount of [redacted] on the list price of pirfenidone.

As a condition of supplying new analyses, NICE requested that analyses for the patient subgroups presented within the initial submission are also re-run and supplied as part of this response to the ACD [email correspondence received from NICE on 14 June 2016]. Therefore, the following subgroups according to FVC %-predicted were explored within the revised economic analysis outlined above:

- ITT: Patients with FVC %-predicted ≥ 50
- Patients with FVC %-predicted ≥ 80
- Patients with 50 ≤ FVC %-predicted < 80
- New analysis: Patients with 50 ≤ FVC %-predicted < 90

The latter subgroup was explored to demonstrate results for the population of patients from which the majority of clinical data are available (since patients with FVC ≥ 90%-predicted were excluded from the ASCEND study). Further details regarding the consideration of this subgroup of patients are presented in the "Additional analysis in subgroup where most data are available (FVC< 90%-predicted)" section.

Scenario analyses are supplied for the Committee's consideration around the key aspects of uncertainty identified by the Committee (survival curve fit, treatment effect duration, stopping rule) in line with the ACD:

- As described above, analyses are performed for three different parameterisations of the OS curve: Weibull, Gompertz and weighted average distribution (see Table 3)
- The duration of treatment effect is explored, with the pessimistic scenario defined as a duration

of treatment effect of 8 years (as discussed in the section “

-
- Evidence that the survival curve selected by the Committee on the advice of the ERG to model long-term outcomes with pirfenidone is not supported by the available data”) and an optimistic scenario defined as a lifetime duration of treatment effect (as presented by the ERG in their report).
- Inclusion/exclusion of the stopping rule for pirfenidone is also considered. The Committee considered these scenarios to form bounds between which the true ICER is likely to lie, if the stopping rule were to be implemented in clinical practice).

In addition to the base case results presented in the tables below, plots of ICERs by duration of treatment effect were produced to assist the Committee in their decision making. These plots show at which point the duration of treatment effect is long enough in order for pirfenidone to appear cost-effective at a threshold of £30,000 per QALY gained.

Results – without stopping rule

Table 4, Table 6 and Table 8 contain results for an **eight-year duration** of treatment effect, with no stopping rule for pirfenidone using the weighted, Weibull and Gompertz curves, respectively. These results present the pessimistic assumption that duration of treatment effect is short, with no stopping rule (and therefore the cost of pirfenidone is applied for all patients regardless of FVC decline ahead of discontinuation). Consistent with earlier analyses, the Gompertz curve results in the least optimistic results (due to the inappropriately low survival curve chosen). ICERs for the ITT population (i.e. the population we would consider most relevant to this appraisal) were £27,565, £30,012 and £34,222 for the weighted, Weibull and Gompertz curves, respectively.

Table 5, Table 7 and Table 9 contain results for a **lifetime duration** of treatment effect, with no stopping rule for pirfenidone using the weighted, Weibull and Gompertz curves, respectively. These results present the slightly less pessimistic assumption that duration of treatment effect is long, but with no stopping rule applied: again, the Gompertz curve results in the least optimistic results (due to the inappropriately low survival curve chosen). ICERs for the ITT population (i.e. the population we would consider most relevant to this appraisal) were £23,544, £25,986 and £30,360 for the weighted, Weibull and Gompertz curves, respectively.

Figure 9, Figure 10 and Figure 11 present the durations of treatment benefit required for the ICER to fall below £30,000 / QALY, when modelling OS using weighted, Weibull and Gompertz parameterisations (respectively). For each presentation, the threshold duration of benefit required for each patient subgroup is shown. In each parameterisation, the required duration for the subgroup of patients with FVC ≥ 80%-predicted (> 11 years) is longer than all other populations (~8 years for the preferred Weibull distribution).

Table 4: Baseline model results: Weighted curve, 8 year Duration of treatment effect, no stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£66,670	9.79	5.66				
PFN	£34,053	7.47	4.48	£32,617	2.32	1.18	£27,565
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£81,739	12.31	6.79				
PFN	£40,259	9.98	5.65	£41,480	2.33	1.14	£36,292
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£63,459	9.41	5.48				
PFN	£33,241	7.19	4.33	£30,217	2.22	1.15	£26,372
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£65,210	9.39	5.49				
PFN	£33,182	7.15	4.34	£32,028	2.24	1.16	£27,685
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Table 5: Baseline model results: Weighted curve, Lifetime Duration of treatment effect, no stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£66,566	9.80	5.67				
PFN	£32,486	6.75	4.22	£34,080	3.05	1.45	£23,544
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£81,725	12.28	6.78				
PFN	£37,955	8.74	5.22	£43,769	3.54	1.56	£28,060
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£63,579	9.42	5.48				
PFN	£31,970	6.52	4.09	£31,609	2.90	1.39	£22,767
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£65,293	9.43	5.50				
PFN	£31,894	6.50	4.10	£33,399	2.92	1.40	£23,779
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Figure 9: Overview of duration of treatment effect impact on model ICER: Weighted curve, no stopping rule for pirfenidone

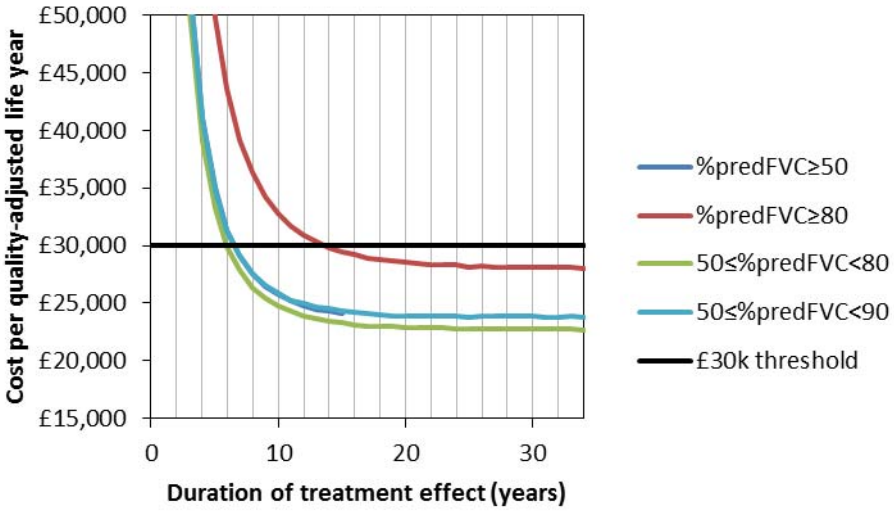


Table 6: Baseline model results: Weibull curve, 8 year Duration of treatment effect, no stopping rule for pirfenidone

	Total			Incremental			ICER
a) Population: FVC ≥ 50%-predicted (ITT)	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£64,916	8.73	5.32				
PFN	£32,804	6.77	4.25	£32,111	1.96	1.07	£30,012
b) Population: FVC %-predicted ≥ 80%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£79,778	11.27	6.48				
PFN	£38,771	9.21	5.41	£41,007	2.06	1.07	£38,474
c) Population: 50% ≤ FVC %-predicted < 80%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£59,775	7.68	4.83				
PFN	£30,358	5.88	3.81	£29,417	1.80	1.02	£28,884
d) Population: 50% ≤ FVC %-predicted < 90%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£63,363	8.31	5.14				
PFN	£31,806	6.45	4.10	£31,556	1.87	1.04	£30,432

Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone

Table 7: Baseline model results: Weibull curve, Lifetime Duration of treatment effect, no stopping rule for pirfenidone

	Total			Incremental			ICER
a) Population: FVC ≥ 50%-predicted (ITT)	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£64,817	8.72	5.32				
PFN	£31,464	6.22	4.04	£33,353	2.50	1.28	£25,986
b) Population: FVC %-predicted ≥ 80%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£80,169	11.28	6.48				
PFN	£37,016	8.17	5.04	£43,153	3.10	1.44	£29,874
c) Population: 50% ≤ FVC %-predicted < 80%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£59,827	7.69	4.83				
PFN	£29,535	5.53	3.66	£30,292	2.16	1.17	£25,979
d) Population: 50% ≤ FVC %-predicted < 90%	**Costs**	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£63,718	8.36	5.16				
PFN	£31,013	6.00	3.92	£32,706	2.36	1.24	£26,439

Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone

Figure 10: Overview of duration of treatment effect impact on model ICER: Weibull curve, no stopping rule for pirfenidone

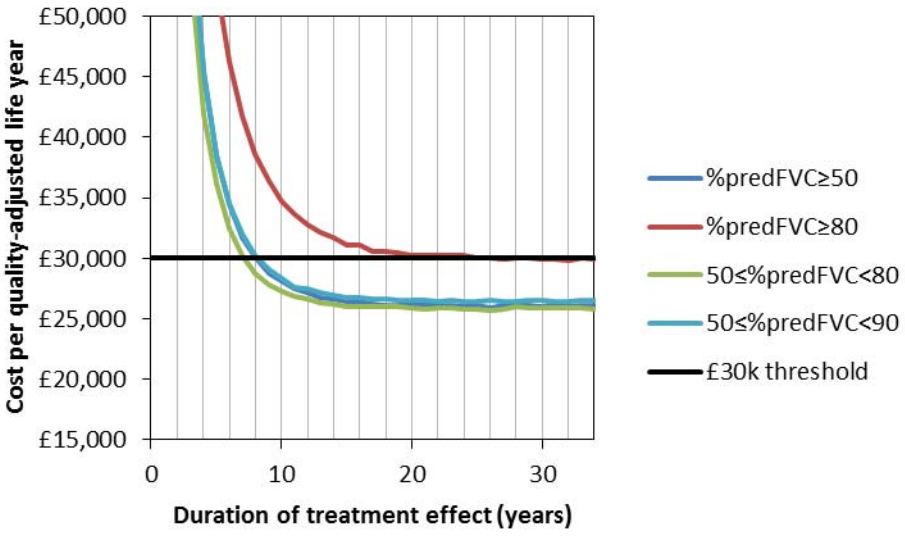


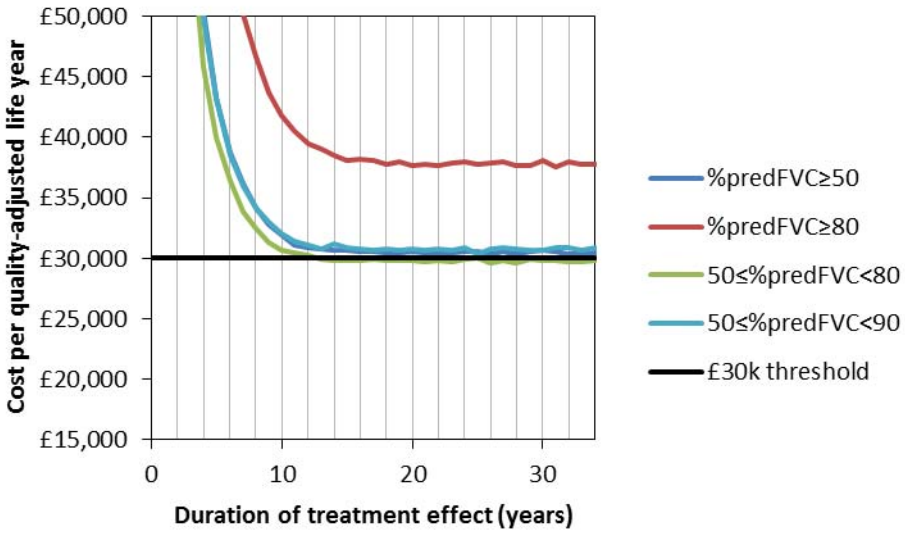
Table 8: Baseline model results: Gompertz curve, 8 year Duration of treatment effect, no stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£62,658	7.68	4.93				
PFN	£31,290	6.12	4.01	£31,368	1.55	0.92	£34,222
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£75,552	9.15	5.73				
PFN	£35,793	7.66	4.87	£39,758	1.49	0.86	£46,171
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£57,584	6.89	4.50				
PFN	£29,075	5.42	3.62	£28,509	1.47	0.88	£32,253
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£61,063	7.37	4.77				
PFN	£30,378	5.86	3.87	£30,685	1.51	0.90	£34,267
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Table 9: Baseline model results: Gompertz curve, Lifetime Duration of treatment effect, no stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£62,611	7.68	4.93				
PFN	£30,513	5.82	3.87	£32,097	1.86	1.06	£30,360
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£75,933	9.25	5.77				
PFN	£34,730	7.22	4.67	£41,203	2.03	1.10	£37,536
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£57,352	6.82	4.47				
PFN	£28,393	5.17	3.50	£28,959	1.65	0.97	£29,771
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£61,288	7.40	4.78				
PFN	£29,931	5.62	3.76	£31,357	1.78	1.02	£30,607
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Figure 11: Overview of duration of treatment effect impact on model ICER: Gompertz curve, no stopping rule for pirfenidone



Results – with stopping rule

Table 10, Table 12 and Table 14 contain results for an **eight-year duration** of treatment effect, with the stopping rule for pirfenidone using the weighted, Weibull and Gompertz curves, respectively. These results present the pessimistic assumption that duration of treatment effect is short, but the optimistic setting including the pirfenidone stopping rule. ICERs for the ITT population (i.e. the population we would consider most relevant to this appraisal) were £18,920, £20,587 and £23,237 for the weighted, Weibull and Gompertz curves, respectively.

Table 11, Table 13 and Table 15 contain results for a **lifetime duration** of treatment effect, with the stopping rule for pirfenidone using the weighted, Weibull x and Gompertz curves, respectively. These results present the optimistic assumption that duration of treatment effect is long, including the pirfenidone stopping rule. ICERs for the ITT population (i.e. the population we would consider most relevant to this appraisal) were £16,533, £18,167 and £21,002 for the weighted, Weibull and Gompertz curves, respectively.

Figure 12, Figure 13 and Figure 14 present the durations of treatment benefit required for the ICER to fall below £30,000 / QALY, when modelling OS using weighted, Weibull and Gompertz parameterisations (respectively). For each presentation, the threshold duration of benefit required for each patient subgroup is shown. In each parameterisation, the required duration for the subgroup of patients with FVC ≥ 80%-predicted (up to 8 years) is longer than all other populations (~4 years for the preferred Weibull distribution).

Table 10: Baseline model results: Weighted curve, 8 year Duration of treatment effect, with stopping rule for pirfenidone

e) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£56,370	9.80	5.66				
PFN	£33,978	7.49	4.48	£22,393	2.32	1.18	£18,920
f) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£66,359	12.30	6.79				
PFN	£40,237	9.98	5.65	£26,122	2.33	1.14	£22,862
g) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£54,369	9.39	5.47				
PFN	£33,247	7.18	4.33	£21,122	2.21	1.14	£18,509
h) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£55,294	9.43	5.50				
PFN	£33,240	7.17	4.34	£22,054	2.26	1.16	£18,943
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Table 11: Baseline model results: Weighted curve, Lifetime Duration of treatment effect, with stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£56,411	9.77	5.65				
PFN	£32,542	6.73	4.21	£23,869	3.04	1.44	£16,533
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£66,431	12.27	6.78				
PFN	£37,912	8.73	5.22	£28,520	3.54	1.56	£18,263
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£54,546	9.45	5.50				
PFN	£32,043	6.55	4.11	£22,504	2.90	1.39	£16,223
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£55,165	9.40	5.49				
PFN	£31,811	6.49	4.09	£23,354	2.90	1.40	£16,676
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Figure 12: Overview of duration of treatment effect impact on model ICER: Weighted curve, with stopping rule for pirfenidone

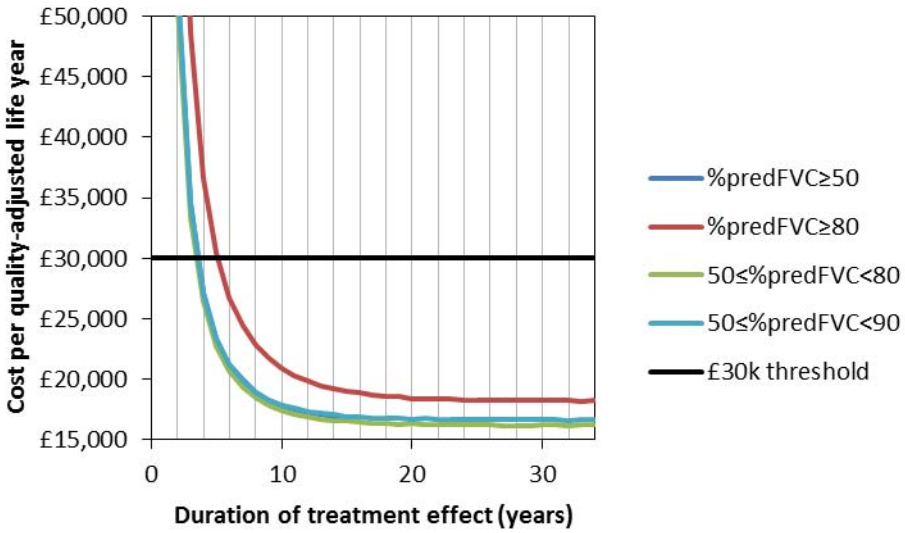


Table 12: Baseline model results: Weibull curve, 8 year Duration of treatment effect, with stopping rule for pirfenidone

e) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£54,646	8.73	5.32				
PFN	£32,627	6.78	4.26	£22,019	1.95	1.07	£20,587
f) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£64,942	11.35	6.51				
PFN	£39,023	9.28	5.44	£25,920	2.07	1.07	£24,295
g) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£51,034	7.68	4.83				
PFN	£30,258	5.88	3.81	£20,775	1.80	1.02	£20,411
h) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£53,487	8.35	5.16				
PFN	£31,857	6.47	4.11	£21,630	1.88	1.04	£20,738
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Table 13: Baseline model results: Weibull curve, Lifetime Duration of treatment effect, with stopping rule for pirfenidone

a) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£54,676	8.69	5.31				
PFN	£31,449	6.21	4.03	£23,227	2.48	1.28	£18,167
b) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£65,124	11.31	6.49				
PFN	£36,995	8.20	5.04	£28,129	3.12	1.45	£19,406
c) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£51,381	7.73	4.85				
PFN	£29,723	5.56	3.68	£21,659	2.17	1.17	£18,508
d) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£53,500	8.34	5.15				
PFN	£30,817	6.00	3.92	£22,683	2.34	1.23	£18,443
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Figure 13: Overview of duration of treatment effect impact on model ICER: Weibull curve, with stopping rule for pirfenidone

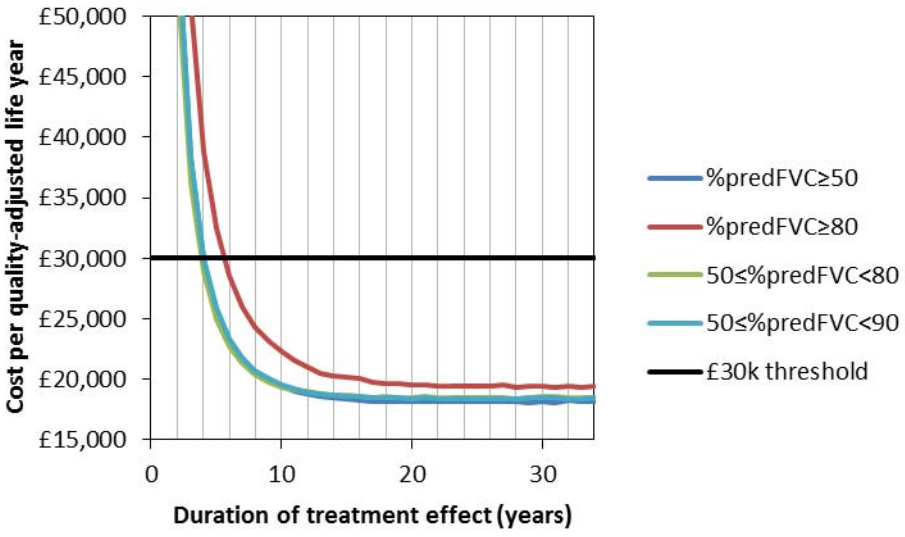


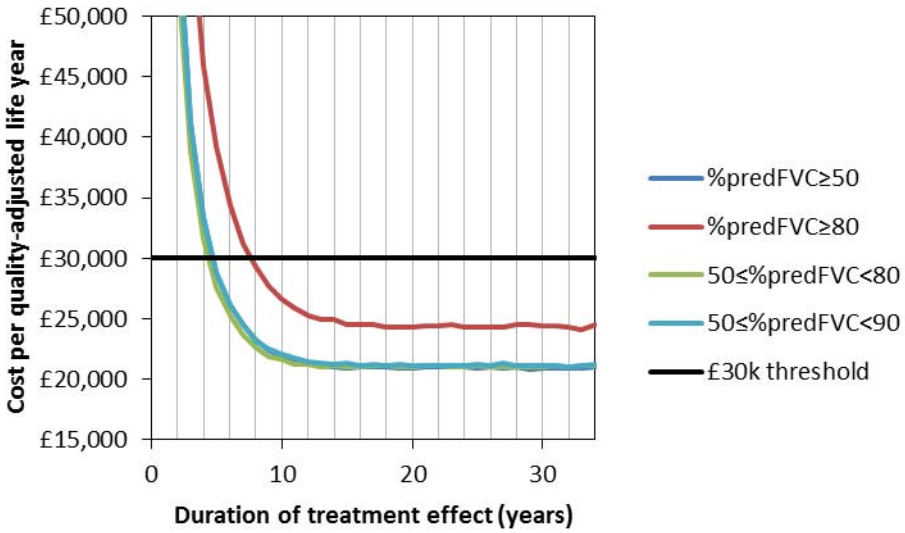
Table 14: Baseline model results: Gompertz curve, 8 year Duration of treatment effect, with stopping rule for pirfenidone

e) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£52,724	7.65	4.91				
PFN	£31,317	6.09	3.99	£21,407	1.56	0.92	£23,237
f) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£60,938	9.14	5.72				
PFN	£35,677	7.65	4.86	£25,261	1.49	0.86	£29,244
g) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£49,135	6.88	4.50				
PFN	£29,059	5.41	3.62	£20,076	1.47	0.89	£22,673
h) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£51,147	7.36	4.76				
PFN	£30,358	5.86	3.86	£20,790	1.51	0.90	£23,188
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Table 15: Baseline model results: Gompertz curve, Lifetime Duration of treatment effect, with stopping rule for pirfenidone

e) Population: FVC ≥ 50%-predicted (ITT)							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£52,578	7.66	4.91				
PFN	£30,449	5.80	3.86	£22,129	1.86	1.05	£21,002
f) Population: FVC %-predicted ≥ 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£60,791	9.12	5.71				
PFN	£34,323	7.14	4.63	£26,469	1.98	1.08	£24,494
g) Population: 50% ≤ FVC %-predicted < 80%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£49,108	6.86	4.49				
PFN	£28,530	5.20	3.51	£20,577	1.66	0.97	£21,119
h) Population: 50% ≤ FVC %-predicted < 90%							
	Total			**Incremental**			**ICER**
	Costs	**LYs**	**QALYs**	**Costs**	**LYs**	**QALYs**	
BSC	£51,293	7.37	4.76				
PFN	£29,724	5.60	3.75	£21,568	1.76	1.01	£21,267
Key: BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality-adjusted life year; PFN, pirfenidone							

Figure 14: Overview of duration of treatment effect impact on model ICER: Gompertz curve, with stopping rule for pirfenidone



References

Albera 2016 [AIC]. Efficacy of pirfenidone in patients with idiopathic pulmonary fibrosis with more preserved lung function. Albera et al, *ERJ* 2016, in press [AIC: unpublished]

Boehringer Ingelheim 2015. Boehringer Ingelheim Ltd. Single technology appraisal: Nintedanib for the treatment of adults with idiopathic pulmonary fibrosis: Company evidence submission. 2015.

Dias, S., et al., NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011, NICE Decision Support Unit (DSU): Sheffield, UK.

Ley 2012. Ley B, Ryerson CJ, E Vittinghoff. A Multidimensional Index and Staging System for Idiopathic Pulmonary Fibrosis. Ann Intern Med. 2012;156:684-691.

NICE 2013a. National Institute for Health and Care Excellence. Pirfenidone for treating idiopathic pulmonary fibrosis [TA282]. Published: 24 April 2013

NICE 2016a. National Institute for Health and Care Excellence. Pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282) [ID837]: Committee Papers

NICE 2016b. National Institute for Health and Care Excellence. Nintedanib for treating idiopathic pulmonary fibrosis [TA379]

Raghu 2015. Raghu G, Rochwerg B, Zhang Y et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med 2015; 192 (2): e3–e19

Wells 2003. Wells AU, Desai SR, MB Rubens et al. Idiopathic Pulmonary Fibrosis: A Composite Physiologic Index Derived from Disease Extent Observed by Computed Tomography. Am J Respir Crit Care Med 2003; 167: 962–969

White 2016. White ES, Borok Z, Brown KK et al. An American Thoracic Society Official Research Statement: Future Directions in Lung Fibrosis Research. 2016

British Thoracic Society

17 Doughty Street, London WC1N 2PL
T: +44 (0) 20 7831 8778 F: +44 (0) 20 7831 8766
bts@brit-thoracic.org.uk
www.brit-thoracic.org.uk
Registered as a charity in England and Wales No. 285174
Scottish Charity No. SC041209
Company Registration No. 1645201

To be submitted via NICE docs

June 2016

Dear Sir,

ACD - Consultees & Commentators: Idiopathic pulmonary fibrosis - pirfenidone (review of TA282) [837]

Thank you for inviting comments from the British Thoracic Society on the Appraisal Consultation Document (ACD).

Has all of the relevant evidence been taken into account?

The stopping rule: The Society is concerned that although change in FVC is considered to be a suitable endpoint in clinical trials, there is considerable intra-subject variability, which limits its usefulness in disease monitoring on an individual basis (Raghu G, Collard H *et al Am J Respir Crit Care Med* 2011)

Data from the recently published paper by Nathan SD *et al* (*Thorax* 2016; 71:429) showed that FVC change during the first 6 months of pirfenidone treatment was not predictive of change during the second interval. 59 patients had decline in FVC of 10% or more during the first 6 months of treatment. Importantly however, in 16 of these patients, FVC stabilised or improved with continued treatment for 6 months. Fewer patients in the pirfenidone treatment group compared to placebo group experienced a second >10% decline in FVC (2/34 in pirfenidone group compared to 19/68 in the placebo group). More patients in the treatment group had no further decline in FVC. This suggests that in IPF patient with progressive disease continued treatment with pirfenidone may reduce the risk of subsequent decline or death.

It is unknown what an individual's decline in FVC would be without treatment. Hence, >10% decline in FVC may be a treatment response depending upon the starting FVC, i.e. treatment changes the trajectory of FVC decline.

The effect of an exacerbation which may result in FVC declining by >10% should also be considered. For nintedanib, the effect of treatment is equivalent regardless of starting FVC.

Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?

We are uncertain if the committee considered the pooled mortality data from ASCEND and CAPACITY studies – pirfenidone reduced all-cause mortality and respiratory-related mortality after 1 year of treatment. This is likely to impact the cost effectiveness model.

Are the provisional recommendations sound and a suitable basis for guidance to the NHS?

The Society would support expansion of the current treatment guideline to include patients with FVC>80%. In many cases the starting FVC is unknown so FVC>80% often reflects clinically significant disease. It's important to note that FVC can be artificially elevated by the presence of co-existent emphysema, so this patient cohort is unfairly disadvantaged.

In addition there is concern about the accuracy of current lung function measurements and analysis in an older population. In some patients, change in TLco (DLco) is a more sensitive measure of disease progression rather than change in FVC.

Data from the BTS ILD registry shows that 40% patients have an FVC >80% , so this represents a substantial number of patients who are not eligible for treatment.

In the Netherlands, treatment guidelines are similar, but, for patients with FVC>80%, who have progressive disease (for example starting FVC 120% with documented evidence of progression radiologically and/or physiologically), their suitability for anti-fibrotic therapy is evaluated by an expert panel. This may be an option to consider; a defined yearly envelope of funding would ensure that there is clear attention given to which patients the expert panel "sign-off", rather than just everyone.

Yours faithfully,

[redacted]
[redacted]
British Thoracic Society

National Institute for Health and Care Excellence

Idiopathic pulmonary fibrosis - pirfenidone (review TA282) [ID837]

Royal College of Nursing

Introduction

The Royal College of Nursing (RCN) was invited to review the Appraisal Consultation Document (ACD) for Pirfenidone for the treatment of Idiopathic pulmonary fibrosis - (review TA282) [ID837]

Nurses caring for people with idiopathic pulmonary fibrosis were invited to review the documents on behalf of the RCN.

Appraisal Consultation Document – RCN Response

The Royal College of Nursing welcomes the opportunity to review this document. The reviewers' response to the questions on which comments were requested is set out below:

i) **Has the relevant evidence has been taken into account?**

The evidence considered seems comprehensive.

ii) **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?**

We would ask that the summaries of the clinical and cost effectiveness of this appraisal should be aligned to the clinical pathway followed by patients with idiopathic pulmonary fibrosis. The preliminary views on resource impact and implications should be in line with established standard clinical practice.

iii) **Are the provisional recommendations sound and a suitable basis for guidance to the NHS?**

Nurses working in this area of health have reviewed the recommendations of the Appraisal Committee and do not have any other comments to add.

The RCN would welcome guidance to the NHS on the use of this health technology.

Iv) **Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of gender, race, disability, age, sexual orientation, religion or belief?**

Given the robustness of the evaluation and taking opinions into consideration the draft recommendations are probably not unexpected although it is somewhat disappointing that the upper limit of forced vital capacity - FVC >80% remains because of the confounder of emphysema so these patients whist preserving their FVC actually can do badly.

Applying a FVC cut-off at 80% removes a significant number of patients that have IPF with some emphysema, they present with a preserved at best and or inflated FVC. As treatment with Pirfendione only slows progression, it does not seem sensible to wait for advanced disease before starting treatment. The role of FVC as an outcome measure for efficacy (there is reasonable evidence for this) but as a sole measure of severity, there is no evidence. There is much better evidence that the combination of DLCO and FVC provides a better measure of severity than FVC alone (Ley B. Ann Int Med 2012;156:684-691).

We agree with the stoppage of the medication if it is not affecting lung function decline.

v) **Are there any equality-related issues that need special consideration that are not covered in the appraisal consultation document?**

We are not aware of any specific issue at this stage. We would ask that any guidance issued should show that an equality impact analysis has been considered and that the guidance demonstrates an understanding of issues relating to all the protected characteristics where appropriate.

Lastly, we already have a drug which meets the >50% <80% criteria, we need to treat the milder patients earlier and raise the upper limit to at least 90%. There is a need for more drugs in the management of IPF, even if their efficacy is similar. Prognosis is poor for the patients, early treatment is paramount

To,
Meindert Boysen
Programme Director,
Centre for Health Technology Evaluation
Level 1A, City Tower, Piccadilly Plaza
Manchester, M1 4BT

24 June 2016

Dear Meindert,

Please find below our comments on the ACD on ID837: Idiopathic pulmonary fibrosis - pirfenidone (review of TA282).

Has all of the relevant evidence been taken into account?

Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease.

Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?

Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease.

Are the recommendations sound and a suitable basis for guidance to the NHS?

Please see our comments under question #4, re. treatment options for early stage (>80% predicted FVC) IPF patients for this progressive and poor-prognosis disease.

Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual orientation, age, gender reassignment, pregnancy and maternity?

ACD sections 4.5, 4.6 and 4.17 focus on the population of patients with percent predicted FVC of >80%, and calls them 'mild' patients. Although this terminology has been used in clinically and in NICE TA282 and TA379, it risks not conveying the seriousness of the condition in these patients. IPF is progressive and poor-prognosis disease with a median survival estimated at around 3y (range: 2-5y) across stages. Early stage (>70% predicted FVC patients) patients have a median survival of around 55 months (less than 5y). This is worse than many early, or even late, stage cancers. And it is progressive. Hence, just as no 'mild cancer' terminology exists, calling such a group of patients 'early stage IPF' rather than 'mild IPF' might be more appropriate.

At present, NICE does not recommend any active treatment for these early stage (nee 'mild') IPF patients; in effect, limiting patient and clinician choice of therapy in this group of early stage disease in a progressive poor-prognosis condition to 'best supportive care' (BSC, which is effectively no active treatment; especially as previous treatment options like N-Acetyl Cysteine have been shown to be potentially harmful to IPF patients). This is so even though active treatments, like nintedanib and pirfenidone, have been shown to be effective and is licenced in this population of early stage IPF

patients.

We understand that in this assessment the case for pirfenidone has not satisfied the appraisal committee for its clinical effectiveness or cost effectiveness versus BSC in this population.

We would, however, urge NICE and the committee to reconsider both the evidence for nintedanib and pirfenidone in this population of early stage IPF patients to provide an option for slowing progression of this poor-prognosis condition even in early stages of IPF instead of letting patients progress to a threshold (80% predicted FVC in this case) and then be able to be started on active treatment. This leaves an unmet need for early stage IPF patients despite two active treatments licensed in this indication. We would also suggest an assessment from an equality standpoint versus comparable poor-prognosis indications (e.g. some cancers).

Thanks for the opportunity to comment on this.

Best wishes,

[redacted]
[redacted] Boehringer Ingelheim Ltd
Ellesfield Avenue, Bracknell, Berkshire. RG12 8YS
Tel: [redacted] | Mobile: [redacted]

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

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Sarah Davis, Senior Lecturer in Health Economics, ScHARR, University of Sheffield, Sheffield, UK
	Rachid Rafia, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Jean Sanderson, Research Associate in statistics, ScHARR, University of Sheffield, Sheffield, UK
Correspondence Author	Sarah Davis, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
Date completed	Date completed (29/07/2016)

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

Declared competing interests of the authors

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

Rider on responsibility for report

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

This report should be referenced as follows:

Davis S, Rafia R, Sanderson J. Second addendum to Pirfenidone for treating idiopathic pulmonary fibrosis (review of TA282): A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2016.

Contributions of authors

Sarah Davis summarised and critiqued the clinical effectiveness data reported within the company's response to the ACD. Jean Sanderson critiqued the company's approach to meta-analysis and provided a revised ERG meta-analysis. Sarah Davis and Rachid Rafia critiqued and validated the revised health economic analysis submitted by the company and re-ran the revised company model using the ERG's revised meta-analysis. All authors were involved in drafting and commenting on the final report.

CONTENTS

TABLES

Abbreviations

ACD	Appraisal consultation document
AIC	Akaike Information Criterion
BSC	Best supportive care
CrI	Credible interval
CS	Company's submission
ERG	Evidence Review Group
FVC	Forced vital capacity
HR	Hazard ratios
ICER	Incremental cost-effectiveness ratio
IPF	Idiopathic pulmonary fibrosis
ITT	Intention to treat
KM	Kaplan-Meier
LY	Life-years
MA	Meta-analysis
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
OS	Overall survival
PAS	Patient access scheme
PFN	Pirfenidone
PFS	Progression-free survival
PrI	Predictive interval
QALY	Quality-adjusted life-year
RCT	Randomised controlled trial
SD	Standard deviation
6MWD	Six minute walking distance
UCSD SOBQ	University of California San Diego Shortness of Breath Questionnaire

1 Background

The company originally submitted clinical and economic evidence of the use of pirfenidone in people with a percent predicted forced vital capacity (FVC) ≥50% and in the subgroups of people considered to have moderate disease (percent predicted FVC of 50% to 80%) and mild disease (percent predicted FVC >80%).[1] The original CS included a patient access scheme (PAS) which was lower than the original PAS included in TA282. The original Evidence Review Group (ERG) report which details the ERG critique of the original submission is available on the National Institute for Health and Care Excellence (NICE) website.[2]

Following preliminary guidance in the appraisal consultation document (ACD),[3] the company submitted a response to the ACD,[4] which included a set of revised cost-effectiveness analyses.

This addendum to the original ERG report provides a critique of the company's submission (CS) in response to the ACD. The critique focuses on areas where new evidence or analyses were provided by the company. A point by point response to all issues raised in the CS in response to the ACD is not provided due to the restricted time frame available to the ERG to produce this addendum.

2 Clinical effectiveness evidence

2.1 Subgroup analysis based on percent predicted FVC above or below 80%

The company's response to the ACD states, *"no difference in the efficacy of pirfenidone by patient subgroup has been identified [..], and we believe the data from the trial populations should be assessed in their entirety."*[4] The company argues that there is no biologically plausible mechanism whereby either the efficacy of pirfenidone and/or the capacity to benefit would differ between patients with differing FVC levels and that the interaction test for a difference in treatment effect by subgroup was non-significant. The company states that information provided previously in their CS regarding which subgroup analyses were pre-specified and which were post-hoc was incorrect and corrected details are provided. No new analyses were provided by the company on the relative treatments effects of pirfenidone compared to best supportive care (BSC) for subgroups of patients defined by percent predicted FVC at baseline.

However, a pooled analysis of the ASCEND and CAPACITY trials is provided which examines whether disease progression differs for patients with more or less progressed disease who received placebo.[5] This analysis shows that patients with moderate disease (percent predicted FVC <80%) are at increased risk of experiencing functional decline, compared to patients with mild disease (percent predicted FVC ≥80%) as measured by a >50m decline in six minute walking distance (6MWD) or

death (hazard ratio [HR] 1.67, 95%CI 1.16-2.41, p=0.0049). Patients with moderate disease are also shown to be at increased risk of experiencing worsening of their breathlessness, as measured by a decline of more than 20 points in University of California San Diego Shortness of Breath Questionnaire (UCSD SOBQ) or death (HR 2.68, 95%CI 1.71-4.21, p<0.0001) over 12 months. However, patients with moderate disease were not found to be at increased risk of experiencing a ≥10% decline in percent predicted FVC or death (HR 1.28, 95%CI 0.85 – 1.92, p=0.2403) over 12 months (Albera 2016). The authors also note that clinically significant disease progression occurred in both subgroups.[5]

ERG comment

The ERG notes that when discussing the evidence for or against a difference in treatment effect for patients according to whether their percent predicted FVC was above or below 80%, the company's response to the ACD focuses on the evidence from 52 weeks. The ERG notes that although the results of the treatment-by-subgroup interaction tests were negative at both 52 weeks and 72 weeks, the inconsistency between the overall survival (OS) midpoint estimates for mild and moderate idiopathic pulmonary fibrosis (IPF) was greater at 72 weeks (0.59 vs 0.48 at 52 weeks and 0.90 vs 0.58 at 72 weeks).[6] In addition, the uncertainty around the HR for OS in the subgroup with percent predicted FVC >80% was much greater at 72 weeks. The results presented in the company's fact check for progression-free survival (PFS) and OS are reproduced in Table 1 below for reference.[6]

The ERG notes that in the cost-effectiveness analysis conducted by the ERG, an identical HRs was applied for the mild (percent predicted FVC of 50 – 80%) and moderate (percent predicted FVC >80%) subgroups and this HR was based on the network meta-analysis (NMA) of the intention to treat (ITT) populations of the ASCEND and CAPACITY trials.[2] Therefore, the approach taken by the ERG in their original report was consistent with the lack of statistical evidence to support a difference in clinical effectiveness and any difference in cost-effectiveness between the mild and moderate subgroups is not being driven by different estimates of relative treatment effect.

Table 1 PFS and OS results by baseline percent predicted FVC at 52 and 72 weeks [data reproduced from Tables 1 and 2 of the company fact check*]

Study & time point	Baseline FVC ≤80% predicted				Baseline FVC >80% predicted				Interaction test: p-value (likelihood ratio)
	n (PFN / pla)	Adjusted HR	95% CI	p-value	n (PFN / pla)	Adjusted HR	95% CI	p-value	
PFS from pooled trials									
52 weeks	472 / 450	0.62	0.52-0.78	<0.0001	146 / 168	0.54	0.35-0.75	0.0069	0.4656
72 weeks		0.64	0.52-0.79	<0.0001		0.53	0.35-0.79	0.0017	0.4106
OS from pooled trials									
52 weeks	477 / 454	0.48	0.27-0.83	0.0071	146 / 170	0.59	0.14-2.51	0.4682	0.6452
72 weeks		0.58	0.36-0.94	0.0240		0.90	0.27-2.99	0.8610	0.4728

*[4]

The difference in cost-effectiveness is instead being driven by differences in the expected rate of disease progression and treatment continuation for these subgroups, which is incorporated in the model via the Kaplan-Meier (KM) curves for PFS, OS and time to discontinuation in the pirfenidone arm. In the mild population, the hazard for progression is [redacted] and therefore patients spend [redacted] in the PFS health state. Patients in the mild subgroup also have [redacted] OS resulting in [redacted] for life-years (LYs) gained and quality-adjusted life-years (QALYs) gained. Based on the time to treatment discontinuation data, they also spend [redacted] receiving pirfenidone resulting in [redacted] life-time treatment costs.[1] However, the balance of additional cost and additional benefits is such that the cost-effectiveness is [redacted] favourable in the mild subgroup. So whilst the cost-effectiveness model does not assume any difference in relative efficacy, there is a difference assumed in prognosis and expected treatment duration for these two groups, based on what was observed in the ASCEND and CAPACITY trials, and this results in differing estimates of cost-effectiveness. The ERG notes that a difference in PFS for patients starting with mild and moderate disease is consistent with the analysis presented by Albera et al., as in the model PFS was defined as either a ≥10% fall in percent predicted FVC or a ≥50m fall in 6MWD and a significant difference was found in the latter for mild versus moderate IPF in a post-hoc analysis of the pooled ASCEND and CAPACITY trial.[5] Whilst the difference in progression based on lung function (≥10% decline in percent predicted FVC or death) was not statistically significant, there was a non-significant trend for FVC decline or death to be more common in patients with moderate disease.[5]

The ERG further notes that the methods guide states that subgroups may be identified based on "*an expectation of differential clinical or cost effectiveness because of known, biologically plausible mechanisms, social characteristics or other clearly justified factors.*"[7] The ERG accepts that it is difficult to provide biological plausibility for a subgroup analysis based on percent predicted FVC due to the fact that, "*IPF is a complex disease that is not yet fully understood, and research is hampered by the lack of a model which fully represents the human disease.*"[4] However, a subgroup analysis may be considered reasonable if there are differences in the cost-effectiveness estimates which are being driven by differences in the prognosis of patients with more or less severe disease prior to starting treatment. We therefore, believe that it is reasonable to consider separately the cost-effectiveness of pirfenidone in patients with percent predicted FVC ≥80% versus those with percent predicted FVC <80% as these were the subgroups examined by Albera et al.[5]

2.2 Subgroup analysis based on percent predicted FVC above or below 90%

In the original submission, the company presented subgroup analyses using a percent predicted FVC of 80% as the cut-off value. In the company's response to the ACD, they note that "*During the Appraisal Committee meeting, there was discussion that the majority of evidence from the ASCEND and CAPACITY trials is from patients with FVC < 90%-predicted"* and additional cost-effectiveness analyses are provided for the subgroup with percent predicted FVC ≥50% and < 90%.[4] The company states, *"there is no evidence of a treatment-interaction when the FVC < 90% subgroup is explored, and there is no clinical or biological rationale why treatment effect should differ in this subgroup"*.[4]

ERG comments

The company does not provide any analysis of clinical effectiveness for the population with percent predicted FVC ≥50% and <90%. Therefore, it is not possible for the ERG to assess the validity of the company's claim that *"there is no evidence of a treatment-interaction when the FVC < 90% subgroup is explored"*.

The ERG does not understand the company's rationale for presenting an analysis for the subgroup of patients with a percent predicted FVC ≤90%. The Committee noted in the ACD that patients with a percent predicted FVC >90% would be treated in clinical practice,[3] and therefore the ERG considers it reasonable to include evidence from this group within the cost-effectiveness analysis. The fact that this population were under-represented in the totality of the population enrolled in the pirfenidone clinical trials, primarily due to their exclusion from the ASCEND study, is not a valid rationale for providing a subgroup analysis excluding this group. It is also inconsistent with the company's argument that there is no difference in clinical efficacy for patients with different levels of percent predicted FVC.

2.3 Duration of treatment effect

The company provides a comparison of OS for pirfenidone versus BSC based on long-term (maximum of 8 years) follow-up data from patients enrolled in the ASCEND and CAPACITY trials for pirfenidone (including the RECAP extension study) and data from the INOVA registry for BSC.[4] The company provides a new parametric curve for BSC based on the HR obtained by comparing patients randomised to pirfenidone in the ASCEND / CAPACITY / RECAP trials against patients receiving BSC in the INOVA registry, after trimming the registry data using the inclusion / exclusion criteria from the pirfenidone randomised controlled trials (RCTs) and using a propensity score model to adjust for remaining imbalances. The company also presents log-cumulative hazard plots for the pirfenidone patients enrolled in the ASCEND / CAPACITY / RECAP trials and for the INOVA

registry and states that these demonstrate that there is no significant interaction between observed treatment effect and time.

ERG comments

The ERG considers that the comparison of data from the pirfenidone arms of the clinical trials and data on BSC from the INOVA registry is potentially subject to bias. This is because it is not based on a randomised comparison and there may be differences in unknown confounders between the groups. This is true even after restricting the registry data to patients meeting the trial inclusion criteria and using propensity score matching to estimate the HR as these cannot correct for an imbalance or unknown confounding variables.

The ERG also notes the concern it raised previously regarding the fact that the short-term data on OS in BSC from the pooled ASCEND / CAPACITY / RECAP datasets does not match the OS observed in the trimmed and propensity score matched registry dataset.[2] This suggests that these populations may not be comparable which may lead to bias in the estimation of relative efficacy obtained by comparing across these two datasets. Furthermore, the ERG notes that Figure 2 of the company response to the ACD shows that both the parametric curves fitted are under predicting OS from around 5 years in the BSC arm.[4] The ERG also notes that the 95% CI for the HR generated by comparing the pirfenidone trial arm to the INOVA registry data [redacted] does not contain the midpoint HR from the trials when using the 72 week data (0.64).[1] This inconsistency raises further doubt regarding whether the comparison with registry data is estimating a treatment effect that is consistent with the trial evidence. The company's assumption of a constant HR over 8 years is also inconsistent with the finding that the midpoint HR for OS from the meta-analysis (MA) of pirfenidone RCTs (see Figures 5 and 6 of the original ERG report[2]) was higher at 72 weeks than at 52 weeks (0.64 versus 0.52), although the ERG accepts that based on the data presented in the CS there was not sufficient evidence to reject the proportional hazards assumption.

The ERG concludes that there remains considerable uncertainty regarding whether the treatment effect observed in the trials persists up to 8 years or beyond.

2.4 Extrapolation of overall survival data

The CS in response to the ACD provides a new parametric survival curve estimated by weighting the six candidate parametric curves (exponential, Weibull, log-normal, gamma, log-logistic, Gompertz) fitted previously (see Figure 8 of the CS in response to the ACD).[4] The weighting of the curves is calculated based on their Akaike Information Criterion (AIC) values.

The CS in response to the ACD provides additional information on the expected all-cause mortality for the modelled cohort based on general population mortality when taking into account the age distribution observed within the trial population rather than assuming that all patients in the cohort have the mean age of the trial population (see Figures 5, 7 and 8 of the CS in response to the ACD).[4] The ERG notes that based on the company's calculations, the OS predicted for a general population cohort with matched ages remains above the OS prediction for the company's base-case parametric survival curve (the Weibull) for patients with IPF.

ERG comments

The ERG has several concerns regarding the calculation of the weighted curve for OS. Firstly all six candidate curves have been included in the weighted curve, whereas it is often desirable to choose a manageable set of models which are both parsimonious and supported by the data[8] This may be done using an arbitrary cut-off in AIC and Richards recommends a 6 point difference between the AIC for the candidate curve and the AIC for the optimal curve based on simulation studies.[9] In this case, this would eliminate the exponential, log-normal and Gompertz curves, although the impact on the weighted curve is likely to be minimal as these candidate curves have a very low weighting. An additional approach recommended by Richards is to remove overly complex curves by eliminating those which are more complicated versions of curves with a lower AIC.[9] In this case, this would eliminate the Gamma curve as the Weibull is a special case of the Gamma and has a lower AIC. Applying these methods would suggest that the Weibull and log-logistic curves would be retained in the weighted parametric curve based on the AIC. It is unclear what impact this would have on the weighted parametric curve and any cost-effectiveness analysis based on this curve.

Furthermore, the ERG notes that the methods used to select models and calculate the weight that should be placed on each is based on their fit to the observed data and ignores whether the long-term extrapolation predicted by each curve is plausible. In particular, the log-logistic curve and Weibull curves have very different predictions of long-term survival which are unlikely to have similar clinical plausibility but they have very similar predicted weightings. The ERG therefore considers that the weighted survival curve estimated by the company has limited credibility.

The calculation of OS for an age matched general population cohort addresses one of the ERG's concerns regarding the plausibility of the company's base-case (Weibull) OS curve. However, the ERG would highlight their previous comment in the ERG report that, "*patients treated in real-life scenarios differ from those treated in RCTs as real-life patients often have comorbidities, more severe*

disease, take concomitant medications and have a higher mortality,"[2] which would suggest that OS in real-life clinical settings may be lower than in clinical trial settings.

The ERG notes that the CS in response to the ACD states that the proportion alive at 17 years in the INOVA registry data (approximately 13%) is more consistent with the proportion predicted to be alive at that time for BSC based on the Weibull curve (3%) than the Gompertz curve (0.3%) (see page 25 of the CS in response to the ACD).[4] However, the ERG does not consider this to be a strong argument for rejecting the Gompertz curve as there appears to be considerable censoring of the data from the INOVA registry from 10 years. Figure 52 of the original CS, presents the number at risk at various time points and it can be seen that although the KM estimate of survival at 200 months (16.7 years) is around 13%, only 3% of the original cohort (8 of 234) are known to be alive.[1] The ERG notes that at 100 months (8.3 years), the Weibull and Gompertz curves predict a similar proportion surviving on BSC (26.4% versus 25.6% respectively) (see Figure 8 of the company response to the ACD[4]).

The ERG is therefore not convinced that the Weibull curve provides a more plausible extrapolation of OS than the Gompertz curve.

3 Revised cost-effectiveness analyses

The company presents revised cost-effectiveness analyses in their response to the ACD.[4] These are based on the ERG's adaptation of the company's model with the following changes:

1. no comparison against nintedanib is presented
2. the discount from the original PAS from TA282 is applied (discount of [redacted] on the list price for all doses)
3. treatment effect is varied from 8 years to lifetime instead of 2 years to lifetime (although graphical results for 2 years to lifetime are also provided)
4. an additional subgroup analysis is presented for patients with percent predicted FVC ≥50% and <90%
5. in addition to the ERG's preferred parametric curve (the Gompertz) the company presents results for two alternative methods for extrapolating OS; the Weibull, and a weighted parametric model.
6. the model is populated with CODA samples for the HR for OS at 72 weeks generated by the company using the posterior distribution from the company's original NMA (instead of the predictive distribution)

As the starting point for the company's revised analysis was the ERG's adaptation of the company model, all other amendments to the company's model included in the ERG base-case have been maintained.

ERG comments

The ERG has no comments to make on points 1 to 2. Points 3 and 4 have been covered above in the clinical effectiveness section (see section 2.2 and section 2.3). A discussion regarding the choice between parametric survival curves (point 5) has already been provided in section 2.4, but section 3.1 provides one additional point regarding the implementation of the weighted parametric curve within the economic analysis. The ERG's comments on the incorporation of uncertainty within the economic analysis using CODA samples (point 6) are provided in section 3.2.

Finally, the ERG notes that the company provides analyses both with and without the stopping rule and the ERG wishes to highlight their original conclusion that analyses which implement the stopping rule are likely to be biased in favour of pirfenidone for the reasons described in the original ERG report.[2]

3.1 Method used to estimate ICER for the weighted survival curve

The cost-effectiveness estimates for the weighted survival curve have been estimated by taking the average costs and QALYs for each of the six candidate curves and then taking the weighted average using the weights presented in Table 3 of the company response to the ACD.[4] The incremental cost-effectiveness ratio (ICER) presented for this scenario is the ratio of these weighted averages for costs and QALYs from these six candidate curves.

ERG comments

The ERG considers that a better method would have been to select one curve for each probabilistic sensitivity analysis run using the weights to determine the proportion of times each curve was selected. However, the direction and size of any bias from using the company's method is unclear.

3.2 Incorporation of uncertainty within the economic analysis using CODA samples

In its original submission to NICE,[1] the company used the median HR (for OS and PFS) estimated from the NMA at 52 weeks. The NMA was necessary in the model associated with the original submission to allow an indirect comparison between pirfenidone and nintedanib.

In the ACD it is stated "*The committee agreed that it preferred to use efficacy estimates from the ERG's network meta-analysis [NMA] because this included the 72-week follow-up data and used the predictive distribution*".[3] However, in the company's response to the ACD, the company states that

the CODA samples produced by the ERG contained HRs that varied in excess of the range that was clinically plausible.[4] The company states *"With such a small network of trials, the predictive distributions will strongly depend on the choice of priors and guidance on choice of informative priors is limited ... Overall, we believe: the between-study heterogeneity is already taken into account by a random effects model; using predictive distributions is excessive, and; our original approach – where posterior medians and 95% credible intervals [CrIs]are used within economic analysis – is in line with previous submissions."*[4]

In the revised economic model submitted in response to the ACD,[4] the company uses the CODA sample from the NMA for OS at 72 weeks presented in its original submission to NICE. The CODA samples were based on the posterior distribution instead of the predictive distribution. It should be noted that for PFS, the company did not use the CODA samples from the NMA but used instead the median HR at 72 weeks.

ERG comments

The ERG has a number of concerns with the approach and argument presented by the company.

- The ERG does not agree with the company that the use of predictive distributions is "excessive". In the presence of heterogeneity it is recommend that the predictive distribution of the treatment effect in a new study should be used, since the mean of the random effect distribution does not relate to any specific patient population and so is not meaningful. The predictive distribution better represents uncertainty about comparative effectiveness for a future rollout of a particular intervention.
- Furthermore, in the revised economic model submitted as part of its ACD response, nintedanib is no longer considered. Therefore, a full NMA including trials of nintedanib versus placebo is no longer needed.
- It is also unclear to the ERG why the CODA samples for OS have been used in the model but not the CODA samples for PFS as it would have been consistent to use the same methods for both model inputs.

4 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

4.1 Methods for revised meta-analyses

Although the ERG considers that the CODA samples used previously by the ERG were appropriate, further analyses were conducted since a comparison against nintedanib is no longer required and so a full NMA is no longer necessary. The analyses were therefore repeated, excluding the nintedanib trials (i.e. to provide a pairwise MA of the pirfenidone trials). Although the point estimates of treatment effect for pirfenidone vs BSC are expected to be very similar in the pairwise MA to those from the full NMA (containing nintedanib), the credible intervals (CrIs) and predictive intervals (PrIs) may be different.

SP3 was also included in the pairwise MA as the ACD stated that this was considered appropriate by the committee in order to maintain consistency with the approach taken in the nintedanib appraisal.[3]

For OS two analyses were conducted i) inverse gamma prior, as used in the companies base case and ii) half-normal prior as used in the ERGs base case.

As before, analyses were conducted in the freely available software package WinBUGS[10] and R[11], using the R2Winbugs[12] interface package. For all outcomes, a burn-in of 500,000 iterations of the Markov chain was used with a further 100,000 iterations retained to estimate parameters. Samples from the posterior distributions exhibited moderate correlation between successive iterations of the Markov chain so were thinned by retaining every 10th sample.

Results from the pairwise MA for OS and PFS are summarised in section 4.2.

4.2 Pairwise meta-analysis results

4.2.1 All-cause mortality at 72 weeks

Summary statistics the HR for OS at 72 weeks in the ERG's revised MA when using the inverse-Gamma and half-normal priors are provided in Table 2 and Table 3 respectively. It can be seen that the half-normal prior used by the ERG provides slightly a narrower CrI and PrI that that provided by the inverse-gamma prior used by the company.

Table 2 Summary statistics for HR for OS for pirfenidone versus placebo when using an inverse-Gamma prior for between study standard deviation (used by company)

	Mean	Sd	2.50%	25%	50%	75%	97.50%
hr[1,2]	0.638	0.191	0.356	0.517	0.616	0.73	1.042
hr.pred[1,2]	0.758	4.801	0.261	0.501	0.62	0.76	1.408
lhr[1,2]	-0.489	0.279	-1.032	-0.659	-0.484	-0.315	0.041
tau	0.235	0.32	0.027	0.07	0.141	0.283	0.98
deviance	9.946	2.262	7.299	8.374	9.356	10.86	15.82

Table 3 Summary statistics for HR for OS for pirfenidone versus placebo when using a half-normal prior for between study standard deviation (used by ERG)

	Mean	sd	2.50%	25%	50%	75%	97.50%
hr[1,2]	0.631	0.163	0.37	0.519	0.611	0.723	1.002
hr.pred[1,2]	0.649	0.274	0.293	0.496	0.609	0.748	1.222
lhr[1,2]	-0.492	0.252	-0.993	-0.656	-0.493	-0.325	0.002
tau	0.2	0.154	0.008	0.08	0.166	0.286	0.583
deviance	11.257	2.165	8.629	9.731	10.71	12.2	16.88

4.2.2 PFS at 72 weeks

Summary statistics for the HR for PFS in the ERG's revised MA are provided in Table 4.

Table 4 Summary statistics for HR for PFS for pirfenidone versus (used by ERG)

	Mean	sd	2.50%	25%	50%	75%	97.50%
hr[1,2]	0.632	0.073	0.502	0.588	0.627	0.672	0.783
hr.pred[1,2]	0.64	0.149	0.424	0.573	0.628	0.69	0.912
Tau	0.126	0.134	0.024	0.052	0.088	0.152	0.448
Deviance	-4.323	1.897	-6.716	-5.599	-4.762	-3.525	0.669

4.3 Cost-effectiveness results based on the revised ERG meta-analysis

The ERG re-ran the company's cost-effectiveness analyses using the CODA samples summarised in Table 3 and Table 4 which were generated using the ERG's original methods but having excluded the nintedanib trials from the network (as the comparison with nintedanib was no longer considered relevant following the ACD[3]) and including the SP3 trial. The ERG used the CODA samples from the predictive distribution of the treatment effects for OS and PFS in the cost-effectiveness analysis.

The ERG found that the ICERs were similar to those produced in the company's response to the ACD when comparing identical scenarios. Therefore the ERG is satisfied that the differences between the ERG's and the company's approach to incorporating uncertainty in the OS and PFS estimates does not make a material difference to the estimates of cost-effectiveness.

5 Overall conclusions

The ERG considers that the evidence to support an 8 years treatment effect is potentially biased and therefore there remains considerable uncertainty regarding whether the treatment effect observed in the clinical trials would be maintained up to 8 years or beyond.

The ERG considers that the Weibull and Gompertz curves for OS have more face validity than the curve based on the weighted average of the 6 curves fitted by the company. The ERG is not convinced that the Weibull curve provides a more plausible extrapolation of OS than the Gompertz curve.

The ERG does not believe that the choice of prior (half-normal or inverse-Gamma) for the NMA of OS has any material impact on the estimates of cost effectiveness. The ERG believes that there would be no significant change to the ICERs presented by the company in response to the ACD if the predictive distribution from a simple pairwise MA of the pirfenidone trials including SP3 was used instead of the estimates from the company's original NMA.

The ERG considers that the analyses which implemented the stopping rule are likely to be biased in favour of pirfenidone for the reasons described in the original ERG report.[2]

6 REFERENCES

1. Roche Product Limited. *Pirfenidone for the treatment of adults with idiopathic pulmonary fibrosis: Company evidence submission*. London, UK: The National Institute for Health and Care Excellence; 2016.
2. Davis S RR, Carroll C, Essat M, Sanderson J, Dracup N, Bianchi S, Thickett D. *Pirfenidone for treating idiopathic pulmonary fibrosis: A Single Technology Appraisal.* Sheffield: The University of Sheffield; 2016.
3. National Institute for Health and Care Excellence. *Idiopathic pulmonary fibrosis - Pirfenidone. Appraisal consultation document (ACD)*. London: National Institute for Health and Care Excellence,; May 2016.
4. Roche Product Limited. *Pirfenidone for the treatment of adults with idiopathic pulmonary fibrosis: ACD response*. London, UK: National Institute for Health and Care Excellence,; 2016.
5. Albera C, Costabel U, Fagan E, Glassberg M, Gorina E, Lancaster L, *et al.* Efficacy of pirfenidone in patients with idiopathic pulmonary fibrosis with more preserved lung function. *European Respiratory Journal* 2016; DOI: 10.1183/13993003.01966-2015:In press (AIC). http://dx.doi.org/DOI: 10.1183/13993003.01966-2015
6. Roche Product Limited. *Pirfenidone for the treatment of adults with idiopathic pulmonary fibrosis: Company Fact check*. London, UK: The National Institute for Health and Care Excellence; 2016.
7. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal*. London, UK: The National Institute for Health and Care Excellence; 2013.
8. Jackson CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc a Stat* 2009;**172**:383-404. http://dx.doi.org/10.1111/j.1467-985X.2008.00573.x
9. Richards SA. Dealing with overdispersed count data in applied ecology. *J Appl Ecol* 2008;**45**:218-27. http://dx.doi.org/10.1111/j.1365-2664.2007.01377.x
10. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat Comput* 2000;**10**:325-37. http://dx.doi.org/Doi 10.1023/A:1008929526011
11. R Developement Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing; 2013.
12. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *J Stat Softw* 2005;**12**:1-16. <Go to ISI>://WOS:000232806500001