NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Submitted by Boehringer Ingelheim Ltd

Nintedanib for the treatment of adults with idiopathic pulmonary fibrosis

Company evidence submission

May 2015

File name	Version	Contains confidential information	Date
Ofev NICE STA for submission.docx	V 1	Yes	1st May 2015

Instructions for companies

This is the template for submission of evidence to the National Institute for Health and Care Excellence (NICE) as part of the single technology appraisal (STA) process. Please note that the information requirements for submissions are summarised in this template; full details of the requirements for pharmaceuticals and devices are in the user guide.

This submission must not be longer than 250 pages, excluding appendices and the pages covered by this template.

Companies making evidence submissions to NICE should also refer to the NICE guide to the methods of technology appraisal and the NICE guide to the processes of technology appraisal.

Contents

1 Executive summary 14
1.1 Statement of decision problem 15
1.2 Description of the technology being appraised 17
1.3 Summary of the clinical effectiveness analysis 18
1.4 Summary of the cost-effectiveness analysis 19
2 The technology 23
2.1 Description of the technology 23
2.2 Marketing authorisation/CE marking and health technology assessment 23
2.3 Administration and costs of the technology 25
2.4 Changes in service provision and management 26
2.5 Innovation 27
3 Health condition and position of the technology in the treatment pathway 28
4 Clinical effectiveness 32
4.1 Identification and selection of relevant studies 32
4.2 List of relevant randomised controlled trials 39
4.3 Summary of methodology of the relevant randomised controlled trials . 42
4.4 Statistical analysis and definition of study groups in the relevant randomised controlled trials 50
4.5 Participant flow in the relevant randomised controlled trials 54
4.6 Quality assessment of the relevant randomised controlled trials 57
4.7 Clinical effectiveness results of the relevant randomised controlled trials 59
4.8 Subgroup analysis 65
4.9 Meta-analysis 66
4.10 Indirect and mixed treatment comparisons 66
4.11 Non-randomised and non-controlled evidence 141
4.12 Adverse reactions 142
4.13 Interpretation of clinical effectiveness and safety evidence 145
4.14 Ongoing studies 152
5 Cost effectiveness 153
5.1 Identification of studies 153

5.2 De novo analysis 158
5.3 Clinical parameters and variables 163
5.4 Measurement and valuation of health effects 189
5.5 Cost and healthcare resource use identification, measurement and valuation 225
5.6 Summary of base-case de novo analysis inputs and assumptions 245
5.7 Base-case results 260
5.8 Sensitivity analyses 278
5.9 Subgroup analysis 302
5.10 Validation 302
5.11 Interpretation and conclusions of economic evidence 303
6 Assessment of factors relevant to the NHS and other parties 305
References 310
7 Appendices 321

Tables and figures

Table 1: The decision problem 16
Table 2: Technology being appraised 17
Table 3: Incremental cost-effectiveness results without nintedanib PAS 21
Table 4: Incremental cost-effectiveness results with nintedanib PAS 21
Table 5: Costs of the technology being appraised 25
Table 6: Eligibility criteria used in the search for RCTs 35
Table 7: Records pertaining to the TOMORROW Phase II study 39
Table 8: Records pertaining to the INPULSIS Studies 40
Table 9: Excluded studies that contained nintedanib 40
Table 10: Relevant Randomised Controlled Studies 41
Table 11: Ongoing RCTs/Extension studies 41
Table 12: Summary of methodology of relevant RCTs 44
Table 13: Eligibility Criteria for TOMORROW and INPULSIS Studies 48
Table 14: Summary of Statistical Analyses in the RCTs 53
Table 15: Critical Appraisal of the TOMORROW and INPULSIS studies 58
Table 16: Characteristics of participants in the TOMORROW studies 60
Table 17: Characteristics of participants in the INPULSIS studies 61
Table 18: Summary of clinical outcomes from TOMORROW Study 63
Table 19: Summary of clinical outcomes for INPULSIS Studies 64
Table 20: Basic demographics of nintedanib trials and the UK population 65
Table 21: Study selection for NMA 67
Table 22: Study design of RCTs considered for inclusion in the NMA 70
Table 23: Patient characteristics of RCTs considered for inclusion in the NMA (1) 76
Table 24: Patient characteristics of RCTs considered for inclusion in the NMA (2) 78
Table 25: Efficacy outcomes of RCTs considered for inclusion in the NMA (1) 80
Table 26: Efficacy outcomes of RCTs considered for inclusion in the NMA (2) 83
Table 27: Safety outcomes of RCTs considered for inclusion in the NMA 85
Table 28: Bias risks in clinical trials 90
Table 29: Overall survival evidence 92
Table 30: Acute exacerbations evidence 93
Table 31: Loss of lung function evidence 94
Table 32: PFS evidence (Noble et al. [2011] definition) 95
Table 33: 6MWT evidence 97
Table 34: Serious cardiac events evidence 98
Table 35: Serious GI events evidence 99
Table 36: Discontinuation due to AEs evidence 99
Table 37: Overall discontinuation evidence 100
Table 38: Summary of evidence considered 102
Table 39: Description of the 8 endpoints considered for the NMA 103

Table 40: Potential treatment effect modifiers identified 104
Table 41: Population and disease characteristics for the nintedanib clinical trials 106
Table 42: Population and disease characteristics for the pirfenidone clinical trials 107
Table 43: Population and disease characteristics for the N-acetylcysteine clinical trials 108
Table 44: Study follow-up duration 113
Table 45: Scenarios considered for overall survival 115
Table 46: Nintedanib vs. placebo; scenario 1 OS 116
Table 47: Pirfenidone vs. placebo, scenario 1 OS 116
Table 48: N-acetylcysteine vs. placebo; scenario 1 OS 116
Table 49: NMA; scenario 1 OS 117
Table 50: Ranking matrix; scenario 1 OS 118
Table 51: Scenarios considered for acute exacerbations 118
Table 52: Nintedanib vs. placebo; scenario 1 acute exacerbations 119
Table 53: Pirfenidone vs. placebo; scenario 1 acute exacerbations 119
Table 54: N-acetylcysteine vs. placebo; scenario 1 acute exacerbations 119
Table 55: NMA; scenario 1 acute exacerbations 120
Table 56: Ranking matrix; scenario 1 acute exacerbations 121
Table 57: Scenarios considered for loss of lung function 121
Table 58: Nintedanib vs. placebo; scenario 1 loss of lung function 122
Table 59: Pirfenidone vs. placebo; scenario 1 loss of lung function 122
Table 60: N-acetylcysteine vs. placebo; scenario 1 loss of lung function 122
Table 61: NMA; scenario 1 loss of lung function 123
Table 62: Ranking matrix; scenario 1 loss of lung function 124
Table 63: NDB vs. PBO; 6MWT change from baseline 125
Table 64: PFN vs. PBO; 6MWT change from baseline 125
Table 65: NAC vs. PBO; 6MWT change from baseline 125
Table 66: NMA; 6MWT change from baseline 125
Table 67: Ranking matrix; 6MWT change from baseline (fixed effects) 126
Table 68: Scenarios considered for serious cardiac events 126
Table 69: Nintedanib vs. placebo; scenario 1 serious cardiac events 127
Table 70: Pirfenidone vs. placebo; scenario 1 serious cardiac events 127
Table 71: N-acetylcysteine vs. placebo; scenario 1 serious cardiac events 127
Table 72: NMA; scenario 1 serious cardiac events 128
Table 73: Ranking matrix; scenario 1 serious cardiac events 129
Table 74: Scenarios considered for serious GI events 129
Table 75: Nintedanib vs. placebo; scenario 1 serious GI events 130
Table 76: Pirfenidone vs. placebo; scenario 1 serious GI events 130
Table 77. N-acetylcysteine vs. placebo; scenario 1 serious GI events 130
Table 78: NMA; scenario 1 serious GI events 130
Table 79: Ranking matrix; scenario 1 serious GI events 132

Table 80: Scenarios considered for discontinuation due to AEs 132
Table 81: Nintedanib vs. placebo; scenario 1 discontinuation due to AEs 133
Table 82: Pirfenidone vs. placebo; scenario 1 discontinuation due to AEs 133
Table 83: N-acetylcysteine vs. placebo; scenario 1 discontinuation due to AEs 134
Table 84: NMA; scenario 1 discontinuation due to AEs 134
Table 85: Ranking matrix; scenario 1 discontinuation due to AEs 136
Table 86: Scenarios considered for overall discontinuation 136
Table 87: Nintedanib vs. placebo; scenario 1 discontinuation overall 137
Table 88: Pirfenidone vs. placebo; scenario 1 discontinuation overall 137
Table 89: N-acetylcysteine vs. placebo; scenario 1 discontinuation overall 137
Table 90: NMA; scenario 1 discontinuation overall 137
Table 91: Ranking matrix; scenario 1 discontinuation overall 139
Table 92: Adverse Events of the TOMORROW Study 143
Table 93: Adverse Events of the INPULSIS Studies 144
Table 94: Eligibility criteria for economic evaluations 154
Table 95: Summary of published cost-effectiveness studies identified 157
Table 96: Model transitions and inputs 162
Table 97: Features of the *de novo* analysis 162
Table 98: Goodness of fit - overall survival 165
Table 99: Coefficients of parametric models - overall survival 166
Table 100: Variance-covariance matrix - overall survival 166
Table 101: Baseline population characteristics across trials 168
Table 102: OR values for OS - NMA results 170
Table 103: Goodness of fit: time-to-first-acute-exacerbation (investigator reported) 172
Table 104: Coefficient for exponential model - exacerbation (investigator reported) 172
Table 105: Goodness of fit: time-to-first-acute-exacerbation (adjudication committee) 172
Table 106: Coefficients for exponential model - exacerbation (adjudication committee) 172
Table 107: OR values for exacerbations - NMA Results 174
Table 108: Patient distribution at the start of the model 174
Table 109: Coefficients for lung function decline (without exacerbation) 176
Table 110: Variance-covariance matrix for lung function decline (without exacerbation) 176
Table 111: 3-month probability of progression (model without exacerbation, placebo arm) 176
Table 112: Coefficients for lung function decline (with exacerbation) 177
Table 113: Variance-covariance matrix for lung function decline (with exacerbation) 177
Table 114: 3-month probabilities of progression (model with exacerbation, placebo arm) 178
Table 115: OR values for loss of lung function - NMA results 179
Table 116: Incidence and risk of SAEs in the PBO arm 181
Table 117: OR values for SAEs - NMA results 181
Table 118: Incidence and risk of clinically important events for the nintedanib arm 182
Table 119: Incidence and risk of clinically important events for the pirfenidone arm 182

Table 120: AIC values for the current UK model (NMA) 183
Table 121: Coefficient for exponential model - overall discontinuation, placebo arm 183
Table 122: OR values for overall discontinuation - NMA results 183
Table 123: Loglogistic model for OS 184
Table 124: Weibull model for OS 184
Table 125: STATA survival analysis output – time to acute exacerbation 186
Table 126: Base-case logistic model for loss of lung function (without exacerbations) 187
Table 127: Logistic model for loss of lung function (with exacerbations considered) 187
Table 128: STATA output for risk of discontinuation 187
Table 129: EQ-5D utility values by FVC%pred group 190
Table 130: Exacerbation-related disutility 192
Table 131: Eligibility criteria for the HRQL search 194
Table 132: Summary of retrieved records pertaining to the HRQL search 197
Table 133: Summary of disutilities due to AEs 224
Table 134: Eligibility criteria for unit costs and resource use search 226
Table 135: Adverse event-related costs 229
Table 136: Concomitant medication - incidence (>10%) 230
Table 137: Concomitant medications - probabilities per cycle 230
Table 138: Most prevalent drug for each concomitant therapy class 230
Table 139: Concomitant therapy cost per cycle 231
Table 140: Number of HCRU observations in each FVC%pred group 232
Table 141: Cost inputs for hospitalisation 233
Table 142: 3-month probabilities of ER visits 235
Table 143: 3-month probabilities of GP visit 236
Table 144: 3-month probabilities of specialist visits 236
Table 145: 3-month probabilities of nurse visits 237
Table 146: 3-month probabilities of physiotherapist visits 238
Table 147: 3-month probabilities of other visits 238
Table 148: 3-month probabilities of chest HRCT 239
Table 149: 3-month probabilities of chest X-ray 239
Table 150: 3-month probabilities of O_2 requirement assessment 240
Table 151: 3-month probabilities of bronchoalveolar lavage 240
Table 152: 3-month probabilities of CT pulmonary angiograms 241
Table 153: 3-month probabilities of right heart catheterisation 241
Table 154: 3-month probabilities of general diagnostic 242
Table 155: Total cost per FVC%pred group 242
Table 156: HCRU services excluded from the analysis 243
Table 157: 3-month probabilities of exacerbation costs 244
Table 158: Exacerbation costs per item and total 244
Table 159: Summary of variables applied in the economic model 246

Table 160: Key model assumptions 254
Table 161: Deterministic results for nintedanib vs. pirfenidone 261
Table 162: Deterministic results for NDB vs. PFN 262
Table 163: Deterministic results for nintedanib vs. BSC 263
Table 164: Deterministic results for NDB vs. BSC 265
Table 165: Base-case results at nintedanib list price 265
Table 166: Base-case results at nintedanib PAS price 267
Table 167: Proportion of patients alive at the end of the first year 270
Table 168: Proportion of patients alive with exacerbations at the end of the first year 271
Table 169. Proportion of patients alive and in each FVC%Pred category at the end of one year: BSC arm 272
Table 170. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB arm 273
Table 171. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB and PFN 274
Table 172: Model outputs by clinical outcomes for BSC arm 277
Table 173: Model outputs by clinical outcomes for nintedanib arm 277
Table 174: Model outputs by clinical outcomes for pirfenidone arm 277
Table 175: Probabilistic sensitivity analysis - parameter inputs 279
Table 176: PSA results for nintedanib vs. pirfenidone 281
Table 177: QALY results for nintedanib and pirfenidone 282
Table 178: PSA results for nintedanib vs. best supportive care 283
Table 179: OWSA scenarios testing parameters between the limits of their 95% CI 286
Table 180: OWSA Scenarios for OS 286
Table 181: OWSA scenarios for acute exacerbations 287
Table 182: OWSA scenarios for loss of lung function 287
Table 183: OWSA Scenarios for safety 287
Table 184: OWSA scenarios for discontinuation 287
Table 185: OWSA for FVC%pred categories 288
Table 186: Differences between the ASCEND-like population model and the main model 289
Table 187: Sensitivity analysis results (nintedanib vs. pirfenidone) 291
Table 188: Deterministic results for nintedanib vs. pirfenidone 293
Table 189: Impact of pirfenidone discount rate on the ICER, nintedanib at list price 294
Table 190: Comparison of average survival time (nintedanib arm) 298
Table 191: Sensitivity analysis results (nintedanib vs. BSC) 299
Table 192: Estimated number of patients eligible for nintedanib treatment 305
Table 193: Existing treatment scenario (without nintedanib) 306
Table 194: New treatment scenario (with nintedanib in the expected mix of treatments) 306
Table 195: Costs of drug and intervention 307
Table 196: Summary of budget impact (based on list price) 308

Table 197: Summary of budget impact (based on discounted price under PAS) 308

List of figures

Figure 1: Model structure 19
[redacted] 22
Figure 3: Scatter plot: incremental cost-effectiveness analysis NDB vs BSC and NDB vs PFN (with nintedanib PAS) ..... **Error! Bookmark not defined.**
Figure 4: Current clinical pathway of care for idiopathic pulmonary fibrosis 29
Figure 5: Study flow diagram for RCT study selection (PRISMA diagram) 38
Figure 6: INPULSIS Study Design 43
Figure 7: Hierarchy of endpoints tested 51
Figure 8: Patient Flow diagram for the TOMORROW Study 54
Figure 9: Patient flow for INPULSIS-1 55
Figure 10: Patient flow for INPULSIS-2 56
Figure 11: Study flow diagram for RCT study selection to be used in NMA (PRISMA diagram) 69
Figure 12: Network for OS 91
Figure 13: Network for acute exacerbations 92
Figure 14: Network for loss of lung function 94
Figure 15: Pairwise comparison for PFS 95
Figure 16: Network for 6MWT 97
Figure 17: Network for serious cardiac events 98
Figure 18: Network for serious GI 98
Figure 19: Network for discontinuation due to AEs 99
Figure 20: Network for overall discontinuation 100
Figure 21: Forest plot; scenario 1 OS fixed effects model 117
Figure 22: Forest plot; scenario 1 OS random effects model 118
Figure 23: Forest plot; scenario 1 acute exacerbations fixed effects 120
Figure 24: Forest plot; scenario 1 acute exacerbations random effects 121
Figure 25: Forest plot; scenario 1 loss of lung function fixed effects 123
Figure 26: Forest plot; scenario 1 loss of lung function random effects 124
Figure 27: Forest plot; change from baseline 6MWT fixed effects 126
Figure 28: Forest plot; scenario 1 serious cardiac events fixed effects 128
Figure 29: Forest plot; scenario 1 serious cardiac events random effects 129
Figure 30: Forest plot; scenario 1 serious GI events fixed effects 131
Figure 31: Forest plot; scenario 1 serious GI events random effects 132
Figure 32: Forest plot; scenario 1 discontinuation due to AEs fixed effects 135
Figure 33: Forest plot; scenario 1 discontinuation due to AEs random effects 135
Figure 34. Forest plot; scenario 1 overall discontinuation fixed effects 138
Figure 35: Forest plot; scenario 1 overall discontinuation random effects 138
Figure 36: Study flow diagram for economic evaluation study selection (PRISMA diagram) 156
Figure 37: Model structure 160

Figure 38: Model structure - overall survival 165
Figure 39: Overall survival - 10 year extrapolation 166
Figure 40: OS model fit vs. Kaplan-Meier for placebo 167
Figure 41: Overall survival extrapolations for the placebo arm compared against Kondoh et al 2010169
Figure 42: Overall survival loglogistic extrapolations for the active arms compared against Kondoh et al 170
Figure 43: Model structure - exacerbations 171
Figure 44: Exacerbation model fit vs. clinical trial Kaplan-Meier 173
Figure 45: Model structure - loss of lung function 175
Figure 46: Progression probabilities (model without exacerbation, placebo arm) 177
Figure 47: Progression probabilities (model with exacerbation) 178
Figure 48: Model synthesis and clinical trial results for the BSC arm (patients per health state at one year) 179
Figure 49: Model synthesis and clinical trial results for the nintedanib arm (patients per health state at one year) 180
Figure 50: Baseline utility values 191
Figure 51: Study flow diagram for HRQL study selection (PRISMA diagram) 195
Figure 52: Study flow diagram for resource use study selection (PRISMA diagram) 227
Figure 53: Diagram of HCRU analysis 233
Figure 54: Cost synthesis: hospitalisation 234
Figure 55: Hospitalisation cost per FVC%pred group 235
Figure 56: Total annual costs for each FVC%pred group 243
[redacted] 261
Figure 58: Base-case total costs (nintedanib and pirfenidone) ..... **Error! Bookmark not defined.**
[redacted] 263
Figure 60: Base-case total costs (nintedanib and BSC) ..... **Error! Bookmark not defined.**
[redacted] 266
Figure 62: Scatter plot: incremental cost-effectiveness analysis NDB vs BSC and NDB vs PFN (with nintedanib PAS) 267
Figure 63. OS models fit vs. clinical trial KM – PBO arm 268
Figure 64. OS models fit vs. clinical trial KM – NDB arm 269
Figure 65: OS models fit vs. clinical trial KM - PFN arm 270
Figure 66. Proportion of patients alive and in each FVC%Pred category at the end of one year: BSC arm 272
Figure 67. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB arm 273
Figure 68. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB and PFN arms (model projections) 274
Figure 69: Proportion in each state in nintedanib arm 275
Figure 70: Proportion in each state in pirfenidone arm 275

Figure 71: Proportion in each state in best supportive care arm 276
[redacted] 282
Figure 73: CEAC (nintedanib vs. pirfenidone) 283
[redacted] 284
Figure 75: CEAC (nintedanib vs. BSC) 284
[redacted] 285
[redacted] 285
[redacted] 294
Figure 79: Comparison of nintedanib survival for the loglogistic model 297
Figure 80: Comparison of nintedanib survival for the Weibull model 298
Figure 81: Comparison of nintedanib survival for the Gompertz model 298
[redacted]: Tornado[redacted] 301

1 Executive summary

Nintedanib, marketed as Ofev, is a small molecule tyrosine kinase inhibitor (TKI) which specifically targets three growth factor receptors: the vascular endothelial growth factor receptor (VEGFR), fibroblast growth factor receptor (FGFR) and platelet-derived growth factor receptor (PDGFR) (Hilberg *et al.*, 2008[2], Wollin *et al.*, 2014[3]). It is believed that these receptors are involved in the process of pulmonary fibrosis and blocking these receptors has the potential to slow the decline in lung function (Hilberg *et al.*, 2008[2]) (Richeldi *et al.*, 2011[1]). Note that nintedanib is also marketed as Vargatef. Vargatef is indicated in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent non-small cell lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy. A submission to NICE for the licenced indication was made to NICE in August 2014.

Nintedanib (Ofev) was granted EU marketing authorisation in January 2015 and is licensed in the UK with an indication for the treatment of idiopathic pulmonary fibrosis (IPF) in adults. In addition, nintedanib (Ofev) was designated as an orphan medicinal product by the European Commission in April 2013. Nintedanib is available as soft gelatine capsules and administered orally. The recommended dosage is 150mg twice daily. A dose reduction to 100mg twice daily is allowed as required for the management of adverse events (AEs). There is no restriction for the use of nintedanib (Ofev) within its marketing authorisation. The list price for nintedanib is £2,151.10 for a 30-day supply (150mgx60). A discount for nintedanib is available under the patient access scheme (PAS).

Pirfenidone, the only other drug to be approved for treatment of adults with mild to moderate IPF, is the main comparator in current clinical practice. It is recommended only in patients with a forced vital capacity (FVC) between 50% and 80% predicted. Patients who are not eligible for, or cannot tolerate, pirfenidone currently receive best supportive care (BSC). As such, there is a high unmet need for IPF patients.

Clinical effectiveness

The two key clinical trials of nintedanib for the treatment of IPF are the replicate randomised, double-blind, placebo-controlled phase III trials, INPULSIS-1 and INPULSIS-2 (Richeldi, 2014[4]). Both studies reached the primary endpoint of measuring the annual rate of decline in FVC. It was shown that nintedanib at a dose of 150mg twice daily consistently reduced the decline in FVC of IPF patients when compared to placebo over 52 weeks (nintedanib vs. placebo: -114.7 mL vs. -239.9 mL in INPULSIS-1, p<0.001; and -113.6 mL vs. -207.3 mL in INPULSIS-2, p<0.001).

Both studies also measured the time to first acute exacerbation as a key secondary endpoint. Despite no significant difference between nintedanib and the placebo group in INPULSIS-1 (Hazard Ratio [HR]: 1.15, p=0.67), there was a significant increase in the time to first acute exacerbation in the nintedanib group as compared with the placebo group in INPULSIS-2 (HR: 0.38, p=0.005).

The adjusted mean change from baseline to week 52 in St. George's Respiratory Questionnaire (SGRQ) score was assessed. As a lower total SGRQ score implies better health-related quality of life (HRQL), a smaller increase in total score reflects less disease deterioration. No significant difference between groups was measured in INPULSIS-1 (difference: -0.05, p=0.97). However, in INPULSIS-2, the adjusted mean change in SGRQ total score was significantly smaller for nintedanib compared with placebo (difference: -2.69, p=0.02). Other patient reported outcomes, such as the University of California in San Diego Shortness of Breath Questionnaire (UCSD-SOBQ), the cough domains of the Cough and Sputum Assessment Questionnaire (CASA-Q) and the Patient's Global Impression of Change (PGI-C) were measured and the results were largely in line with the SGRQ score, indicating an overall modest improvement in HRQL compared with placebo (Brun *et al.*, 2014[5], Brun *et al.*, 2014[6]).

In a prespecified pooled analysis, no significant between group difference in death from any cause, death from a respiratory cause or death that occurred between randomisation and 28 days after the last dose of study drug was measured (Richeldi *et al.*, 2011[1]). The most common adverse event for the nintedanib group in both studies was diarrhoea. Among the patients who had diarrhoea, the majority of events were of mild to moderate intensity (93.7% in INPULSIS-1 and 95.2% in INPULSIS-2) and rarely led to discontinuation of nintedanib (4.4% of patients). The rate of serious adverse events was similar in the nintedanib and placebo group in both trials (INPULSIS-1: 31.1% vs. 27.0%; INPULSIS-2: 29.8% vs. 32.9%).

Indirect treatment comparison vs. pirfenidone

As there is no head-to-head comparison between nintedanib and pirfenidone, a network meta-analysis (NMA) was performed to compare the clinical effectiveness and safety of nintedanib, pirfenidone and placebo.

The results indicate that the overall survival, loss of lung function, acute exacerbation and progression free survival for nintedanib and pirfenidone were comparable. In addition, nintedanib always had a lower discontinuation rate due to adverse events than pirfenidone. These results suggest that nintedanib could provide a much needed additional treatment option for IPF patients.

Base-case cost-effectiveness results

In the base-case analysis, nintedanib was associated with a total cost of £85,087. These were £87,479 and £25,359 for pirfenidone and BSC, respectively. The quality-adjusted life year (QALY) gains for nintedanib, pirfenidone and BSC were 3.67, 3.62 and 3.27, respectively. This indicated that nintedanib dominates pirfenidone (that is nintedanib costs less and provides greater benefits). When nintedanib was compared to BSC, the estimated ICER was £149,361 per QALY gained.

1.1 *Statement of decision problem*

Table 1: The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	Adults with idiopathic pulmonary fibrosis	Same as final scope issued by NICE	N/A
Intervention	Nintedanib	Same as final scope issued by NICE	N/A
Comparator (s)	1. Pirfenidone (where appropriate) 2. Best supportive care	Same as final scope issued by NICE	N/A
Outcomes	The outcome measures to be considered include: • pulmonary function parameters • physical function • exacerbation rate • progression-free survival • mortality • adverse effects of treatment • health-related quality of life.	Same as final scope issued by NICE	N/A
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective. The availability of any patient access schemes for the intervention or comparator technologies should be taken into account.	Same as final scope issued by NICE	N/A
Subgroups to be considered	N/A	N/A	N/A
Special considerations including issues related to equity or equality	No special considerations	No special considerations	N/A

1.2 *Description of the technology being appraised*

Table 2: Technology being appraised

UK approved name and brand name	UK approved name: Nintedanib Brand name: Ofev
Marketing authorisation/CE mark status	Nintedanib (Ofev) was granted EU marketing Authorisation (EU/1/14/979) on 15th January 2015. Nintedanib (Ofev) was also designated as an orphan medicinal product (EU/3/13/1123) by the European Commission on 26th April 2013.
Indications and any restriction(s) as described in the summary of product characteristics	Nintedanib (Ofev) is indicated in adults for the treatment of (IPF). There is no restriction under the current EU Marketing Authorisation (EU/1/14/979).
Method of administration and dosage	Nintedanib is administered orally. The recommended dosage is 150mg twice daily in adults ≥18 years. A dose reduction to 100mg twice daily is allowed as required for the management of AEs.

1.3 *Summary of the clinical effectiveness analysis*

- Two replicate randomised, double-blind, placebo-controlled phase III trials, INPULSIS-1 and INPULSIS-2 investigated the efficacy and safety of nintedanib for the treatment of IPF (Richeldi *et al.*, 2011[1]).
- In both INPULSIS studies, it was shown that nintedanib at a dose of 150mg twice daily consistently reduced the decline in FVC of IPF patients when compared to placebo over 52 weeks (nintedanib vs. placebo: -114.7 mL vs. -239.9 mL in INPULSIS-1, p<0.001; and -113.6 mL vs. -207.3 mL in INPULSIS-2, p<0.001).
- Time to first acute exacerbation was measured as a key secondary endpoint. Despite no significant difference between nintedanib and the placebo group in INPULSIS-1 (HR: 1.15, p=0.67), there was a significant increase in the time to first acute exacerbation in the nintedanib group as compared with the placebo group in INPULSIS-2 (HR: 0.38, p=0.005).
- The adjusted mean change from baseline to week 52 in SGRQ score was assessed. No significant difference between groups was measured in INPULSIS-1 (difference: -0.05, p=0.97). In INPULSIS-2, the adjusted mean change in SGRQ total score was significantly smaller for nintedanib compared with placebo (difference: -2.69, p=0.02).
- A numerical reduction in all-cause mortality with nintedanib versus placebo (HR 0.70 [95% CI: 0.46, 1.08]; p=0.0954). The proportion of patients who died was 5.8% in the nintedanib group vs 8.3% in the placebo group
- The most common adverse event for the nintedanib group in both studies was diarrhoea. Among the patients who had diarrhoea, the majority of diarrhoea adverse events were of mild to moderate intensity (93.7% in INPULSIS-1 and 95.2% in INPULSIS-2) and rarely led to discontinuation of nintedanib (4.4% of patients).
- The rate of serious adverse events was similar in the nintedanib and placebo groups in both trials (INPULSIS-1: 31.1% vs. 27.0%; INPULSIS-2: 29.8% vs. 32.9%).

Indirect treatment comparison – nintedanib vs. pirfenidone and placebo

A network meta-analysis was performed to compare nintedanib, pirfenidone and placebo for overall survival (OS), acute exacerbations, loss of lung function, progression free survival (PFS), 6 minute walk test, serious cardiac events and serious gastrointestinal (GI) events. Details of the studies included and the structure of the network are discussed in detail in section 4.9 and 4.10.

With reference to overall survival, no intervention was statistically significantly different to placebo and nintedanib was comparable to pirfenidone. No intervention was found to be statistically different to placebo for acute exacerbations. However, the results changed when homogeneity is considered; indicating statistical significance in favour of nintedanib. For loss of lung function, the results were similar for nintedanib and pirfenidone and were statistical significantly different to placebo, indicating greater clinical benefit.

The safety profiles are comparable between nintedanib and pirfenidone. Discontinuations due to adverse events for nintedanib were always lower than for pirfenidone.

These results indicate that nintedanib could provide a much needed additional treatment option for IPF patients.

Strengths and limitations of the evidence

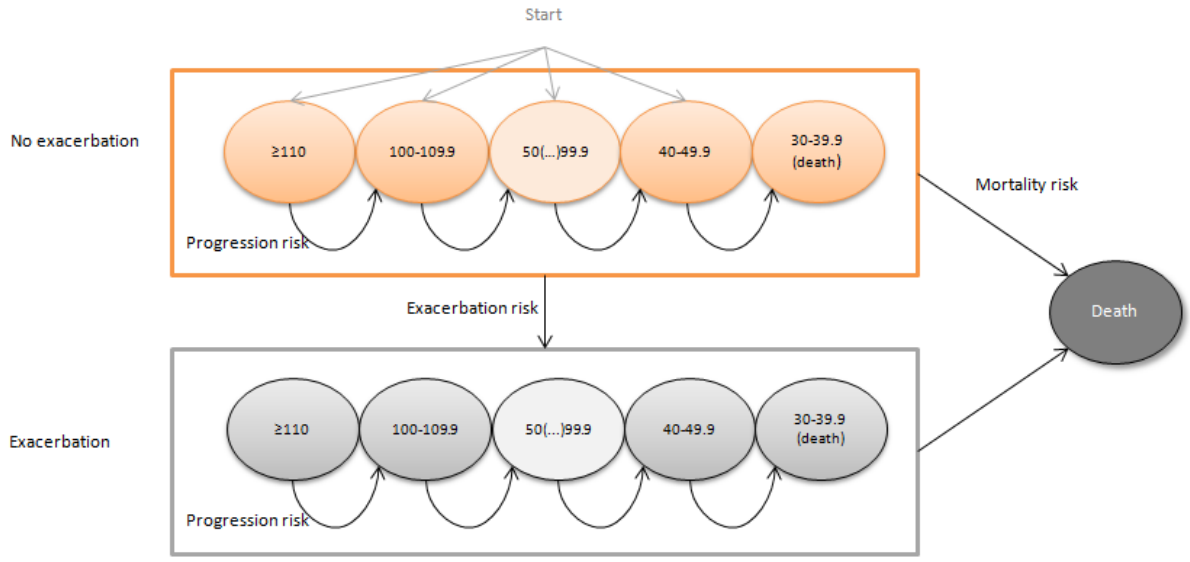
The strength of the evidence is that both INPULSIS-1 and INPULSIS-2 are well-designed randomised, double-blind, placebo-controlled clinical trials, thus limiting potential biases. However, there is no head-to-head comparison between active treatments. As such, systematic review was conducted to identify all published evidence for the indication and comparators. Relevant evidence was then used in a NMA in an attempt to provide evidence in the absence of direct treatment comparisons.

1.4 *Summary of the cost-effectiveness analysis*

Cost-effectiveness model

A Markov model was developed for the cost-effectiveness analysis for nintedanib vs. pirfenidone and BSC, using time-dependent and constant transition probabilities to describe decline in lung function and incidence of first acute exacerbations over the cohort's lifetime. The model structure is illustrated in Figure 1.

Figure 1: Model structure



Possible transitions were defined between each of the mutually exclusive health states and the probability of each transition occurring within each cycle was assigned.

The cohort entered the model at different FVC % predicted health states without exacerbation. The possible transitions were:

- death,
- loss of lung function (progression to a health state with lower FVC % predicted),
- exacerbation,

- loss of lung function combined with exacerbation,
- remaining in the same health state.

Transitions were evaluated over a 3-month period. It was assumed that once progressed to a lower FVC % predicted state, the cohort could not regress back to health states with improved lung function (higher FVC % predicted). Moreover, once an exacerbation occurred, the cohort could not move back to a health state without exacerbation and would continue in the health states with exacerbation history. Death could occur in two ways – at any point in the model (and from any health state) based on the survival analysis of the clinical trial data; or, at the point that patients reached a level of FVC % predicted of 30%, which was assumed to be an unsustainable level of lung function.

Base-case cost-effectiveness results

The incremental cost-effectiveness results are shown in Table 3 and Table 4.

1. Base-case deterministic results for nintedanib vs. pirfenidone

The total costs for the nintedanib arm were £85,087, and £87,479 for the pirfenidone arm. With an incremental total cost of -£2,392 and incremental total QALYs of 0.0471, the nintedanib strategy dominated the pirfenidone one (that is, nintedanib costs less and provides greater benefits). With the nintedanib PAS in place, the total costs for the nintedanib arm dropped to [redacted], resulting in an incremental cost of [redacted] vs. pirfenidone.

2. Base-case deterministic analysis for nintedanib vs. BSC

The total cost for the nintedanib treatment strategy was estimated to be £85,087 and £25,359 for BSC. With an incremental total cost of £59,728 and an estimated incremental QALY gained of 0.40, the estimated ICER for nintedanib vs. BSC was £149,361 per QALY gained. With the nintedanib PAS in place, the total costs for the nintedanib arm dropped to £[redacted], resulting in an incremental cost of [redacted] and an ICER of £[redacted] gained vs BSC.

Sensitivity analysis in section 5.8 shows that nintedanib always dominated pirfenidone when one way sensitivity analyses were performed around the 95% confidence values of all model parameters and several model assumptions. When nintedanib was compared with best supportive care, sensitivity analysis showed that mortality risk estimates were very strong drivers of the model results. Results were also sensitive to discontinuation probabilities, treatment cost and progression probabilities.

Figure 2: Scatter plot: incremental cost-effectiveness analysis NDB vs BSC and NDB vs PFN (without nintedanib PAS) shows the cost effectiveness of nintedanib vs pirfenidone and nintedanib vs BSC without the nintedanib PAS in place. **Error! Reference source not found.** shows the cost effectiveness of nintedanib vs pirfenidone and nintedanib vs BSC with the nintedanib PAS in place. It can be seen that nintedanib is below the £20,000/QALY willingness-to-pay threshold both with and without the PAS in place. In both cases nintedanib is cheaper and more effective than pirfenidone.

Strength and weaknesses of the cost-effectiveness analysis

A Markov model is considered appropriate for the analysis as it reflects the progressive nature of IPF and is consistent with NICE modelling in IPF. Given that FVC "is a widely used measure of disease status and a common endpoint in clinical trials in IPF patients" (du Bois *et al.*, 2011[7]), it was selected to be the main factor of disease progression. In particular, FVC % predicted was used for Markov model health state, as it reflected the absolute state of patient condition adjusted for body capacity; age, gender, and height. This removed some of the heterogeneity of the health-state members. Furthermore, treatment effects for pirfenidone from the CAPACITY trials have been published in FVC % predicted, allowing for an indirect treatment comparison with the data from the nintedanib INPULSIS trials.

However, as with many cost-effectiveness analyses using a Markov model, limitations come from the assumption that the transition probabilities remain the same over time. In addition, short-term data from clinical trials were extrapolated over lifetime horizon in the model.

Table 3: Incremental cost-effectiveness results without nintedanib PAS

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs. baseline (vs. BSC) (QALYs)	ICER (£) incremental (QALYs)
BSC	£25,359.09	4.3602	3.2721					
PFN	£87,479.15	4.8558	3.6249	62,120.05	0.4956	0.3528	£176,081	Dominated by NDB
NDB	£85,087.49	4.8593	3.6720	-2,391.66	0.0035	0.0471	£149,361	£149,361

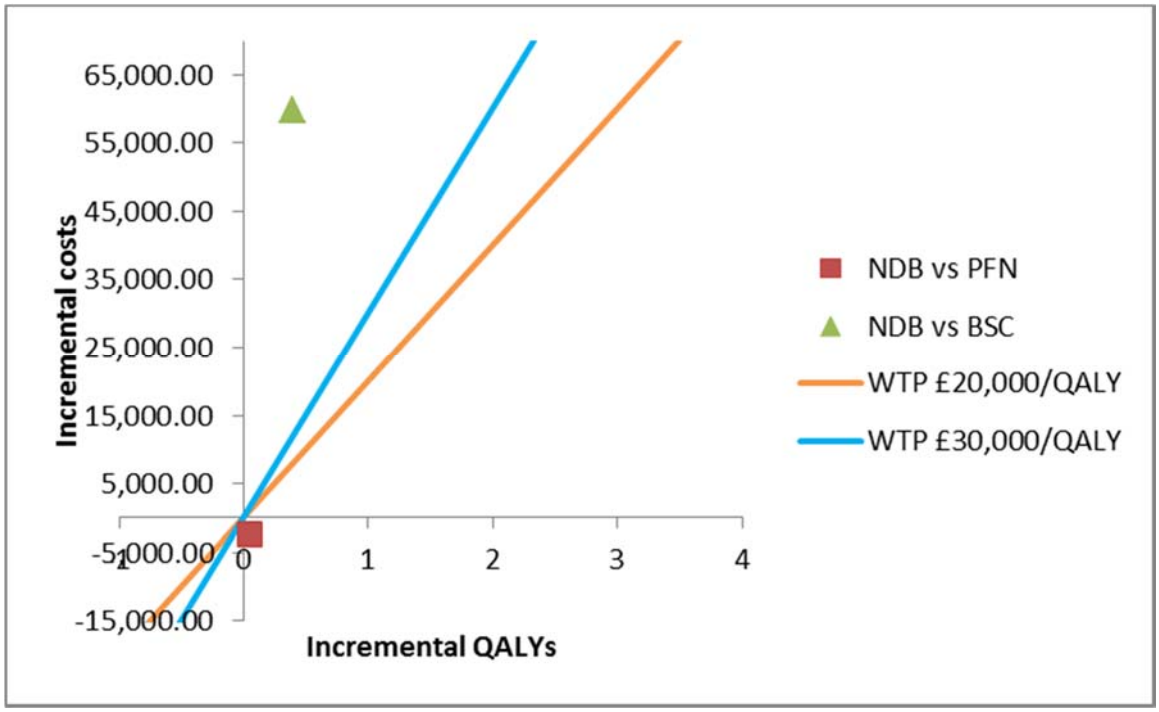
* ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 4: Incremental cost-effectiveness results with nintedanib PAS

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs. baseline (vs. BSC) (QALYs)	ICER (£) incremental (QALYs)
BSC	£25,359.09	4.3602	3.2721					
PFN	£87,479.15	4.8558	3.6249	62,120.05	0.4956	0.3528	£176,081	Dominated by NDB
NDB	[redacted]	4.8593	3.6720	[redacted]	0.0035	0.0471	[redacted]	[redacted]

*ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Figure 2: Scatter plot: incremental cost-effectiveness analysis NDB vs BSC and NDB vs PFN (without nintedanib PAS)



[redacted]

2 The technology

2.1 *Description of the technology*

2.1.1 Give the brand name, UK approved name, the therapeutic class and a brief overview of the mechanism of action. For devices, provide details of any different versions of the same device.

- Brand name: Ofev
- UK approved name: Nintedanib
- Therapeutic class: Tyrosine kinase inhibitor

Nintedanib is a small molecule TKI which specifically targets three growth factor receptors: the VEGFR, FGFR and PDGFR (Hilberg *et al.*, 2008[2], Wollin *et al.*, 2014[3]). It has been shown that these receptors are potentially involved in the pathomechanisms of pulmonary fibrosis. By blocking these signalling pathways that are involved in the fibrotic processes, it is believed that nintedanib has the potential to reduce disease progression, slowing the decline of lung function (Hilberg *et al.*, 2008[2]) (Richeldi *et al.*, 2011[1]).

2.2 *Marketing authorisation/CE marking and health technology assessment*

2.2.1 Indicate whether the technology has a UK marketing authorisation/CE marking for the indications detailed in this submission. If so, give the date on which this was received. If not, state the current UK regulatory status, with relevant dates (for example, date of application and/or expected date of approval from the Committee for Human Medicinal Products).

Nintedanib (Ofev) has EU Marketing Authorisation and is indicated in adults for the treatment of IPF. The EU marketing authorisation (EU/1/14/979) was granted on 15th January 2015. In addition, Ofev was designated as an orphan medicinal product on 26th April 2013 (EU/3/13/1123).

2.2.2 Give the (anticipated) indication(s) in the UK. For devices, provide the date of (anticipated) CE marking, including the indication for use. If a submission is based on the company's proposed or anticipated marketing authorisation, the company must advise NICE immediately of any variation between the anticipated and the final marketing authorisation approved by the regulatory authorities.

In the UK, nintedanib (Ofev) is indicated in adults for the treatment of IPF.

2.2.3 Summarise any (anticipated) restrictions or contraindications that are likely to be included in the (draft) summary of product characteristics (SmPC).

There are no restrictions for the use of nintedanib (Ofev) within its Marketing Authorisation.

Nintedanib (Ofev) is contraindicated in patients with hypersensitivity to nintedanib, to peanut or soya, or to any of the excipients.

2.2.4 Include the (draft) SmPC for pharmaceuticals or information for use (IFU) for devices in an appendix.

The SmPC for nintedanib (Ofev) is included in Appendix E.

2.2.5 Provide the (draft) assessment report produced by the regulatory authorities (that is, the European public assessment report for pharmaceuticals) and a (draft) technical manual for devices in an appendix.

The assessment report by the European Medicines Agency (EMA) is included in Appendix F.

2.2.6 Summarise the main issues discussed by the regulatory authorities (preferably by referring to the [draft] assessment report [for example, the European public assessment report]). State any special conditions attached to the marketing authorisation (for example, if it is a conditional marketing authorisation)

The Committee for Medicinal Products for Human Use (CHMP) discussed in the assessment report both non-clinical and clinical aspects, as well as the benefit-risk balance of nintedanib.

Based on the CHMP review of data on quality, safety and efficacy, the CHMP considered by consensus that the risk-benefit balance of nintedanib (Ofev) when indicated in adults for the treatment of IPF is favourable and therefore recommended the granting of the marketing authorisation.

No special conditions were attached to the marketing authorisation.

2.2.7 If the technology has not been launched, supply the anticipated date of availability in the UK.

Nintedanib (Ofev) commercial stock was made available in April 2015.

2.2.8 State whether the technology has regulatory approval outside the UK. If so, please provide details.

Nintedanib (Ofev) has EU marketing authorisation (EU/1/14/979) and is approved for use in the European Union. In addition, Ofev was approved in the United States by the U.S. Food and Drug Administration (FDA) on 15th October 2014.

2.2.9 State whether the technology is subject to any other health technology assessment in the UK. If so, give the timescale for completion.

A full submission was made to the Scottish Medicines Consortium (SMC) in April 2015, and advice is expected in August 2015.

2.3 *Administration and costs of the technology*

2.3.1 For pharmaceuticals, complete the table 'Costs of the technology being appraised' in the company evidence submission template, including details of the treatment regimen and method of administration. Indicate whether the acquisition cost is list price or includes a patient access scheme, and the anticipated care setting. Specify the sources of information and data used to complete the table, for example SmPC or trial data. For more information see section 5.5 of the NICE guide to the methods of technology appraisal.

Table 5: Costs of the technology being appraised

	Data	**Source**
Pharmaceutical formulation	Soft gelatine capsule	SmPC
Acquisition cost (excluding VAT)*	The list price for nintedanib is £2,151.10 for 60 x 150mg capsules	Anticipated list price
Method of administration	Oral	SmPC
Doses	150mg	SmPC
Dosing frequency	Twice daily	SmPC
Average length of a course of treatment	Nintedanib is administered until treatment progression or unacceptable adverse events. In the INPULSIS studies, patients received nintedanib for 52 weeks.	INPULSIS studies (Ref 4)
Average cost of a course of treatment	£2,151.10 every 30 days	Anticipated list price
Anticipated average interval between courses of treatments	Continuous daily treatment until disease progression or unacceptable adverse events.	INPULSIS studies (Ref 4)
Anticipated number of repeat courses of treatment	Continuous daily treatment until disease progression or unacceptable adverse events.	INPULSIS studies (Ref 4)
Dose adjustments	Dose reduction to 100mg twice daily is permitted in order to manage adverse events. Upon resolution of an adverse event, the dose can be reinstated to 150mg twice daily.	INPULSIS studies (Ref 4)
Anticipated care setting	Primary Care	INPULSIS studies (Ref 4)
* Indicate whether this acquisition cost is list price or includes an approved patient access scheme. When the marketing authorisation or anticipated marketing authorisation recommends the intervention in combination with other treatments, the acquisition cost of each intervention should be presented.		

2.3.2 Provide details of any patient access scheme that has been referred to NICE for inclusion in the technology appraisal by ministers and formally agreed by the company with the Department of Health before the date of evidence submission to NICE for the technology. For more information see section 5 of the NICE guide to the processes of technology appraisal.

A patient access scheme is in place. The patient access scheme is in the form of a simple discount [redacted] from the list price.

2.3.3 For devices, provide the list price and average selling price in a table similar to the table presented in the template, 'Costs of the technology being appraised'. If the unit cost of the device is not yet known, provide details of the anticipated unit cost, including the range of possible unit costs.

Not applicable.

2.4 *Changes in service provision and management*

2.4.1 State whether additional tests or investigations are needed (for example, diagnostic tests to identify the population for whom the technology is indicated in the marketing authorisation) or whether there are particular administration requirements for the technology. For more information see section 5.9 of the NICE guide to the methods of technology appraisal.

None

2.4.2 Identify the main resource use to the NHS associated with the technology being appraised. Describe the location or setting of care (that is, primary and/ or secondary care, commissioned by NHS England specialised services and/or clinical commissioning groups), staff costs, administration costs, monitoring and tests. Provide details of data sources used to inform resource estimates and values.

Nintedanib will be used primarily in the primary care setting, commissioned by clinical commissioning groups. Nintedanib is taken orally and no additional monitoring is required.

Liver function tests were assumed to be routinely performed on patients receiving nintedanib. The cost of a liver panel blood test was estimated at £3.01 (NHS Reference Costs 2012/13, DAPS05 Haematology (Department of Health, 2013[8]) and is included as one of the cost inputs in the cost-effectiveness analysis. It was assumed that all patients on active treatment would incur this cost, at a quarterly frequency (every 3 months). This frequency schedule is the same as the maintenance test frequency recommended by the pirfenidone SPC (Electronic Medicines Compendium (eMC), Accessed May 2014[9]).

2.4.3 Specify if the technology requires additional infrastructure in the NHS to be put in place.

No additional infrastructure required.

2.4.4 State if and to what extent the technology will affect patient monitoring compared with established clinical practice in England.

The use of nintedanib (Ofev) will not affect established clinical practice in England.

2.4.5 State whether there are any concomitant therapies specified in the marketing authorisation or used in the key clinical trials (for example, for managing adverse reactions) administered with the technology.

Nintedanib (Ofev) is a monotherapy and no concomitant therapies were specified in the marketing authorisation or administered with the technology in the key clinical trials.

2.5 *Innovation*

Treatment options for IPF are very limited. The only treatment recommended by NICE is pirfenidone, which is only recommended for IPF patients with a FVC value between 50% and 80% predicted (NICE, 2014[10]). As pirfenidone is only indicated in mild to moderate IPF patients, if the FVC predicted value is outside of the 50% to 80% range, the only alternative is best supportive care. Therefore, there is high unmet need to improve the treatment options for IPF patients.

Nintedanib is a unique TKI simultaneously acting on three growth factor receptors (VEGFR, FGFR and PDGFR) (Richeldi *et al.*, 2011[1], Hilberg *et al.*, 2008[2], Wollin *et al.*, 2014[3]). Nintedanib is licensed for all adults with IPF regardless of the FVC predicted value. In the UK, nintedanib fits into the clinical pathway not only as an alternative to pirfenidone, but also as a treatment option for IPF patients whose FVC predicted value is outside of the 50% to 80% range (as demonstrated in sub-analyses that looked into patient groups with different FVC predicted values (Kolb M, 2015[11]). In addition, the recommended dose of nintedanib is one 150mg capsule twice daily. This simple regime significantly reduces the “pill burden” when compared to pirfenidone, for which the recommended dose is three 267mg capsules three times a day (Electronic Medicines Compendium (eMC), Accessed May 2014[9]).

3 Health condition and position of the technology in the treatment pathway

3.1 Provide a brief overview of the disease or condition for which the technology is being used. Include details of the underlying course of the disease.

IPF is a chronic lung disease characterised by scarring or thickening of lung tissues, which leads to worsening dyspnoea and decline in lung function (Richeldi *et al.*, 2011[1], Hilberg *et al.*, [2]). IPF is usually progressive and results in significant disabilities. Median survival of IPF patients is 2 to 5 years from the time of diagnosis (Richeldi et al., 2011[1]). An estimated 15,000 patients are affected by IPF in the UK. Each year, an additional 5,000 new cases are diagnosed and 5,000 patients with this disease spectrum will die (Richeldi *et al.*, 2014[12]).

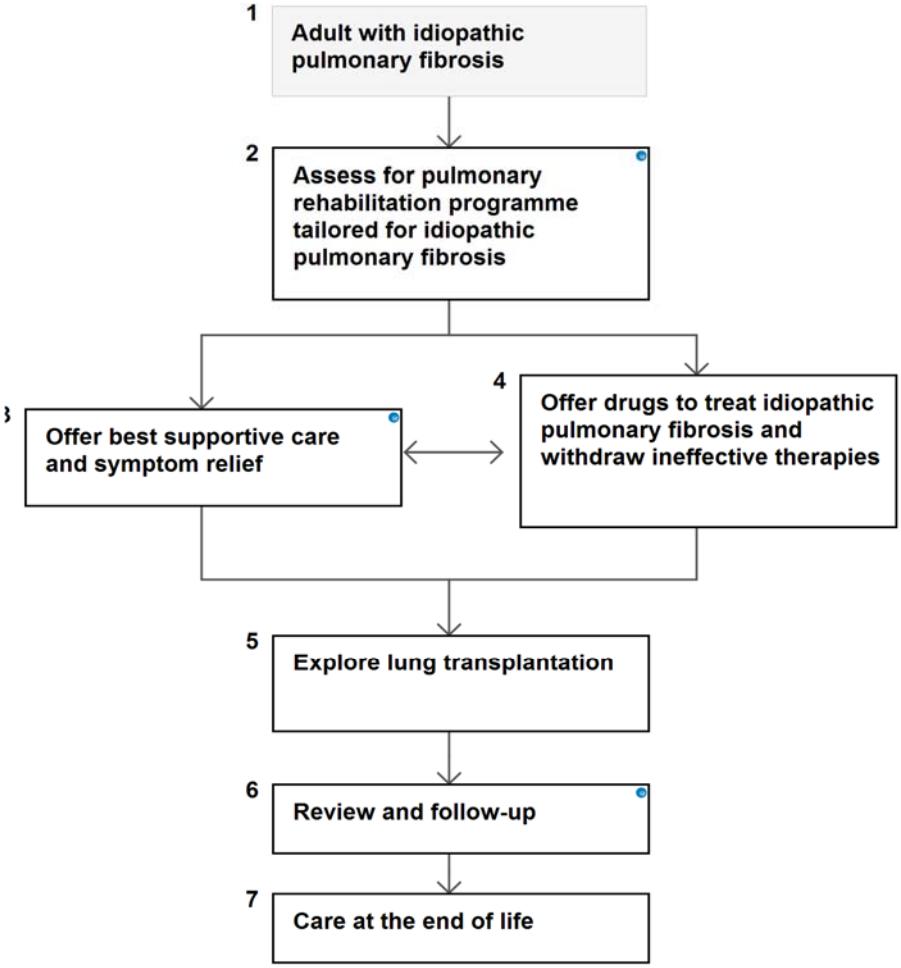
The cause of IPF remains unknown and currently there is no cure for the disease. Management options are limited and the only licensed therapy in England is pirfenidone, which is approved for restricted use in the treatment of mild to moderate IPF in adults with a predicted FVC between 50% and 80%. If pirfenidone fails or is not tolerated, patients are typically provided with best supportive care (Electronic Medicines Compendium (eMC), Accessed May 2014[9]). As such, a new treatment option for IPF is extremely valuable.

3.2 Describe the effects of the disease or condition on patients, carers and society.

IPF is a progressive disease and the median survival in the UK is between 2 and 5 years from the time of diagnosis (NICE, 2014[10]) (Raghu *et al.*, 2011[13]). IPF patients demonstrate impaired HRQL in many life domains such as physical health. Respiratory symptoms, energy levels and degrees of independence are seriously impacted, and disability increases with the severity of the disease (for example due to the need for supplementary oxygen) (Swigris *et al.*, 2011[14]). IPF also significantly impairs the psychological and emotional well-being of patients. Depressive symptoms, poor sleep quality, feelings of isolation and end-of-life decision making are some of the emotional challenges faced by IPF patients (Swigris *et al.*, 2011[14]). Moreover, IPF patients are often elderly (mean age at diagnosis is 66 years (Rafii *et al.*, 2013[15]) and with the additional burden of the disease, it is more likely that they will need carer support. There are also significant societal costs associated with IPF as patients tend to be unemployed, or retire early compared to age-matched controls (Brun *et al.*, 2012[16]).

3.3 Present the clinical pathway of care that shows the context of the proposed use of the technology. This information may be presented in a diagram. Explain how the new technology may change the existing pathway. If a relevant NICE clinical guideline has been published, the response to this point should be consistent with the guideline and any differences should be explained.

Figure 3: Current clinical pathway of care for idiopathic pulmonary fibrosis



(NICE, 2013[17], NICE, 2013[18])

Interventions with best supportive care or pharmacological agents are recommended before consideration of lung transplantation.

1. Pharmacological intervention

Pirfenidone is the only drug recommended for the treatment of IPF, and only if:

- the person has a forced vital capacity (FVC) between 50% and 80% predicted **and**
- the manufacturer provides pirfenidone with the discount agreed in the patient access scheme.
- treatment with pirfenidone should be discontinued if there is evidence of disease progression (a decline in per cent predicted FVC of 10% or more within any 12 month period).

2. Best supportive care

Offer best supportive care to people with IPF from the point of diagnosis.

Proposed use of nintedanib in IPF

Nintedanib fits well in the existing clinical pathway as an additional option of pharmacological intervention and additionally can meet the current unmet need that pirfenidone is only recommended in patients who have a FVC between 50% and 80% predicted.

3.4 Provide information about the life expectancy of people with the disease or condition in England and the source of the data. Please provide information on the number of people with the particular therapeutic indication for which the technology is being appraised. If the marketing authorisation also includes other therapeutic indications for the technology, provide information about the numbers of people with these diseases or conditions in England and provide the source of the data. This is to assess whether the technology may be suitable for consideration as a 'life-extending treatment at the end of life' as described in section 6.2.10 of the NICE guide to the methods of technology appraisal.

The life expectancy of IPF patients is approximately 2 to 5 years from the time of diagnosis. It is estimated that there are currently 15,000 patients affected by IPF in the UK (Raghu *et al.*, 2011[13], Meltzer *et al.*, 2008[19], Navaratnam *et al.*, 2011[20]).

3.5 Provide details of any relevant NICE guidance, pathways or commissioning guides related to the condition for which the technology is being used. Specify whether any subgroups were explicitly addressed.

NICE clinical guidance CG163 – Idiopathic pulmonary fibrosis: The diagnosis and management of suspected idiopathic pulmonary fibrosis (NICE, 2013[18])

NICE technology appraisal (TA) 282 – Pirfenidone for treating idiopathic pulmonary fibrosis (NICE, 2014[10])

NICE quality standard (QS) 79 – Idiopathic pulmonary fibrosis (NICE, 2013[21])

NICE pathway – Idiopathic pulmonary fibrosis (NICE, 2013[22])

3.6 Provide details of other clinical guidelines (for example, UK guidance from the royal societies or European guidance) and national policies.

BTS Guidelines

Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. Issued 2008. (Society, [23]) Available at:

https://www.brit-thoracic.org.uk/document-library/clinical-information/interstitial-lung-disease/ild-guidelines/bts-interstitial-lung-disease-guideline/

ATS/ERS/JRS/ALAT

Raghu G, Collard HR, Egan JJ, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. Issued 2011. (Raghu *et al.*, 2011[13])

3.7 Describe any issues relating to current clinical practice, including any variations or uncertainty about established practice.

Management options for IPF remain limited. The current recommendation by NICE states that pirfenidone should only be used in patients with a FVC between 50% and 80% predicted. In contrast, nintedanib is indicated in adults for the treatment of IPF regardless of the percentage FVC predicted. As such, nintedanib not only offers a different treatment option to the only drug available for IPF so far, but is also a much needed treatment for those who are not eligible for pirfenidone.

3.8 NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. For further information about equality issues see NICE's equality scheme.

Provide an assessment of whether the use of this technology is likely to raise any equality issues. Please document if there are any potential issues that:

- could exclude from full consideration any people protected by the equality legislation who fall within the patient population for whom the technology is or will be licensed
- could lead to recommendations that have a different impact on people protected by the equality legislation compared with the wider population, for example by making it more difficult in practice for a specific group to access the technology
- could lead to recommendations that have any adverse impact on people with a particular disability or disabilities.

Boehringer Ingelheim Ltd. does not believe that nintedanib will be associated with any equality issues.

Please provide any evidence that would enable the Committee to identify and consider the impact of equality issues. State how the analysis has addressed these issues.

Not applicable.

4 Clinical effectiveness

SECTION SUMMARY

- The aim of this section is to demonstrate the efficacy and safety of nintedanib versus other comparators used in English clinical practice for treatment of IPF.
- A systematic literature review identified 12 RCTs of interest which enabled an indirect comparison on three outcomes to be considered for overall survival, acute exacerbations, and loss of lung function.
- The indirect comparisons indicated that there was little difference between treatments in terms of overall survival, but that nintedanib showed better or equivalent efficacy in terms of slowing lung function decline and acute exacerbations.

4.1 *Identification and selection of relevant studies*

A systematic review of the literature was conducted to identify relevant studies for this indication involving comparators currently used in UK clinical practice. Following discussions at the draft scoping meeting regarding current clinical practice in England this review included off-label use of n-acetylcysteine (NAC). NAC was not included in the final scope, so has not been implemented in the cost effectiveness modelling.

Table 1 presents the decision problem relevant to English practice that is addressed in this submission and was used to develop the search strategy and subsequently identify the evidence for data analysis.

Search objective

The clinical search aimed to review evidence on the efficacy and safety of nintedanib and its comparator(s) for the treatment of IPF patients. Four clinical searches were conducted to identify:

- Randomised controlled trials (RCTs) for nintedanib;
- RCTs for nintedanib or its comparators (to be used in a potential network meta-analysis (NMA)) (see section 5 below);
- Non-RCTs for nintedanib; and
- Studies designed primarily to assess safety outcomes for nintedanib (see section 4).

Identification of studies

Databases

In accordance with Centre for Reviews and Dissemination (CRD) advice (Dissemination, 2009[24]) the following computerised bibliographic databases were searched on 18/09/2014:

- MEDLINE;
- MEDLINE In-Process;
- Embase;
- Cochrane Library (including Cochrane Central Register of Controlled Trials [CENTRAL].

MEDLINE and Embase were searched via the Ovid® search platform, and the Cochrane Library was searched directly.

Search strategy

A single clinical search strategy was performed across the four clinical searches as the same population and disease terms were required, and it was expected that some studies would be required for more than one of the four searches. In addition to population and disease terms, the search strings included intervention terms for the intervention of interest (relevant for all four clinical searches) and also for the three comparator regimens (relevant for the NMA search). The search was not limited by outcome terms, country, language, study design or date but studies published in languages other than English were not reviewed in full. Full details of the search strategies are provided in Appendix A.

Conference, clinical trial registry and agencies responsible for the scientific evaluation of medicines search

To identify potentially relevant data for products still in development the following sources were searched: conference proceedings; the European Medicines Agency (EMA) website; The U.S. Food and Drug Administration (FDA) website; and the ClinicalTrials.gov register.

Conference proceedings search

Conference proceedings were searched from January 2012 to December 2014 to identify publications in adults with IPF, treated with nintedanib or its comparators. The proceedings of the following conferences were searched:

- European Respiratory Society Congress (ERS);
- British Thoracic Society Meeting (BTS);
- American Thoracic Society (ATS).

Clinical trials register (ClinicalTrials.gov) search

The clinical trials register of ClinicalTrials.gov was searched to identify completed (with results) interventional (phase I, II, III or IV) and observational studies conducted in adults with IPF, treated with nintedanib or its comparators. The advanced interface was used to gain a higher precision in the search. The following terms were applied (input in the “Search Term” line): "Idiopathic pulmonary fibrosis" OR IPF OR "interstitial pneumonia" OR "interstitial lung disease". Studies in children and animals or with “unknown” status were excluded. No restrictions on intervention, comparator, country or date were applied.

Regulatory agencies search

The EMA and FDA websites were searched to identify potentially relevant information on products that are still in development or in current regulatory submission stages.

Study selection

Study selection for RCTs

The process for study selection was performed in two stages. The first stage involved the review of the records title and abstract by two independent researchers against the pre-determined study eligibility criteria. In the cases where exclusion based on the titles/abstracts was not possible, the full publication was retrieved and evaluated against the eligibility criteria in the second stage.

In the second stage, full-text publications were evaluated against the pre-set eligibility criteria. This was performed by two independent researchers. The selection was conducted in line with pre-set eligibility criteria for Population, Intervention(s)/Comparator(s), Outcome(s) and Study design (PICOS), presented in Table 6. Any judgment based on titles/abstracts or full-text selection where consensus was not achieved was reviewed again by both researchers and, when agreement could not be reached, by a third independent researcher.

Table 6: Eligibility criteria used in the search for RCTs

Criteria	Inclusion criteria	Exclusion criteria
Population	• Adult patients with IPF (including subgroups of patients, e.g.subgroups based on 6MWT SpO2 and %VC)	• Paediatric and younger than 18; • Healthy participants; • Non-human.
Intervention/Comparator	Nintedanib in comparison with: • Pirfenidone; • N-acetylcysteine (monotherapy); • Best supportive care* (placebo).	Any studies not containing nintedanib; NAC triple therapy; Lung transplantation.
Outcomes	**Efficacy** • Change from baseline in FVC (L); • Change from baseline in percent predicted FVC (%); • Annual rate of change in FVC (ml/year); • Change from baseline in DLco (ml/min/mmHg); • Change from baseline in 6MWT distance (m); • Change from baseline in worst SpO_2 during 6MWT; • Progression free survival[1]. • Exacerbation rate: acute IPF exacerbation; **Safety** • Adverse effects of treatment; • Mortality (hazard rate, mortality rate). **HRQL** • Change from baseline in HRQL as measured by disease-specific (e.g. SGRQ) or generic measure of HRQL (e.g. EQ-5D, SF-36D)	• No outcomes of interest
Study design	• Blinded or open-label RCTs (phase II, III, IV) of any duration; • Parallel-design extension of RCTs; • Post-hoc analyses of RCTs; • Pooled analyses of RCTs; • Studies published as abstracts or conference presentations were eligible for the primary analysis of clinical effectiveness if adequate data are provided.	• RCTs (phase I); • Non-RCT studies including observational studies, experimental studies and single arm prospective trials; • Reviews and systematic literature reviews** of RCTs; • Case reports; • Critical appraisals; • Update or commentaries on data

		published elsewhere; • Notes, letters or editorials.
Language	• English language.	• Non-English language.

Abbreviations: 6MWT, 6 Minute Walk Test; EQ-5D, DLCO, diffusion capacity of the lung for carbon monoxide; EuroQol 5-dimensions questionnaire; HRQL, Health-Related Quality of Life; IPF, idiopathic pulmonary fibrosis; RCT, Randomised Controlled Trial; SF-36D, Short Form Health Survey 36; SGRQ, St. George's Respiratory Questionnaire; SpO2, peripheral capillary oxygen saturation.

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

**Systematic literature reviews will be excluded from the main database search but their reference lists were reviewed for RCTs not captured by the main database search and included as a hand search in the review.

Data Extraction

Following full-text review, records were divided in two groups: primary and related references. Primary references presented original evidence and were thus selected for data extraction. Related references presented evidence already captured in the primary reference studies. Data extraction was performed by one researcher, documented in Microsoft Excel data extraction forms and reviewed by a second researcher. Results published as figures or graphs were extracted using the DigitizeIT software. Data on study design, patient characteristics, efficacy outcomes, safety outcomes, and discontinuation were extracted.

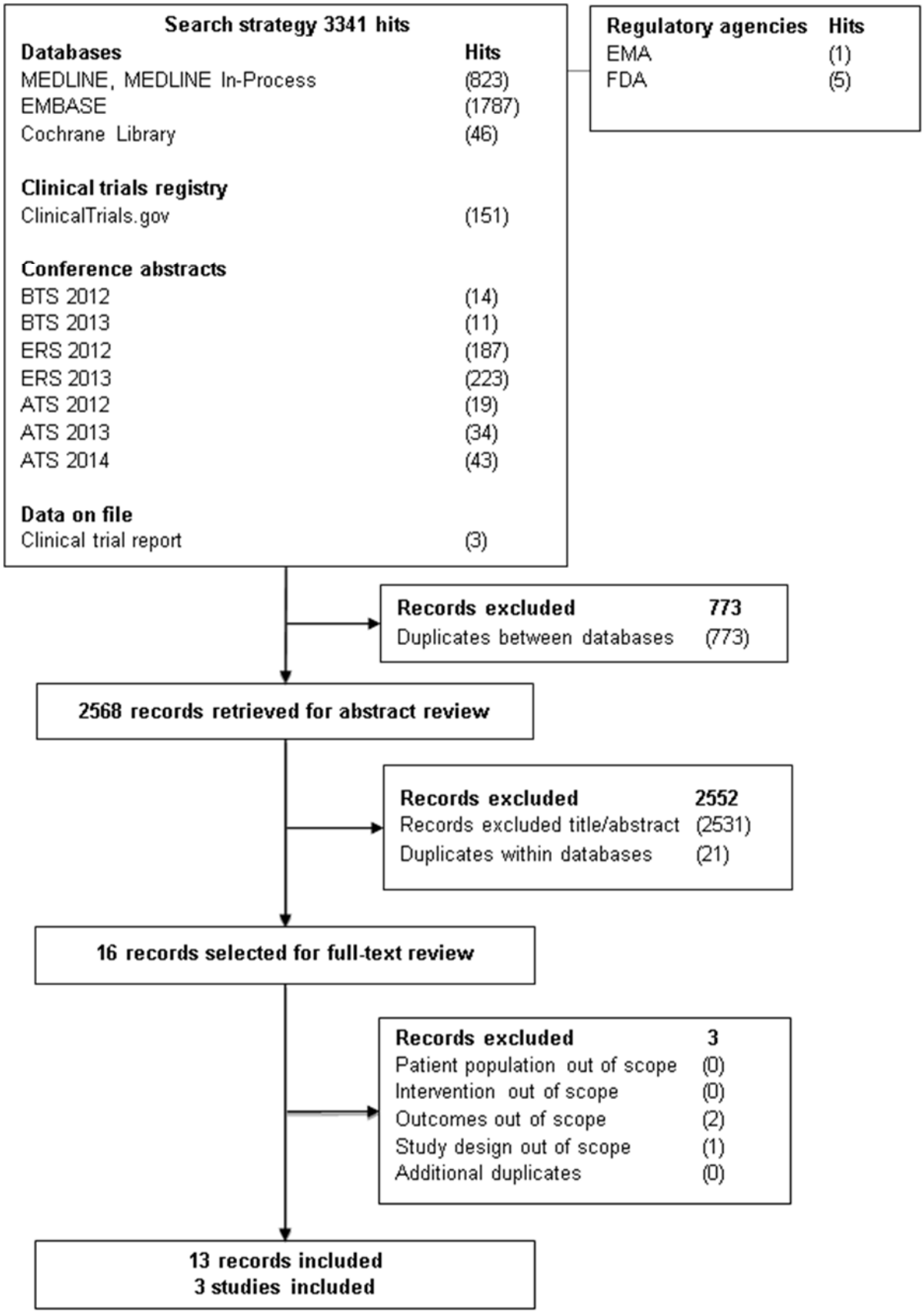
Results

The search identified 3,338 records from databases and registries. Three unpublished clinical study reports (for the Phase II TOMORROW (Brun *et al.*, 2012[16]) trial, and the Phase III INPULSIS-1 (Brun *et al.*, 2014[5]) and INPULSIS-2 (Brun *et al.*, 2014[6]) trials) were also added, accounting for a total of 3,341 records. Three systematic literature reviews (Agarwal *et al.*, 2008[25], Richeldi, 2012[26], Spagnolo, 2010[27]) were identified in the search and their reference lists were checked to confirm that there were no other RCTs which had been missed in the search. Since no further studies were identified this additional hand search step further confirmed that all relevant RCTs had been captured. A total of 13 records were identified, summarising the results of 3 RCTs (TOMORROW, INPULSIS-1 and INPULSIS-2). These RCTs investigated the intervention of interest (nintedanib) and are referred to in the records presented in Table 7 and Table 8. Studies excluded following full-text review are listed in along with the reason for exclusion. A flow diagram of the numbers of studies included and excluded at each stage of the selection process is provided in Figure 5.

Data was extracted for the full-text publications reporting evidence for the TOMORROW and INPULSIS-1 and 2 trials in Table 16 through Table 19.

The search of the EMA and FDA websites confirmed that all relevant studies for nintedanib and its comparators in IPF have been identified.

Figure 4: Study flow diagram for RCT study selection (PRISMA diagram)



4.2 *List of relevant randomised controlled trials*

The TOMORROW (Phase II), INPULSIS-1 and INPULSIS-2 (Phase III) studies are the only studies providing evidence of nintedanib in IPF, with evidence for these studies presented in this dossier extracted from the sources presented in Table 7 for TOMORROW, and Table 8 for the INPULSIS trials. For completeness, Table 9 presents those studies identified that contained nintedanib, but were excluded as they were either on-going studies, or did not present study data.

Table 7: Records pertaining to the TOMORROW Phase II study

Clinical Trial Reports (CTRs) for TOMORROW	
Clinical Trial Report (Brun *et al.*, 2012[16])	A 52 week, double blind, randomized, placebo-controlled trial evaluating the effect of BIBF 1120 administered at oral doses of 50 mg qd, 50 mg bid, 100 mg bid and 150 mg bid on Forced Vital Capacity decline during one year, in patients with Idiopathic Pulmonary Fibrosis, with optional active treatment extension until last patient out
Published articles	
Richeldi et al. 2011 (Richeldi *et al.*, 2011[1])	Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis
Clinical trials records	
NCT00514683 (1199.3) (ClinicalTrials.gov, 2007[28])	Safety And Efficacy of BIBF 1120 in Idiopathic Pulmonary Fibrosis
Conference abstracts	
Richeldi et al. 2011 (Richeldi, 2011[29])	The Oral Triple Kinase Inhibitor BIBF 1120 Reduces Decline In Lung Function In Patients With Idiopathic Pulmonary Fibrosis (IPF): Results From The Tomorrow Study American Journal of Respiratory and Critical Care Medicine 2011: A-5303
Richeldi et al. 2011 (Richeldi, 2011[30])	Treatment with BIBF 1120 reduces acute exacerbations and improves quality of life in patients with IPF: Results from the tomorrow study American Journal of Respiratory and Critical Care Medicine 2011: 183
Rangarajan et al. 2014 (Rangarajan *et al.*, 2014[31])	Non-Canonical Autophagy And The Anti-Fibrotic Effects Of BIBF-1120 In IPF American Thoracic Society 2014: A-5395

Table 8: Records pertaining to the INPULSIS Studies

CTRs for INPULSIS trials	
Clinical Trial Report 1199.32(Brun *et al.*, 2014[5])	A 52 week, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF)
Clinical Trial Report 1199.34 (Brun *et al.*, 2014[6])	A 52 week, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF)
Published articles	
Richeldi et al. 2014 (Richeldi, 2014[4])	Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis
Clinical trials records	
NCT01335477 (1199.34|2010-024252-29) (ClinicalTrials.gov, 2011[32])	Safety and Efficacy of BIBF 1120 at High Dose in Idiopathic Pulmonary Fibrosis Patients II
NCT01335464 (1199.32,2010-024251-87) (ClinicalTrials.gov, 2011[33])	Safety and Efficacy of BIBF 1120 at High Dose in Idiopathic Pulmonary Fibrosis Patients
Conference abstracts	
Richeldi et al. 2014 (Richeldi *et al.*, 2014[12])	Efficacy And Safety Of Nintedanib In Patients With Idiopathic Pulmonary Fibrosis: Results Of Two 52-Week, Phase III, Randomized, Placebo-Controlled Trials (INPULSIS 1 and 2) American Thoracic Society 2014: A-6603
Rangarajan et al. 2014 (Rangarajan *et al.*, 2014[31])	Non-Canonical Autophagy And The Anti-Fibrotic Effects Of BIBF-1120 In IPF American Thoracic Society 2014: A-5395

Table 9: Excluded studies that contained nintedanib

Study	**Nintedanib arm**	**Reason for exclusion**
Inoue et al. 2011 (Inoue, 2011[34])	BIBF 1120 alone or in combination with pirfenidone	Limited evidence available/pharmacokinetic study
NCT01136174 (ClinicalTrials.gov, 2010[35])	Nintedanib 50, 100, 150 mg	No data available
NCT01170065 (1199.35|2009-013788-21) (ClinicalTrials.gov, 2010[36])	Roll Over Study From 1199.30 BIBF 1120 in Idiopathic Pulmonary Fibrosis (IPF)	No data available
NCT01979952 (1199.187) (ClinicalTrials.gov, 2013[37])	Nintedanib 150 mg twice daily	Study in progress, no data
NCT01417156 (1199.4) Follow up Study From 1199.31 (NCT01136174) (ClinicalTrials.gov, 2011[38])	Nintedanib 50, 100, 150 mg	Study in progress, no data
NCT01619085 (1199.33|2011-002766-21) (ClinicalTrials.gov, 2012[39])	Nintedanib twice daily	Study in progress, no data

Nintedanib in IPF Clinical Development Programme Summary

The TOMORROW study was a 12 month phase IIb trial, assessing the efficacy and safety of four different oral doses of nintedanib. The INPULSIS trials were two replicate phase III trials of nintedanib in patients with idiopathic pulmonary fibrosis. Details of the TOMORROW and INPULSIS studies can be found in Table 10 and ongoing/extension studies are presented in Table 11.

Table 10: Relevant Randomised Controlled Studies

Trial name/phase	Population (key Inclusion Criteria)	Interventions/ comparator	Study Design	Duration	Primary study reference
TOMORROW/ Phase II	Patients aged ≥40 years with a diagnosis of IPF (received less than 5 years before screening), confirmed according to standard guidelines*	Nintedanib (50 mg once a day, 50 mg twice a day, 100 mg twice a day and 150 mg twice a day) vs. placebo.	Randomised, double-blind, placebo-controlled, phase II study assessing the efficacy and safety of four oral doses of nintedanib.	52 weeks	Richeldi et al. 2011
INPULSIS I and INPULSIS II/ Phase III	Patients aged ≥40 years with a diagnosis of IPF, confirmed according to standard guidelines*, within 5 years of randomisation.	Nintedanib (150 mg twice daily) vs. placebo	Replicate, randomised, double-blind, phase III study evaluating the efficacy and safety of 150 mg nintedanib twice daily	52 week treatment period plus 28 day follow up	Richeldi et al. 2014

Abbreviations: CSR: clinical study report, FVC: forced vital capacity, HRCT: High-resolution computerised tomography; *standard guidelines, refers to criteria published by the American Thoracic Society (ATS), the European Respiratory Society (ERS)

Table 11: Ongoing RCTs/Extension studies

Trial name	Population	Interventions	Primary study reference
TOMORROW	Adult patients with IPF	Nintedanib Placebo	No data reported
INPULSIS	Adult patients with IPF	Nintedanib Placebo	No data reported

All relevant RCTs, presented in Table 10, compare nintedanib with placebo. The INPULSIS-1, INPULSIS-2 and the TOMORROW studies are the only relevant RCTs to the decision problem, and are placebo controlled. There are no head-to-head trials of nintedanib vs. pirfenidone.

All identified and relevant studies have been included and contained within this submission.

Of the 5 included records, two abstracts referring to the TOMORROW study, which do not present additional data or analyses, are not regarded as a primary reference for

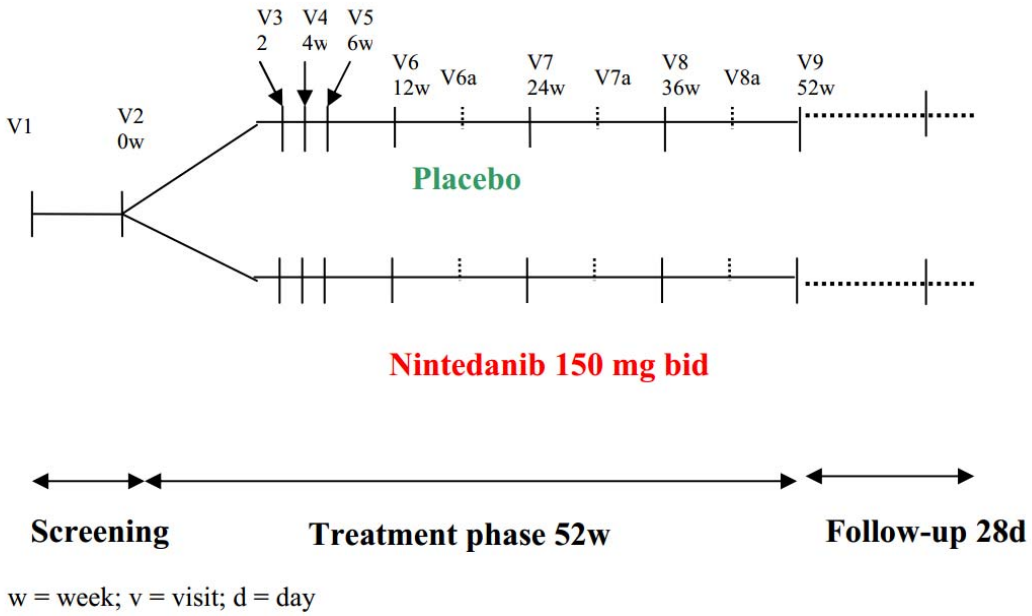
this study and are therefore excluded from further discussion. Evidence for the TOMORROW study is taken from Richeldi 2011 (Richeldi *et al.*, 2011[1]) and the CSR (Brun *et al.*, 2012[16]). One study investigating the pharmacokinetic parameters of pirfenidone alone or in combination with nintedanib (Inoue, 2011[34]) is also excluded from further discussion as it does not report on any of the outcomes relevant to the decision problem.

4.3 *Summary of methodology of the relevant randomised controlled trials*

The TOMORROW study was a randomised, double-blind, placebo-controlled phase II study performed at 92 sites in 25 countries. The diagnosis of IPF was confirmed at the screening visit which took place 4 days or more before randomisation. Patients were assigned to receive placebo capsules or one of the following doses of nintedanib: 50 mg once a day, 50 mg twice a day, 100 mg twice a day, or 150 mg twice a day. The trial used a dose-escalation scheme. The first cohort of patients was randomised to placebo or the lowest (50 mg qd) dose. The second cohort was randomized to placebo or the 50 mg qd or the 50 mg bid dose, etc. Any randomisation to the next higher dose treatment arm was only possible after a Data Monitoring committee (DMC decision on the safety of the previous treatment arms / previous cohort, and the recommendation to continue the trial. Study methodology is presented in Table 12.

The two INPULSIS trials were randomised, double-blind, placebo-controlled, phase III replicate studies performed at 205 sites in 24 countries. Following a screening period, eligible patients were randomly assigned (in a 3:2 ratio) to receive 150 mg nintedanib twice daily or placebo for 52 weeks as presented in Figure 5. Patients were allowed to receive certain background therapies (not including pirfenidone) in cases of deterioration after 6 months of study treatment or if there was an investigator reported acute exacerbation. Study methodology is presented in Table 12.

Figure 5: INPULSIS Study Design



The methodology of the TOMORROW and INPULSIS studies is summarised in Table 12 (Richeldi *et al.*, 2011[1], Richeldi *et al.*, 2014[12]):

Table 12: Summary of methodology of relevant RCTs

	TOMORROW	INPULSIS™-1 and INPULSIS™-2
Purpose	To evaluate the efficacy and safety of four different dose strategies of nintedanib treatment for 52 weeks compared to placebo in patients with IPF.	To evaluate the efficacy and safety of treatment with 150 mg of nintedanib, twice daily in patients with IPF.
Primary Endpoint	Annual rate of FVC decline	Annual rate of FVC decline.
Locations	Europe, Americas, Asia and Australia.	Europe, Americas, Asia and Australia.
Study design	Multicentre, randomised, double-blind, placebo-controlled, phase II study.	Multicentre, randomised, double-blind, placebo-controlled replicate phase III studies.
Duration	52 weeks	52 weeks plus 28 day follow up period
Timing of assessments	Assessment of outcomes was undertaken at randomisation, and at weeks 2, 4, 6, 12, 24, 36 and 52.	Assessment of outcomes was undertaken at randomisation and at 2, 4, 6, 12, 24, 36 and 52 weeks and then at the follow-up visit. SGRQ at randomisation and at 6, 12, 24 and 52 weeks
Method of Random-isation	The TOMORROW trial used a dose-escalation scheme. The first cohort was randomized to placebo or the lowest (50 mg qd) dose. The second cohort was randomized to placebo or the 50 mg qd or the 50 mg bid dose, etc. Any randomisation to the next higher dose treatment arm was only possible after a Data Monitoring Committee (DMC) decision on the safety of the previous treatment arms / previous cohort, and the recommendation to continue the trial. Randomisation was performed by phone using an Interactive Voice Response System (IVRS).	Following a screening period, eligible patients for the INPULSIS™ studies were randomly assigned, in a 3:2 ratio, to receive 150 mg nintedanib twice daily or placebo, for 52 weeks. The process of randomisation was undertaken using an interactive telephone and Web-based response system.
Masking	Double-blind Study medication was identified by a medication code number. Packaging and labelling were otherwise identical.	Double-blind
Intervention	Low dose nintedanib 50 mg, once daily Low dose nintedanib, 50 mg twice daily Intermediate dose nintedanib, 100 mg twice daily High dose nintedanib, 150 mg twice daily Patients received capsules orally.	Patients received capsules (orally), containing nintedanib 150mg twice daily (300mg) for 52 weeks. Patients were instructed to swallow the trial medication whole with a glass of water (~250 mL), and were to observe a dose interval of 12 hours. Trial medication was to be taken at the same time every day (between 06:00 and 11:00 in the morning, and between 18:00 and 23:00 in the evening). Dose interruption or reduction of the dose from 150 mg twice daily to 100 mg twice daily was permitted in order to manage adverse events. Upon resolution of an adverse event the dose was reinstated to 150 mg twice daily.
Comparator	Placebo	Placebo

	TOMORROW	INPULSIS™-1 and INPULSIS™-2
Previous/ existing treatment	Concomitant therapy with up to 15 mg of prednisone per day, or equivalent, was permitted if the dose had been stabilised for at least 8 weeks prior to screening.	Concomitant therapy with up to 15 mg of prednisone per day, or equivalent, was permitted if the dose had been stabilised for at least 8 weeks prior to screening. In cases of deterioration, after 6 months of study treatment patients could receive the following medications, at the discretion of the investigator: azathioprine, cyclophosphamide, cyclosporine, *N*-acetylcysteine, or an increased dose of prednisolone (>15 mg). At any time during the trial, in cases of investigator reported acute exacerbation any treatment (with the exception of pirfenidone) could be initiated or dose increased as deemed appropriate.
Duration of follow-up	52 weeks	52 week treatment period with a 4 week follow up visit.
Primary Outcomes (further discussed section 6.3.5)	Rate of decline in FVC (expressed in mL per year), evaluated from baseline until 12 months of treatment, compared to placebo.	Rate of decline in FVC (expressed in mL per year), evaluated from baseline until 12 months of treatment, compared to placebo.

	TOMORROW	INPULSIS™-1 and INPULSIS™-2
Secondary Outcomes	**FVC changes after 52 weeks of treatment presented as follows:** Change of % predicted FVC from baseline Change of absolute FVC value from baseline Percent change of absolute FVC value from baseline Percent change in FVC percent of predicted from baseline FVC changes from the absolute value at baseline at each time point and by following categories: Decrease of >10% or 200 ml Increase of >10% or 200 ml Change within ≤10% AND ≤200 mL **Survival (all causes of death and lung-transplant free)** **SpO2 (oxygen saturation) at rest** Change from baseline at each time point and at following categories: Decrease of >4% (absolute change from the percent value; e.g. decrease by 5% would be from SpO2= 95% to SpO2= 90%) Increase of >4% Change within ±4% **PaO2**, **PaCO2 and calculated P(A-a)O2** Change from baseline and at 52 weeks and by following categories (only for PaO2 and P(A-a)O2): Decrease of >4 mmHg Increase of >4 mmHg Change within ±4mmHg	**Key Secondary Outcomes:** Time to first acute exacerbation (investigator reported) Absolute change from baseline at 52 weeks, in SGRQ total score **Other Secondary Outcomes** **FVC changes after 52 weeks of treatment presented as follows:** Absolute and relative change from baseline in FVC (mL) and FVC% predicted over 52 weeks Absolute categorical change of FVC% predicted up to 52 weeks: decrease by >5%, increase by >5%, and change within ≤5% Absolute categorical change of FVC% predicted up to 52 weeks: decrease by >10%, increase by >10%, and change within ≤10% Proportion of FVC responders using 5% threshold at 52 weeks, defined as patients with absolute decline in FVC% predicted no greater than 5% and with an FVC evaluation at 52 weeks FVC responders using 10% threshold at 52 weeks, defined as patients with absolute decline in FVC% predicted no greater than 10% and with an FVC evaluation at 52 weeks **Patient reported outcomes** Change from baseline in SGRQ domains over 52 weeks (points) Change from baseline in IPF specific version of SGRQ (SGRQ-I) [R10-5500] total score (exploratory calculation from SGRQ data) over 52 weeks (points) Change from baseline in Shortness of Breath Questionnaire (SOBQ) over 52 weeks Change from baseline in cough impact and cough symptom of the Cough and Sputum Assessment Questionnaire (CASA-Q[CD]) score over 52 weeks Proportion of Patient's Global Impression of Change (PGI-C) responders at 52 weeks (defined as 'Very much better'/'Much better'/'A little better'/'No change') Change from baseline in EuroQol 5-Dimensional Quality of Life Questionnaire (EQ-5D) health state up to 52 weeks

	TOMORROW	INPULSIS™-1 and INPULSIS™-2
	DLCO Change from baseline at 6 months and 12 months and by following categories: Decrease of >15% (from the absolute value at baseline) or > 1 mmol·min-1·kPa-1 Increase of > 15% (from the absolute value at baseline) or > 1 mmol·min-1·kPa-1 Change within ≤15% (from the absolute value at baseline) AND ≤1 mmol·min-1·kPa-1 **6-Minute Walk Test** Change from baseline in distance walked (m) Dyspnoea rating on Borg scale: change from baseline. **Other spirometric parameters** FEV1/FVC actual change from baseline at 3 months, 6 months and 12 months **Patient reported outcomes** SGRQ change from baseline at 3, 6 and 12 months (total score domains) Medical Research Council (MRC) dyspnoea scale, change from baseline at 3 months, 6 months and 12 months. **Full Body Plethysmography** change from baseline at 6 months and 52 weeks Total lung capacity (TLC) Residual volume (RV) Total gas volume (TGV) Vital capacity (VC) Inspiratory capacity (IC) Exacerbations of IPF Incidence Number of occurrences per patient per year Time to first occurrence **Time to progression with progression defined with at least one of the following:** 5 mm Hg increase in the alveolo-arterial pressure difference in oxygen (P(A-a)O2) 10% decrease in FVC (FVC (baseline) – FVC (progression) ≥10%) **Death**	**Acute exacerbations** Proportion of patients with at least one acute exacerbation during the 52 weeks. Exacerbation days (defined by number of days with an exacerbation per patient) during the 52 weeks. Number of exacerbations during the 52 weeks. **Survival Analyses** Time to death over 52 weeks Time to death due to respiratory cause over 52 weeks (adjudicated) Time to on-treatment death Time to death or lung transplant over 52 weeks **Others:** Change from baseline in SpO2 (oxygen saturation, expressed in percent) at rest evaluated from baseline up to 52 weeks. Change from baseline in DLCO at rest at rest evaluated from baseline up to 52 weeks of treatment.

Patients were eligible to participate in the TOMORROW and INPULSIS trials if they were ≥40 years of age and had received a diagnosis of IPF, confirmed according to specified guidelines, within 5 years of randomisation. Additional eligibility criteria are presented in Table 13.

Table 13: Eligibility Criteria for TOMORROW and INPULSIS Studies

TOMORROW	INPULSIS™-1 and INPULSIS™-2
Inclusion Criteria	
Patient ≥ 40 years of age	Patient ≥ 40 years of age
Written informed consent, signed prior to entry into the study	Written informed consent, signed prior to entry into the study
IPF diagnosis (according to American Thoracic Society (ATS) / European Respiratory Society (ERS) criteria), within 5 years of screening visit.	IPF diagnosed (according to most recent ATS, ERS, Japanese Respiratory Society (JRS), Latin American Thoracic Association (ALAT) IPF guideline for diagnosis and management) within 5 years;
HRCT within 52 weeks of randomisation and lung biopsy (the latter if needed to fulfil ATS/ERS criteria) centrally reviewed and consistent with diagnosis	HRCT within 52 weeks of randomisation and lung biopsy (the latter if needed to fulfil ATS/ERS/JRS/ALAT criteria) centrally reviewed and consistent with diagnosis
FVC ≥50 % of predicted value predicted normal values were calculated according to ESCS: Males: FVC predicted (L) = 5.76 x height (meters) 0.026 x age (years) -4.34 Females: FVC predicted (L) = 4.43 x height (meters)- 0.026 x age (years) -2.89	FVC ≥50 % of predicted value predicted normal values were calculated according to ESCS: Males: FVC predicted (L) = 5.76 x height (meters) 0.026 x age (years) -4.34 Females: FVC predicted (L) = 4.43 x height (meters)- 0.026 x age (years) -2.89
Carbon monoxide diffusion capacity (DLCO) (corrected for haemoglobin [Hb] (visit 1)): 30-79% predicted of normal at randomisation.	Carbon monoxide diffusion capacity (DLCO) (corrected for haemoglobin [Hb] (visit 1)): 30-79% predicted of normal at randomisation.
PaO2 ≥55 mmHg (sea level to 1500 m) or50 mmHg (above 1500 m) room air Y	*PaO2 was not explicitly defined as part of the inclusion criteria for the INPULSIS trials.*
Exclusion Criteria	
AST, ALT > 1.5 x ULN	AST, ALT > 1.5 x ULN
Bilirubin > 1.5 x ULN	Bilirubin > 1.5 x ULN
Relevant airways obstruction (i.e. pre-bronchodilator FEV1/FVC < 0.7)	Relevant airways obstruction (i.e. pre-bronchodilator FEV1/FVC < 0.7)
Continuous oxygen supplementation at randomisation (defined as > 15 hours supplemental oxygen per day).	*No criteria based on oxygen supplementation*
Active infection at screening or randomisation.	*No criteria based on infection*
Neutrophils < 1500 / mm3	*No criteria based on neutrophils*
International normalised ratio (INR) > 1.5 and/or Partial thromboplastin time (PTT) > 1.5 x ULN ;	*No criteria based on INR*
Platelets < 100 000 /mL	*No criteria based on platelet count*
Haemoglobin < 9.0 g/dL	*No criteria based on haemoglobin levels*
In the opinion of the Investigator, patient is likely to have lung transplantation during study	In the opinion of the Investigator, patient is likely to have lung transplantation during study

TOMORROW	INPULSIS™-1 and INPULSIS™-2
Life expectancy for disease other than IPF < 2.5 years (Investigator assessment).	*No criteria based on life expectancy*
Other disease that may interfere with testing procedures or in judgement of the investigator may interfere with trial participation or may put the patient at risk when participating in this trial. Myocardial infarction during the previous 6 months Unstable angina during the previous month	Other disease that may interfere with testing procedures or in judgement of the investigator may interfere with trial participation or may put the patient at risk when participating in this trial. Myocardial infarction during the previous 6 months Unstable angina during the previous month
Other investigational therapy received within 8 weeks prior to screening visit.	N-acetylcysteine, prednisone > 15mg/day or equivalent received within 2 weeks of visit 1; pirfenidone, azathioprine, cyclophosphamide, cyclosporine A received within 8 weeks of visit 1;
Pregnant women or women who are breast feeding or of child bearing potential not using a highly effective method of birth control for at least one month prior to enrolment.	Pregnant women or women who are breast feeding or of child bearing potential not using a highly effective method of birth control for at least one month prior to enrolment.
Sexually active males not committing to taking precaution during the course of the study (except where their partner is not of childbearing potential).	Sexually active males not committing to taking precaution during the course of the study (except where their partner is not of childbearing potential).
Known or suspected active alcohol or drug abuse.	Known or suspected active alcohol or drug abuse.
Bleeding risk : Known inherited predisposition to bleeding, patients who require full-dose anticoagulation, patients who require full-dose antiplatelet therapy, history of hemorrhagic CNS event within 12 months prior to screening , any of the following within 3 months prior to screening : gross / frank haemoptysis or haematuria, active gastro-intestinal bleeding or ulcers, major injury or surgery	Bleeding risk : Known inherited predisposition to bleeding, patients who require full-dose anticoagulation, patients who require full-dose antiplatelet therapy, history of hemorrhagic CNS event within 12 months prior to screening , any of the following within 3 months prior to screening : gross / frank haemoptysis or haematuria, active gastro-intestinal bleeding or ulcers, major injury or surgery
Thrombotic risk	Thrombotic risk
Surgical procedures planned to occur during trial period.	Surgical procedures planned to occur during trial period.
Coagulopathy	Coagulopathy
Uncontrolled systemic arterial hypertension	*No criteria based on arterial hypertension*
Known hypersensitivity to lactose or any component of the study medication	*No criteria based on lactose*

4.4 *Statistical analysis and definition of study groups in the relevant randomised controlled trials*

TOMORROW:

The primary efficacy endpoint was the annual rate of decline in FVC and was calculated using all FVC assessments from baseline to 52-weeks using a mixed linear regression model.

A secondary objective was to evaluate the treatment effect on patient's survival. All patients in active arms were encouraged to remain under their randomized treatment until every evaluable entered patient had either completed visit 9 or prematurely discontinued.

The objective of the study was to demonstrate the superiority of at least one dose strategy of nintedanib compared to placebo in IPF patients. This was tested by using the four sets of hypotheses indicated below:

H_{0i}: RD(Ti) = RD(P)

H_{1i}: RD(Ti) ≠ RD(P)

Where RD(Ti) is the mean rate of FVC decline for nintedanib and RD(P) is the mean rate of FVC decline for placebo, and i indicates the nintedanib target dose (50mg qd, 50 mg bid, 100mg bid, and 150 mg bid.

The analysis on the efficacy and safety endpoints was based on the following populations:

Randomised set (RS): This patient set includes all randomised patients whether treated or not

Treated set (TS): This patient set includes all patients who were dispensed study medication and were documented to have taken at least one dose of the investigational treatment

Full analysis set with at least 50% of the intended doses (FAS50): Includes all patients who were randomized, assessed at least at baseline and once post-baseline after at least 3 months continuous study treatment for the primary efficacy endpoint with compliance being of at least 50% during the treatment period.

The primary analysis used the assumption that the decrease of FVC was linear within each subject along the study, and gender, age and height were included as covariates. No replacement of missing data was planned.

INPULSIS:

The primary trial endpoint was the annual rate of decline in FVC estimated from readings over 52 weeks and analysed using random coefficient regression (random slopes and intercepts) model including gender, age and height as covariates.

The key secondary endpoint, change from baseline in the SGRQ total score at 52 weeks, was analysed using Mixed Effects Model for Repeated Measures (MMRM) analysis including treatment and visit as fixed effects and baseline SGRQ total score as a covariate, treatment-by-visit and baseline SGRQ total score-by-visit as interaction terms. The other key additional secondary endpoint, time to first acute IPF exacerbation, was analysed using the Cox's proportional hazards model, adjusted for gender, height, and age and a log-rank test.

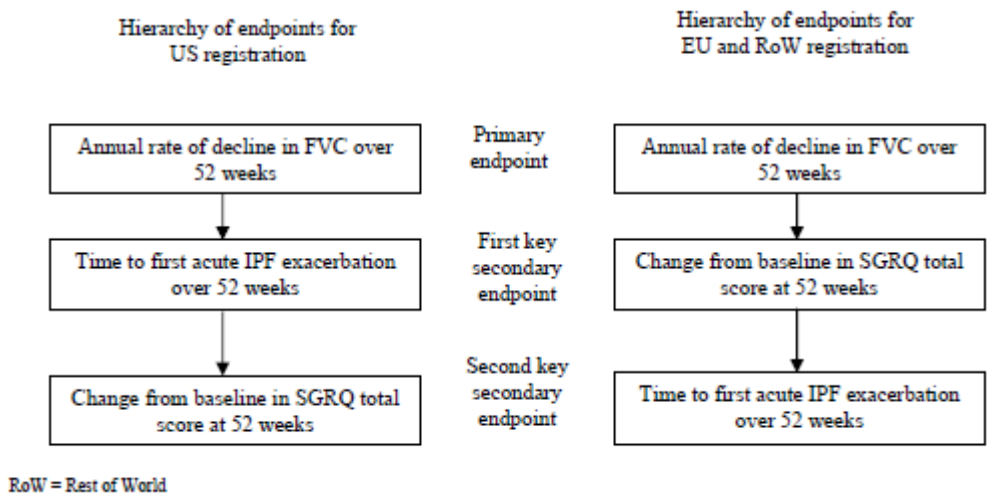
The objective of the trial was to assess the superiority of nintedanib 150mg bid compared to placebo on the annual rate of decline in FVC and was tested using the following set of hypotheses:

H_0: No difference in the annual rate of decline in FVC evaluated from baseline until 52 weeks between nintedanib 150 mg bid and placebo.

H_A: The annual rate of decline in FVC is smaller for patients taking nintedanib 150 mg bid than those taking placebo over 52 weeks.

A hierarchical procedure was used in order to demonstrate the superiority of nintedanib over placebo for the primary and 2 key secondary endpoints. The consecutive steps were considered only if the previous step was significant at the 1-sided 2.5% level and the results were in favour of nintedanib. Two alternative hierarchies with different ordering of the key secondary endpoints were tested for the US and EU regulatory submissions; these are presented in Figure 6 below.

Figure 6: Hierarchy of endpoints tested



All data, from baseline to follow-up visit were included in the primary analysis. Follow-up visit data were excluded, except for patients who prematurely discontinued trial medication and did not complete the planned observation time. FVC decline was assumed to be linear for each patient over 52 weeks. However, this was explored further in sensitivity analyses where quadratic and exponential models were used and checked by graphical displays and statistical tests on the residuals.

For sensitivity to data handling assumptions, the following models were performed:

- Primary model including only on-treatment evaluations (up to last trial drug intake + 1 day)
- Primary model including data collected post-lung transplant

Further sensitivity analyses including multiple imputation approaches to investigate missing data effects. Patients were classified into four distinct groups depending on data availability:

- Patients with a 52-week FVC value who received trial drug until week 52;
- Patients with a 52-week FVC value who prematurely discontinued trial drug but were followed up until Week-52;
- Patients without a 52-week FVC value who were alive at Week 52;
- Patients without a 52-week FVC value who died before Week 52.

Missing data from other visits prior to week 52 or non-monotone missing data were not imputed. Multiple imputation was used to handle missing data at week 52.

A summary of the statistical analyses undertaken is provided in Table 14.

Table 14: Summary of Statistical Analyses in the RCTs

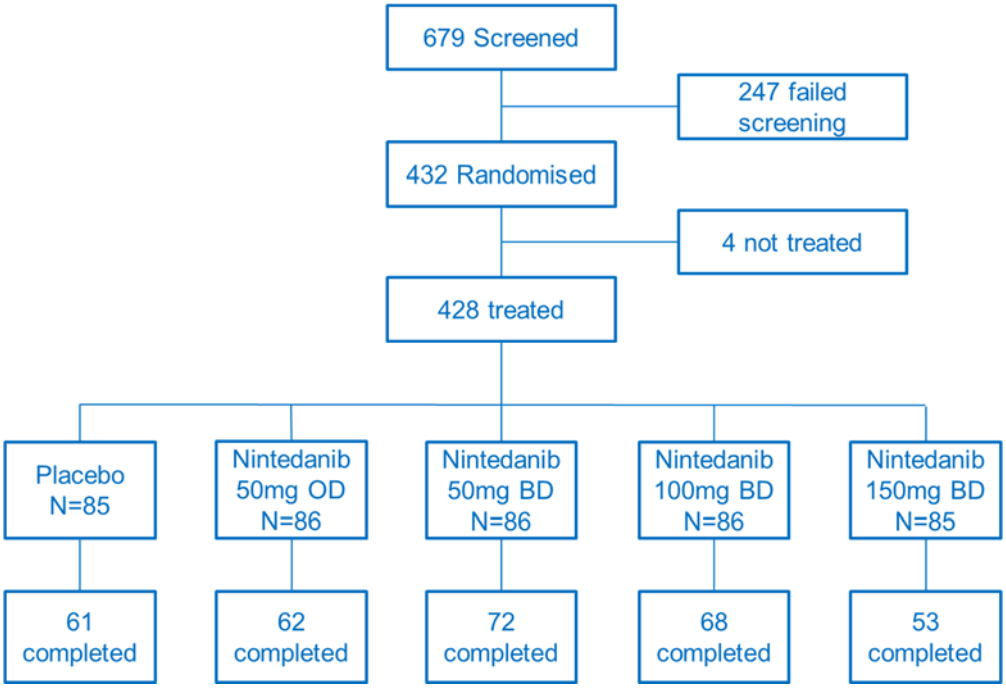
Trial no. (acronym)	Hypothesis objective	Statistical analysis	Sample size, power calculation	Data management, patient withdrawals
TOMORROW	All efficacy analyses were based on the randomised set of 432 patients on an intention to treat basis. Only on-treatment measurements (data collected up to 1 day after the last dose was received) were included in the primary efficacy analyses). Patients were evaluated according to randomisation at the study initiation.	The decrease in FVC over time was calculated by modelling a linear decrease for every patient from the date on which the first dose of the assigned study drug was taken to the date of the last measurement obtained during the study, with all data taken into consideration. Analysis was performed using a closed testing procedure for multiplicity correction and a hierarchical testing procedure.	This was calculated to provide 80% power to detect a difference of 0.1 litres in the annual decrease in FVC between patients receiving nintedanib and those receiving placebo.	In the evaluation of secondary endpoints when data for the complete 52 week assessment period were missing, the last observation carried forward approach was applied.
INPULSIS 1 and 2	• For both INPULSIS trials the null and alternative hypotheses were as follows: • Hn: There is no difference in the annual rate of decline in FVC evaluated from baseline until 52 weeks between nintedanib 150 mg bid and placebo. • Ha: The annual rate of decline in FVC is smaller for patients taking nintedanib 150 mg bid than those taking placebo over 52 weeks.	Analyses of the primary endpoint was performed using a random coefficient regression model (with random slopes and intercepts) that included sex, age and height as covariates. The treatment effect was determined by using estimated slopes for each study group (on the basis of the time-by-treatment interaction term from the mixed model).	For each of the INPULSIS trials, the sample size was calculated to provide 90% power in order to detect a between-group difference of 100 ml in the annual rate of FVC decline. On the basis of evidence from the phase II study (Richeldi 2011), it was assumed that the standard deviation for the change in FVC from baseline was 300 ml for both groups. Assuming that it would not be possible to evaluate data for 2% of patients, the sample size was calculated as 194 patients in the placebo arm and 291 patients in the nintedanib arm for a two-group t-test at a one-sided significance level of 2.5%. Since the primary analysis was based on a random coefficient regression model with functionality for adjustment of several variables and considered information over time it was expected that the power would be greater than the 90% calculated for the t-test.	Regression model data: All available FVC values from baseline to week 52 were used in the primary model. The statistical model used for the primary analysis allowed for missing data with the assumption that they were missing at random; missing data were not imputed for the primary analysis but data collected following discontinuation of the study drug were included in the analysis. Significance tests were two-sided, with an alpha value of 0.05. The superiority of nintedanib over placebo for the primary and key secondary endpoints was tested using a hierarchical procedure to account for multiple comparisons. Sensitivity analyses were performed to test the robustness of the results for the primary and key secondary endpoints. Multiple imputation sensitivity analyses were conducted to assess the effect of missing data and provide an estimate of the treatment effect under different assumptions.

4.5 *Participant flow in the relevant randomised controlled trials*

TOMORROW:

In TOMORROW, a total of 432 patients with confirmed IPF diagnosis were included. The trial used a dose-escalation scheme, with the first cohort randomized to placebo or the lowest (50mg qd) dose. The second cohort was randomized to placebo or the 50mg qd or the 50mg bid dose, etc. Any subsequent randomisation to a higher dose treatment arm was only possible after the Data Monitoring Committee decision on the safety of the previous treatment arms/previous cohort and the recommendation to continue the trial. Randomisation was performed by phone using an Interactive Voice Response System (IVRS). All patients had the opportunity for a one step dose reduction in case of intolerance. A CONSORT flow chart presenting the patient disposition in the TOMORROW study is shown in Figure 7 (Brun *et al.*, 2012[16]).

Figure 7: Patient Flow diagram for the TOMORROW Study



INPULSIS 1 and 2:

Patient disposition for the INPULSIS-1 and INPULSIS-2 trials are shown in Figure 8 and Figure 9, respectively (Brun *et al.*, 2014[5], Brun *et al.*, 2014[6]). A total of 1,066 patients were enrolled into the INPULSIS trials across 205 sites across 24 countries, including seven sites in England and one in Scotland. The number of patients randomized in each country ranged from 1 to 56 patients per treatment group. Figure 8 and Figure 9 present the patient flow from the INPULSIS trials.

Figure 8: Patient flow for INPULSIS-1

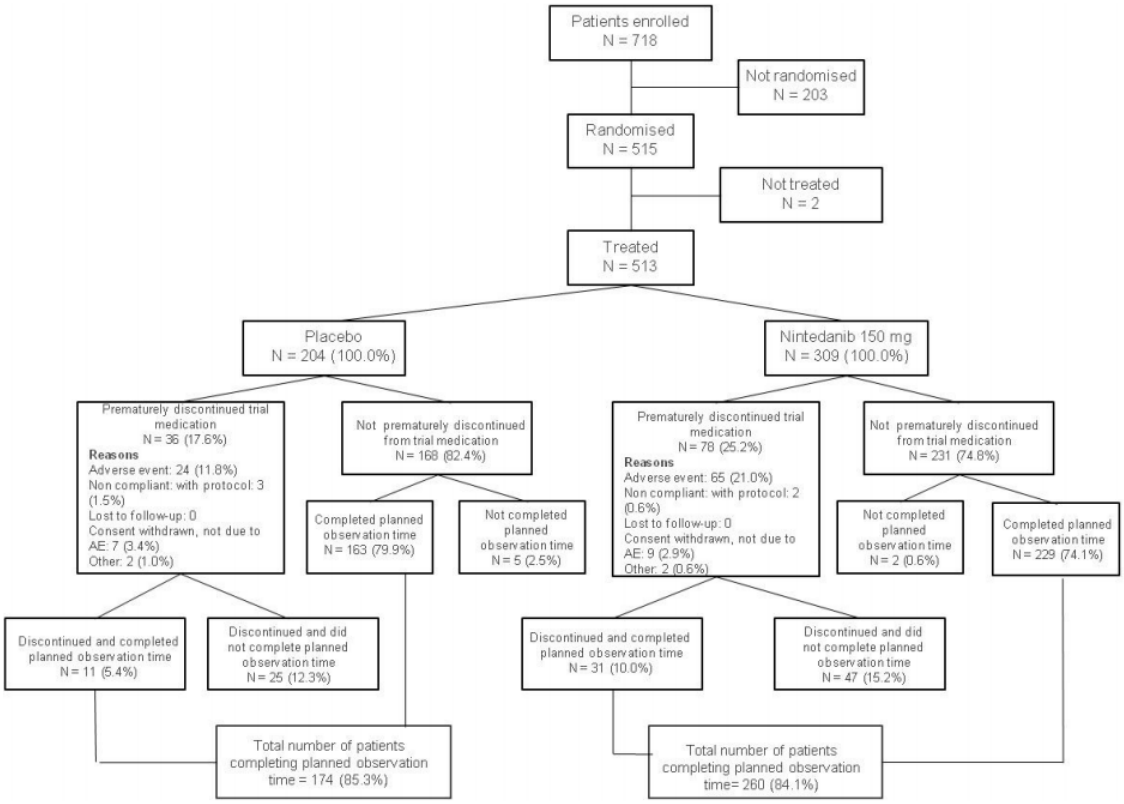
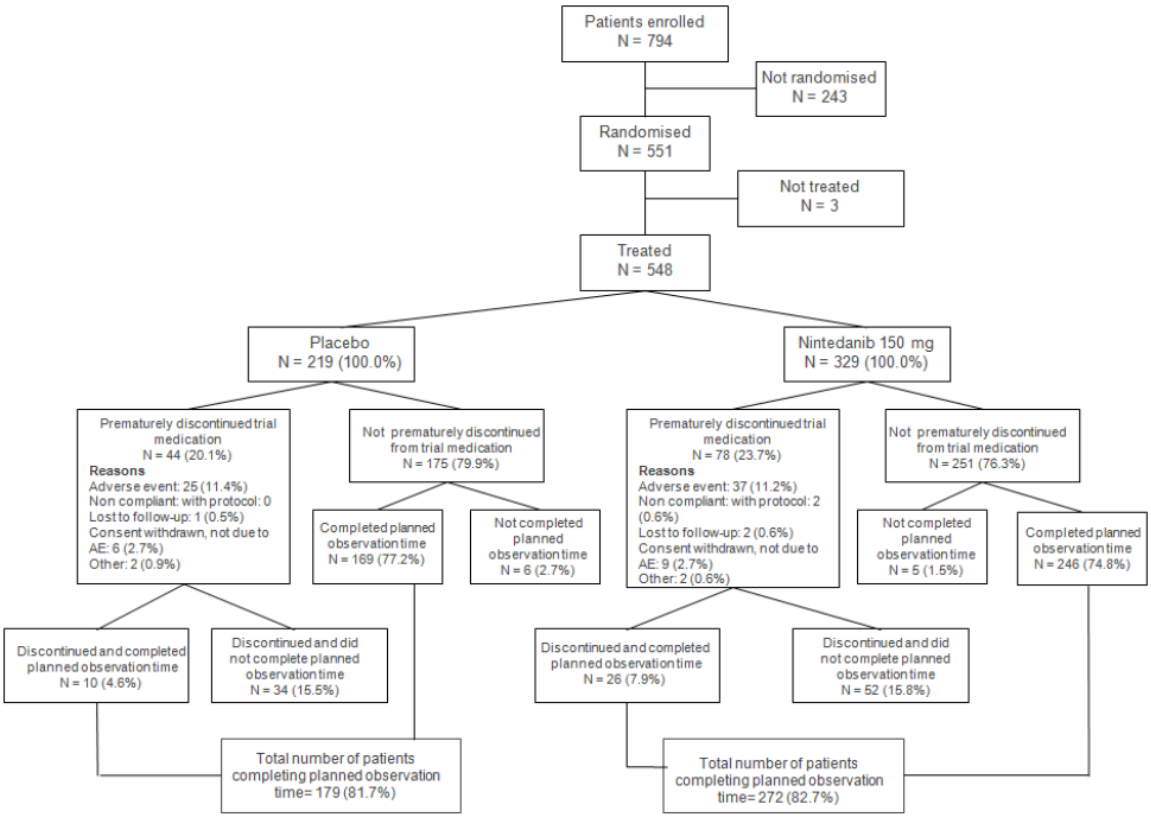


Figure 9: Patient flow for INPULSIS-2



4.6 *Quality assessment of the relevant randomised controlled trials*

Critical appraisal of the RCTs for nintedanib is presented in Table 15.

For TOMORROW, randomisation was conducted by phone using an Interactive Voice Response System (IVRS). Allocation concealment was carried out using an interactive voice response system. Baseline characteristics were similar in all treatment arms. Patients, investigators, and the study sponsor were blind to the study-group assignments throughout the study. There was no imbalance between dropouts across the study groups apart from the highest number of discontinuations occurring in the group receiving the highest dose of BIBF 1120. All pre-specified outcomes appeared to have been reported. All randomised patients were included in the intent-to-treat population and appropriate methods to account for missing data were used.

For INPULSIS, the method of randomisation and allocation concealment were consistent with standard practice (interactive voice response system used). Baseline characteristics were similar in all treatment arms. There were no unexpected imbalances found in the INPULSIS trials. Patients, investigators, and the study sponsor were blind to the study-group assignments throughout the studies. All pre-specified outcomes appear to have been reported. All randomised patients were included in the intent-to-treat population and appropriate methods to account for missing data were used.

Table 15: Critical Appraisal of the TOMORROW and INPULSIS studies

	TOMORROW		INPULSIS	
	Reviewer's judgment	**Justification**	**Reviewer's judgment**	**Justification**
Was randomisation carried out appropriately?	Yes	Randomisation conducted by phone using an Interactive Voice Response System.	Yes	The investigator describes a random component in the sequence generation: interactive telephone and web-based response system were used.
Was the concealment of treatment allocation adequate?	Yes	Participants and investigators enrolling participants could not foresee assignment because an interactive telephone and web-based response system were used.	Yes	Participants and investigators enrolling participants could not foresee assignment because an interactive telephone and web-based response system were used.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Participants had similar characteristics at baseline; treatment arms were well balanced.	Yes	Participants had similar characteristics at baseline; treatment arms were well balanced.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Blinding of participants and key study personnel (investigators and study sponsor) were blinded to the study-group assignments throughout of the study.	Yes	Blinding of participants and key study personnel (investigators and study sponsor) were blinded to the study-group assignments throughout of the study.
Were there any unexpected imbalances in drop-outs between groups?	No	Although there were some differences, those were unlikely to be unexpected. Highest number of discontinuations occurring in the group receiving the highest dose of BIBF 1120.	No	Although there were some differences, those were unlikely to be unexpected. There was slightly higher discontinuation in the nintedanib groups compared to placebo.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	All pre-specified outcomes appear to have been reported.	No	All pre-specified outcomes appear to have been reported.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes/Yes	All efficacy analyses were based on the intent-to-treat population and appropriate: last observation carried forward was used to account for missing data.	Yes/Yes	Analysis was performed on the intention-to-treat population. To minimize the amount of missing data, patients who discontinued the study drug prematurely were asked to attend all scheduled visits and to undergo all examinations as originally planned. For patients who discontinued the drug prematurely but did not agree to attend all visits, data on vital status were collected at week 52.

4.7 *Clinical effectiveness results of the relevant randomised controlled trials*

Summary

There are no head-to-head trials investigating nintedanib against the comparators specified in the decision problem. The INPULSIS studies, initiated in April 2011, were placebo-controlled. Pirfenidone was not a viable choice for a comparator in the INPULSIS trials, as it did not receive positive approval from the FDA until October 2014, 1 year after the completion of the INPULSIS trials. While it is recognised by international guidelines that there is a need for new therapies and randomised clinical trials for IPF, the guidelines do not specify any treatment as the standard of care (Raghu *et al.*, 2011[13]).

Nintedanib is the first targeted treatment for IPF to demonstrate efficacy in two identically designed Phase III trials and reach its primary endpoint in both studies. The phase III evidence for nintedanib confirmed the trend of efficacy observed in the phase II trial. Pirfenidone, studied in the concurrent CAPACITY studies (Noble, 2011[40]), failed to reach its primary endpoint in one of these studies, but went on to meet its primary endpoint in a third RCT (ASCEND) (King *et al.*, 2014[41]).

There is evidence from INPULSIS-1 and INPULSIS-2, to support the claim that nintedanib provides a clinically meaningful benefit for IPF patients, as demonstrated by the annual rate of change in FVC compared to placebo. A reduction in the rate of decline in FVC is considered consistent with the slowing of disease progression (Wells, 2013[42]). The average IPF patient has a lung function loss, measured by forced vital capacity (FVC), of 150–200mL per year compared to an elderly patient without IPF who would be expected to decline by 30-60 mL per year (Alexeeff *et al.*, 2007[43], Griffith *et al.*, 2001[44]). Nintedanib reduced the decline in FVC by 50%, when compared to placebo over 52 weeks:

- INPULSIS-1: -114.7mL (nintedanib) vs. -239.9mL (placebo);
- INPULSIS-2: -113.6mL (nintedanib) vs. -207.3mL (placebo).

In the pooled analysis of the INPULSIS trials nintedanib showed a non-significant increase in the time to first investigator reported acute exacerbation and a significant increase to the time to first adjudicated exacerbation. Acute exacerbations in IPF are associated with a high mortality and morbidity. These results indicate that nintedanib provides a benefit to IPF patients with high unmet need who currently have limited treatment options.

The INPULSIS studies confirm the efficacy and safety of nintedanib over a 52 week period as defined by the pre-specified endpoints and although the trend suggests that the improvement imparted to IPF patients will be sustained beyond one year, evidence from the on-going extension trial will better inform this question.

Baseline patient characteristics

Patient characteristics at baseline for the TOMORROW study and INPULSIS studies are summarised in Table 16 and Table 17, respectively. Baseline characteristics, including age, gender, time since diagnosis of IPF and key outcome measurements were similar across treatment groups.

Table 16: Characteristics of participants in the TOMORROW studies

Baseline characteristic	Placebo (N=85)	Nintedanib				Total (N=428)
		50 mg Once daily (N=86)	50 mg Twice daily (N=86)	100 mg Twice Daily (N=86)	150 mg Twice Daily (N=85)	
Male, no. (%)	63 (74.1)	65 (75.6)	62 (72.1)	65 (75.6)	65 (76.5)	320 (74.8)
Age in years, mean (SD)	64.8 (8.6)	65.3 (9.4)	64.9 (8.5)	65.1 (8.6)	65.4 (7.8)	65.1 (8.6)
BMI, mean (SD)	27.40 (3.8)	27.66 (3.7)	28.14 (4.3)	27.16 (3.4)	26.43 (3.9)	27.36 (3.9)
Weight in Kg, mean (SD)	77.3 (13.3)	78.8 (13.4)	79.0 (16.2)	76.0 (14.5)	74.9 (14.6)	77.2 (14.5)
Smoking Status						
Never smoked	27 (31.4)	29 (33.7)	32 (37.2)	25 (29.4)	28 (32.9)	27 (31.4)
Former smoker	55 (64.0)	53 (61.6)	52 (60.5)	58 (68.2)	51 (60.0)	55 (64.0)
Current smoker	4 (4.7)	4 (4.7)	2 (2.3)	2 (2.4)	6 (7.1)	4 (4.7)
Time since IPF diagnosis, years, mean (SD)	1.4 (1.5)	1.4 (1.3)	1.1 (1.2)	1.2 (1.2)	1.0 (1.2)	1.2 (1.3)
Specimen from surgical lung biopsy available, no. (%)	19 (22.4)	25 (29.1)	27 (31.4)	20 (23.3)	29 (34.1)	120 (28.0)
FVC (L)						
Mean (SD)	2.8 (0.8)	2.8 (0.8)	2.7 (0.7)	2.9 (0.8)	2.7 (0.8)	2.8 (0.8)
Median FVC (L)	2.7	2.8	2.7	2.8	2.7	2.8
Percentage of predicted value,	77.6	79.8	80.4	83.0	78.1	80.2
Median DLCO-mmol/min/kPa,	3.7	3.5	3.6	3.7	3.5	3.6
Median PaO2-mm Hg	75.0	75.8	78.4	80.0	78.3	77.5
Median SpO2 %, median	96.0	96.0	96.0	96.0	96.0	96.0

Table 17: Characteristics of participants in the INPULSIS studies

Baseline characteristic	INPULSIS-1		INPULSIS-2	
	Nintedanib (N=309)	Placebo (N=204)	Nintedanib (N=329)	Placebo (N=219)
Male, no. (%)	251 (81.2)	163 (79.9)	256 (77.8)	171 (78.1)
Age in years, mean (SD)	66.9 (8.4)	66.9 (8.2)	66.4 (7.9)	67.1 (7.5)
BMI in kg/m^2, mean (SD)	28.6 (4.5)	28.1 (4.6)	27.6 (4.6)	27.2 (4.5)
Weight in Kg, mean (SD)	82.0 (16.8)	81.2 (16.3)	76.6 (15.9)	76.3 (16.5)
Smoking status, no. (%)				
Never smoked	71 (23.0)	51 (25.0)	103 (31.3)	71 (32.4)
Former smoker	217 (70.2)	144 (70.6)	218 (66.3)	139 (63.5)
Current smoker	21 (6.8)	9 (4.4)	8 (2.4)	9 (4.1)
Time since IPF diagnosis, years, mean (SD)	1.7 (1.4)	1.6 (1.4)	1.6 (1.3)	1.6 (1.3)
Specimen from surgical lung biopsy available, no. (%)	60 (19.4)	33 (16.2)	84 (25.5)	52 (23.7)
Systemic corticosteroid therapy, no. (%)*	68 (22.0)	43 (21.1)	68 (20.7)	46 (21.0)
FVC				
Mean-ml, (SD)	2757 (735)	2845 (820)	2673 (776)	2619 (787)
Median-ml	2700	2721	2615	2591
Percentage of predicted value, mean (SD)	79.5 (17.0)	80.5 (17.3)	80.0 (18.1)	78.1 (19.0)
FEV_1:FVC (%)	81.5 (5.4)	80.8 (6.1)	81.8 (6.3)	82.4 (5.7)
DLco				
mmol/min/kPa, mean (SD)	4.0 (1.2)	4.0 (1.1)	3.8 (1.2)	3.7 (1.3)
Percentage of predicted value, mean (SD)	47.8 (12.3)	47.5 (11.7)	47.0 (14.5)	46.4 (14.8)
SpO2 %, mean (SD)	95.9 (2.0)	95.9 (1.9)	95.8 (2.6)	95.7 (2.1)
Total SGRQ score$	39.6 (17.6)	39.8 (18.5)	39.5 (20.5)	39.4 (18.7)

Abbreviations: FVC, forced vital capacity; FEV1, Forced Expected Volume; DLco: diffusion capacity of the lung for carbon monoxide; SD, Standard deviation; IPF, idiopathic pulmonary fibrosis; SGRQ, St. George's Respiratory Questionnaire

*Prednisone at a dose of no more than 15 mg per day or the equivalent was permitted if the dose had been stable for at least 8 weeks before screening. § The percentage of the predicted value for the diffusion capacity of the lung for carbon monoxide (DLCO) was calculated with the use of the equation described by the European Community for Steel and Coal in Cotes et al. 1993. In INPULSIS-2 data were available for 218 patients in the placebo group. $ In INPULSIS-1, the total score on the St. George's Respiratory Questionnaire (SGRQ) was available for 298 patients in the nintedanib group and 202 patients in the placebo group; in INPULSIS -2, the total SGRQ score was available for 326 patients in the nintedanib group and 217 patients in the placebo group. The total score ranges from 0 to 100, with higher scores indicating worse HRQL.

Efficacy results

The efficacy results for the outcomes in the TOMORROW study are summarised in Table 18 and for the INPULSIS studies in Table 19. The 150 mg dose in the TOMORROW study showed a non-significant reduction in the rate of decline of FVC vs. placebo using the pre-specified primary analysis method of a closed testing procedure for multiplicity correction (P=0.06); a statistically significant reduction was seen using the pre-specified alternate hierarchical testing procedure. A significant difference between 150 mg of nintedanib twice daily and placebo was seen in some clinically important secondary outcomes with a smaller percentage of patients having an acute exacerbation and a significantly improved SGRQ score in the nintedanib arm. There was no significant difference in terms of the primary outcome between placebo and the lower dose of nintedanib investigated irrespective of the analysis used, and hence these doses were not further investigated.

The INPULSIS studies showed a statistically significant reduction in rate of FVC decline, both when the studies were analysed individually and when a pre-specified analysis of the pooled data was undertaken. In the pooled analysis of investigator reported acute exacerbations there was a non-significant increase in time to first exacerbations, in the pre-specified pooled sensitivity analysis of time to first independently adjudicated exacerbation there was a significant increase in time to first exacerbation.

INPULSIS 1 and INPULSIS 2 showed a no significant difference, and a significant difference in favour of nintedanib respectively for the key secondary outcome of total change in SGRQ from baseline. There was a non-significant difference in favour of nintedanib on pooled analysis. 5 tools for assessing patient reported outcomes were used (SGRQ, UCSD-SOBQ, PGI-C, CASA-Q cough score, and EQ-5D). All instruments used produced similar results (i.e. no significant difference, but a consistent numerical advantage for nintedanib).

The INPULSIS trials showed a non-significant trend for reduced all-cause mortality. A pooled analysis of mortality data from the INPULSIS and TOMORROW trials showed:

- a numerical reduction in all-cause mortality with nintedanib vs. placebo (HR 0.70 [95% CI: 0.46, 1.08]; p=0.0954). The proportion of patients who died was 5.8% in the nintedanib group vs. 8.3% in the placebo group.
- a numerical risk reduction in respiratory mortality with nintedanib vs. placebo (HR 0.62 [95% CI: 0.37, 1.06]; p=0.0779). The proportion of patients who died due to respiratory cause was 3.6% in the nintedanib group vs. 5.7% in the placebo group.
- a **significant reduction** in on-treatment mortality vs. placebo (HR 0.57 [95% CI: 0.34, 0.97]; p=0.0274). The proportion of patients who died was 3.5% in the nintedanib group vs. 6.7% in the placebo group.

Table 18: Summary of clinical outcomes from TOMORROW Study

Clinical endpoint	**Nintedanib 50 mg once a day (N=87)**	**Nintedanib 50 mg twice a day (N=86)**	**Nintedanib 100 mg twice a day (N=86)**	**Nintedanib 150 mg twice a day (N=86)**	**Placebo (N=87)**
Absolute change in FVC at 52 weeks, mean (SE)	-0.18 (0.04)	-0.19 (0.04)	-0.13 (0.04)	-0.06 (0.04) P value: <0.01	-0.23 (0.04)
Patients with reduction in mean FVC of >10% or 200mL (n)	35 (41.2)	41 (47.7)	30 (35.3)	20, (23.8) P value: <0.05	37 (44)
Annual Rate of Decline in FVC (ml/year)	-0.17 (0.04)	-0.21 (0.04)	-0.16 (0.04)	-0.06, (0.04) P value: <0.05	-0.19 (0.04)
Absolute change in FVC% predicted	-4.58 (1.03)	-4.9, (0.98)	-3.15, (1)	-1.04, (0.99)	-6 (1.02)
Absolute change in worst SpO2 during 6MWT from baseline % (SE)	-0.86% (0.38%)	-0.97% (0.36%)	0.06% (0.36%)	-0.18%, (0.36%)	-1.29% (0.37%)
Absolute change in DLco*	-0.422 (0.1155) -	-0.629 (0.1173)	-0.542 (0.0995)	-0.609 (0.1034)	-0.511 (0.1035)
Number (%) with at least one exacerbation	10 (11.5%)	10 (11.6%)	6 (7.0%)	2 (2.3%)	12 (13.8%)
SGRQ score (change from baseline vs. placebo)	-0.79 (-5.22 – 3.64) p=0.7250	-3.28 (-7.63 – 1.07) p=0.1389	-3.98 (-8.35 – 0.39) p=0.0741	-6.12 (-10.57 - -1.67) p=0.0071	NA
Mortality	11 (12.6%)	3 (3.5%)	4 (4.7%)	7 (8.1%)	9 (10.3%)

Table 19: Summary of clinical outcomes for INPULSIS Studies

Clinical endpoint	INPULSIS-1			INPULSIS-2		
	Nintedanib (N Rnd =309)	**Placebo (N Rnd = 206)**	**MD/OR/HR (95% CI) p-value**	**Nintedanib (N Rnd =331)**	**Placebo (N Rnd = 220)**	**MD/OR/HR (95% CI) p-value**
Annual rate of decline in FVC (mL/yr)	-114.7	-239.9	MD: 125.3 (77.7 – 172.8) P<0.001	-113.6	-207.3	MD: 93.7 (44.8 – 142.7) P<0.001
Adjusted absolute mean change from baseline FVC (mL)	-95.1	205	MD: 109.9 (71.3 to 148.6) p<0.001	95.3	-205	MD: 109.8 (70.9 to 148.6) p<0.001
Adjusted absolute mean change from baseline in FVC - % of predicted value	-2.8	-6.0	MD: 3.2 (2.1 to 4.3) p<0.001	-3.1	-6.2	MD: 3.1 (1.9 to 4.3) p<0.001
Patients (%) with an FVC decline ≤5 percentage points	163 (52.5%)	78 (38.2%)	OR: 1.85 (1.28 to 2.66) p=0.001	175 (53.2%)	86 (39.3%)	OR: 1.79 (1.26 to 2.55) p=0.001
Patients (%) with an FVC decline ≤10 percentage points	218 (70.6%)	116 (56.9%)	OR: 1.91 (1.32 to 2.79) p<0.001	229 (69.6%)	140 (63.9%)	OR: 1.29 (0.89 to 1.86) p=0.18
Proportion of patients with ≥1 investigator reported acute exacerbations in 52 weeks	19 (6.1%)	11 (5.4%)	HR: 1.15 (0.54 to 2.42) p=0.673	12 (3.6%)	21 (9.6%)	HR: 0.38 (0.19 to 0.77) p=0.005
Adjudicated acute exacerbations	7 (2.3%)	8 (3.9%)	HR: 0.55 (0.20 to 1.54)	5 (1.5%)	16 (7.3%)	HR: 0.20 (0.07 to 0.56)
SGRQ score (change from baseline)	4.34	4.39	MD: -0.05 (-2.50 – 2.40) p=0.97	2.80	5.48	MD: -2.69 (-4.95 - -0.43) p=0.02
Mortality (%)	13 (4.2%)	13 (6.4%)	HR: 0.63 (0.29 to 1.36)	22 (6.7%)	20 (9.1%)	HR: 0.74 (0.40 to 1.35)
Absolute change from baseline in SpO2 over 52 weeks (%)	-0.24%	-0.53%	MD: 0.29% (95% CI: -0.07, 0.64; p = 0.1138)	-0.39%	-0.66%	MD 0.27% (95% CI: -0.15, 0.69; p = 0.2032)
Absolute change from baseline in DLCO over 52 weeks (mmol/min/kPa)	-0.380	-0.365	MD: -0.015 [95% CI: -0.191, 0.161; p = 0.8650]	-0.286 for	-0.400	MD 0.113 (95% CI: -0.084, 0.310; p = 0.2600),

Safety results

In both INPULSIS trials, the proportion of patients with serious adverse events was similar in the nintedanib and placebo groups. In INPULSIS-1, serious adverse events were reported in 31.1% of patients in the nintedanib group and in 27.0% of patients in the placebo group; in INPULSIS-2, the percentages were 29.8% and 32.9%, respectively.

Diarrhoea was the most frequent gastro-intestinal adverse reaction reported in 62.4% vs. 18.4% of patients treated with Ofev and placebo, respectively. In most patients the adverse reaction was of mild to moderate intensity and occurred within the first 3 months of treatment. Diarrhoea led to dose reduction in 10.7% of the patients and to discontinuation of nintedanib in 4.4% of the patients.

Liver enzyme elevations were reported in 13.6% of nintedanib treated patients. Elevations of liver enzymes were reversible and not associated with clinically manifest liver disease.

Applicability of results to England

Patients within the clinical trial programme for nintedanib are broadly representative of the UK IPF population. A significant number of UK patients were recruited into the programme, with the TOMORROW (Richeldi, 2011[29]) trial recruiting in 5 UK centres, and the INPULSIS (Richeldi, 2014[4]) trials featuring 8 UK trial sites, including 7 in England (Richeldi, 2014[4], ClinicalTrials.gov, 2011[33]). The basic demographics comparing the trial programmes are presented in Table 20.

Table 20: Basic demographics of nintedanib trials and the UK population

	TOMORROW	INPULSIS 1	INPULSIS 2	UK Primary Care Data
Mean Age (yrs)	64	66.9	67.1	71
% Male	63%	79.9%	78.1%	62%

More detailed demographic comparison is hindered by a paucity of epidemiological data; however the broad inclusion criteria in the INPULSIS trials are likely to reflect 'real-world' IPF. Specifically to this point, patients with a clinical diagnosis of IPF and an HRCT consistent with UIP (i.e. showing either honeycombing OR traction bronchiectasis in the absence of atypical features suggesting an alternative diagnosis) (Raghu *et al.*, 2011[13]) were eligible to take part in the INPULSIS trials in the absence of a surgical lung biopsy (Richeldi, 2014[4]). This contrasts to the ASCEND and CAPACITY trials for pirfenidone where a lung biopsy was required for patients unless a much stricter diagnosis of definite IPF was met on HRCT. In addition the INPULSIS exclusion criteria set no upper limit on FVC at recruitment and no restriction based on the existence of concomitant emphysema.

4.8 *Subgroup analysis*

Pre specified analyses INPULSIS

FVC ≤70% vs. >70%

Subgroup analyses of patients with baseline FVC ≤70% vs. >70% of predicted value were conducted using pooled data from the two INPULSIS trials were pre-specified. Analyses were conducted on the primary and key secondary endpoints by repeating the primary analysis of each endpoint in each subgroup.

- For the primary endpoint, the terms subgroup and an interaction term treatment-by time-by-subgroup were included in the model.
- For the key secondary endpoints, the terms subgroup and an interaction term treatment-by-subgroup were included in the model.

- Safety was assessed by subgroup via clinical and laboratory evaluation and the recording of adverse events.

The results of the subgroup analysis did not find any statistically significant differences in outcomes by subgroup, with nintedanib slowing disease progression in both groups. (Costabel, 2014[45]).

FVC ≤90% vs. >90%

A post-hoc subgroup analysis of patients with baseline FVC >90% vs. ≤90% predicted was undertaken using pooled data from the INPULSIS trials to investigate whether patients with marginally impaired FVC receive the same benefit from nintedanib.

There was no significant treatment-by-subgroup interaction for the primary endpoint (p=0.5300); in patients with baseline FVC >90% predicted, the adjusted annual rate of decline in FVC was -91.5 mL/year with nintedanib and -224.6 mL/year with placebo (difference: 133.1 mL/year [95% CI: 68.0, 198.2]) while in patients with baseline FVC ≤90% predicted, it was -121.5 mL/year with nintedanib and -223.6 mL/year with placebo (difference: 102.1 mL/year [95% CI: 61.9, 142.3]). No significant treatment-by-subgroup interaction was observed for the key secondary endpoints. The frequency of adverse events and serious adverse events was comparable between the treatment arms of each subgroup (BI data on file 2014).

Emphysema at baseline

In a post-hoc subgroup analysis of patients with or without emphysema at baseline in the pooled INPULSIS trials, nintedanib slowed disease progression by reducing decline in lung function in both groups. Similarly, time to first investigator reported acute exacerbation and change from baseline in SGRQ total score were consistent between patients with and without emphysema at baseline (Cottin V, 2014[46]).

4.9 *Meta-analysis*

The ISPOR task force for NMA recognised that the terms indirect treatment comparison (ITC), mixed treatment comparison (MTC), and NMA are sometimes used interchangeably and proposed specific definitions which are used in this section (Section 6.7).

4.10 *Indirect and mixed treatment comparisons*

The following steps were taken to identify and retrieve relevant clinical data on comparators:

1. Search based on the NICE decision problem;
2. Study selection;
3. Feasibility assessment (see section k below).

NMA Systematic Literature Review

A search strategy was developed in line with the decision problem in Table 1 and the study selection is described in 4.1 above.

Study selection for NMA

The process of study selection for abstract and full-text review was the same as detailed in section 4.1. However, the PICOS criteria were amended to capture RCTs of nintedanib or its comparators. The PICOS criteria used can be found in Table 21.

Table 21: Study selection for NMA

Criteria	Inclusion criteria	Exclusion criteria
Population	• Adult patients with IPF (including subgroups of patients e.g.subgroups based on 6MWT SpO2 and %VC)	• Paediatric and younger than 18; • Healthy participants; • Non-human.
Intervention/Comparator	• Nintedanib; • Pirfenidone; • N-acetylcysteine (monotherapy); • Best supportive care[1] (placebo).	NAC triple therapy Lung transplantation
Outcomes	**Efficacy** • Change from baseline in FVC (L); • Change from baseline in percent predicted FVC (%); • Annual rate of change in FVC (ml/year); • Change from baseline in DLco (ml/min/mmHg); • Change from baseline in 6MWT distance (m); • Change from baseline in worst SpO_2 during 6MWT; • Progression free survival[1]. • Exacerbation rate: acute IPF exacerbation; **Safety** • Adverse effects of treatment; • Mortality (hazard rate, mortality rate). **HRQL** • Change from baseline in HRQL as measured by disease-specific (e.g. SGRQ) or generic measure of HRQL (e.g. EQ-5D, SF-36D)	• No outcomes of interest
Study design	• Blinded or open-label RCTs (phase II, III, IV) of any duration; • Parallel-design extension of RCTs; • Post-hoc analyses of RCTs; • Pooled analyses of RCTs; • Studies published as abstracts or conference presentations were eligible for the primary analysis of clinical effectiveness if adequate data are provided	• RCTs (phase I); • Non-RCT studies including observational studies, experimental studies and single arm prospective trials; • Reviews and systematic literature reviews[2] of RCTs; • Case reports; • Critical appraisals; • Update or commentaries on data published elsewhere; • Notes, letters or editorials.
Language	• English language	• Non-English language

Abbreviations: HRQL, Health Related Quality of Life; RCT, Randomised Controlled Trial, SGRQ, St. George's Respiratory Questionnaire; EQ-5D, Euroqol 5-dimension; SF-36, Short Form 36 item questionnaire; DLco, DLco: diffusion capacity of the lung for carbon monoxide; SpO2: Peripheral capillary oxygen saturation; 6MWT: 6-minute walk test; IPF, Idiopathic pulmonary fibrosis

[1]Note: Best supportive care was defined as information and support, symptom relief, management of comorbidities, withdrawal of therapies suspected to be ineffective or causing harm, end of life care, oxygen therapy and/or pulmonary rehabilitation.

[2]Note: Systematic literature reviews were excluded from the main database search but their reference lists were searched for RCTs not captured by the main database search and were included as hand search in the review.

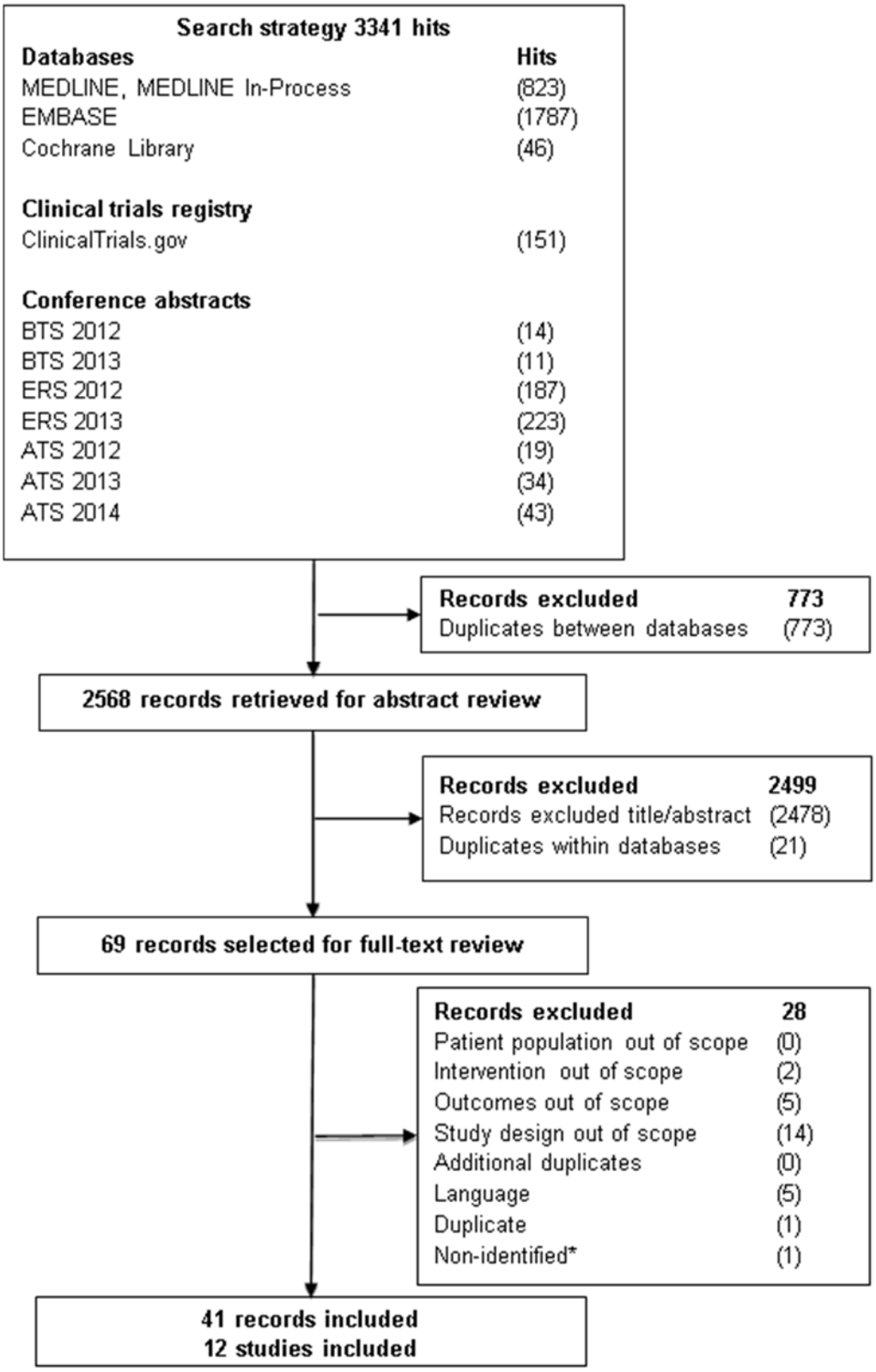
Results:

A total of 3,341 records, both published and unpublished, were identified from the systematic review of the literature. Subsequent titles and abstract review was performed and 67 records were retrieved for full-text review. Forty records were identified, reporting results of 12 RCTs (including INPULSIS I and II, CAPACITY I and II, ASCEND, SP2, SP3 and PANTHER IPF). Fourteen primary references, 12 full-text publications and 2 abstract, were selected for critical appraisal and data extraction. A study flow diagram for the NMA systematic literature review is illustrated in Figure 11.

Information on study design is presented in Table 22. Selected patient characteristics and treatment outcomes are presented in Table 23, Table 24, Table 25, Table 26, and Table 27. Full detail of the data extraction results are presented in Appendix C**.**

Please note: studies and NMA outputs for N-acetylcysteine are included within this section as the initial draft scope suggested it may be included as a comparator, and the NMA process was already underway when the final scope was received. For completeness this has been documented within this section. N-acetylcysteine is not included in the cost effectiveness section or model.

Figure 10: Study flow diagram for RCT study selection to be used in NMA (PRISMA diagram)



*The abstract (Sakae, 2014[47]) of a record identified in the 2014 ATS conference search in San Diego was not identified by the reviewer. Thus, email correspondence with the publisher was initiated. The abstract was not identified by the publisher.

Table 22: Study design of RCTs considered for inclusion in the NMA

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
RCTs for pirfenidone (PFN)							
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Prospective, double-blinded, randomised, placebo-controlled phase III trial	110 centres in 13 countries in Australia, Europe and North America	Pirfenidone (high dose)	Pirfenidone 2,403 mg/day	72 weeks	Inclusion criteria included predicted FVC of at least 50%, predicted carbon monoxide diffusion using capacity (DLco) of at least 35%, either predicted FVC or predicted DLco of 90% or less, and 6-min walk test (6MWT) distance of at least 150 m. Patients younger than 50 years and those not meeting protocol criteria for definite idiopathic pulmonary fibrosis by use of high-resolution CT (HRCT) were required to have a lung biopsy sample showing usual interstitial pneumonia.	Exclusion criteria included obstructive airway disease, connective tissue disease, alternative explanation for interstitial lung disease and being on a waiting list for a lung transplant.
			Pirfenidone (low dose)	Pirfenidone 1,197 mg/day			
			Placebo	Placebo			
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]			Pirfenidone (high dose)	Pirfenidone 2,403 mg/day			
			Placebo	Placebo			

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Prospective, double-blind, randomized, placebo-controlled phase III trial	25 sites in Japan	Pirfenidone	Pirfenidone 1,800 mg/day	36 weeks	Eligible patients were 20 to 75 years of age with adequate oxygenation at rest (PaO2 70 mm Hg) and demonstrated SpO2 of 90% or less during exertion while breathing air, within 1 month before enrolment.	Criteria for exclusion were a decrease in symptoms during the preceding 6 months, use of immunosuppressives and/or oral prednisone greater than 10 mg/day during the preceding 3months, clinical suspicion of idiopathic interstitial pneumonia other than IPF, coexisting emphysema (HRCT images of low attenuated areas in upper lung fields), pulmonary hypertension, asthma, tuberculosis, sarcoidosis, bronchiectasis other than traction associated, aspergillosis or respiratory infection; uncontrolled diabetes, comorbid conditions including malignancy, severe hepatic, renal, or cardiac disease; pregnancy (or its pursuance), breastfeeding; previous use of pirfenidone, suspicion of poor compliance in adherence to protocol, or being unable to understand protocol/written informed consent.
			Placebo	Placebo			

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
Taniguchi et al. 2010 (Taniguchi, 2010[49])*(SP3)	Multicentre, double-blind, placebo-controlled, randomised phase III clinical trial	73 centres in Japan	Pirfenidone (high dose)	Pirfenidone 1,800 mg/day	52 weeks	Eligible patients were adults (20–75 yrs of age) with IPF diagnosis and meeting the following arterial oxygen saturation measured by pulse oximetry (SpO2) criteria: 1) oxygen desaturation of >5% difference between resting SpO2 and the lowest SpO2 during a 6-min steady-state exercise test (6MET); and 2) the lowest Sp,O2 during the 6MET of >85% while breathing air. The 6MET procedure was in accordance with previous study protocols.	Exclusion criteria were: 1) a decrease in symptoms during the preceding 6 months; 2) use of immunosuppressants and/or oral corticosteroids at a dose >10 mg/day during the preceding 3 months; 3) clinical features of idiopathic interstitial pneumonia other than IPF; 4) evidence of known coexisting pulmonary hypertension, asthma, tuberculosis, bronchiectasis, aspergillosis or severe respiratory infection.
			Pirfenidone (low dose)	Pirfenidone 1,200 mg/day			
			Placebo	Placebo			
King et al. 2014 (King, 2014[50]) (ASCEND)	Multicentre, double-blind, placebo-controlled, randomised phase III clinical trial	127 centres in 9 countries	Pirfenidone	Pirfenidone 2,403 mg/day	52 weeks	Eligible patients were between the ages of 40 and 80 years and had received a centrally confirmed diagnosis of idiopathic pulmonary fibrosis. The diagnostic criteria, based on published consensus guidelines, were findings on high-resolution computed tomography (HRCT) that indicated either definite or possible usual interstitial pneumonia; the latter was confirmed on surgical lung biopsy. Other criteria for enrolment included a range of 50% to 90% of the predicted FVC, a range of 30% to 90% of the predicted carbon monoxide diffusing capacity, a ratio of the forced	Disease-related, medical, laboratory, and medication exclusion criteria
			Placebo	Placebo			

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
						expiratory volume in 1 second (FEV) to the FVC of 0.80 or more, and a 6-minute walk distance of 150 m or more.	
RCTs for N-acetylcysteine **(NAC)**							
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et al.*, 2012[51]) (PANTHER-IPF)	Multicentre, double-blind, placebo-controlled, randomised phase III clinical trial	25 centres in USA	NAC triple therapy	Prednisone: max 0.15 mg/kg for 25 weeks Azathioprine: max 150 mg/day NAC: 600 mg three times/day	60 weeks	Patients with idiopathic pulmonary fibrosis between the ages of 35 and 85 years who had mild to-moderate lung-function impairment (defined as a forced vital capacity [FVC] of ³≥50% and a carbon monoxide diffusing capacity of ≥30% of the predicted value) were potentially eligible.	Detailed study exclusion criteria listed in the study protocol.
			NAC	600 mg three times/day			
			Placebo	Placebo			

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	Multicentre, double-blind, placebo-controlled, randomised phase III clinical trial	25 centres in USA	NAC	N-Acetylcysteine 600 mg three times/day	60 weeks	Patients between the ages of 35 and 85 years who had idiopathic pulmonary fibrosis were eligible to participate in the study if they had mild-to-moderate impairment in pulmonary function, which was defined as a FVC of 50% or more of the predicted value and a carbon monoxide diffusing capacity of 30% or more of the predicted value. All patients were required to meet the modified criteria of the American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association for a diagnosis of idiopathic pulmonary fibrosis, 1,5 which was determined on the basis of high-resolution computed tomography (HRCT) or biopsy. All patients had received a diagnosis of idiopathic pulmonary fibrosis within 48 months before enrolment.	Patients were excluded if they had nonidiopathic fibrotic lung disease; an extent of emphysema that was greater than the extent of fibrotic change on HRCT; physiological evidence of airflow obstruction, which was defined as a ratio of the forced expiratory volume in 1 second (FEV1) to FVC of less than 0.65 or a residual volume of more than 120%; or any current signs or symptoms of severe, progressive, or uncontrolled coexisting illnesses, as determined by the site investigator; if they were on an active waiting list for lung transplantation; or if they had received combination therapy with azathioprine, prednisone, and acetylcysteine for more than 12 weeks in the previous 4 years. Patients who were originally randomly assigned to the discontinued three-drug regimen were not allowed to participate in the two-group study.
			Placebo	Placebo			
Tomioka et al. 2005 (Tomioka, 2005[53])	Open, randomised controlled phase III clinical trial	Japan	NAC	N-Acetylcysteine 352 mg/day	12 months	IPF patients who had not received any form of immunosuppressive therapy	Exclusion criteria were: (i) aged 80 years or over; (ii) a grave complication that would influence the clinical course of IPF; (iii) enrolment in a rehabilitation programme;
			Bromhexine	Bromhexine hydrochloride			

Author and year for primary records (Trial name)	Study design	Countries/ centres	Treatments	Drug name and daily dose	Duration of treatment	Inclusion criteria	Exclusion criteria
			hydrochloride	4 mg/day			and (iv) non-compliance in taking medications
Homma et al. 2012 (Homma, 2012[54])	Multicentre, prospective, randomized, controlled trial	Japan	NAC	N-Acetylcysteine 352 mg/day	48 weeks	Eligible patients were aged between 50 and 79 years and had firm clinical and radiological diagnoses of IPF, with the severity of disease classified as stage I or stage II, and a lowest arterial oxygen saturation value of >90% during the 6-min walk distance (6MWD) test. Eligible patients with stage I or stage II disease and no desaturation during the 6MWD test were considered to be in the early stage of IPF.	The exclusion criteria were an improvement in symptoms during the preceding 3 months; use of NAC, immunosuppressive agents, oral prednisone or pirfenidone; and clinical suspicion of idiopathic interstitial pneumonia other than IPF.
			Placebo	No therapy			

*Taniguchi 2011(Taniguchi, 2011[55]) and Azuma 2011 (Azuma *et al.*, 2011[56]) use the same data as Taniguchi 2010 (Taniguchi, 2010[49]) and are therefore not reported here.

Table 23: Patient characteristics of RCTs considered for inclusion in the NMA (1)

Author and year for primary records (Trial name)	Treatment	Diagnosis criteria	ITT n	% male	Mean age, years (SD)	Ethnicity				Smoking status			Duration of disease, years (SD)
						Caucasian	Black	Asian	Other	Non-smokers	Ex-smoker	Smoker	
RCTs for pirfenidone													
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Pirfenidone (high dose)	High resolution CT	174	68	65.7 (8.2)	NR	NR	NR	NR	32.0 %	63.0%	5%	NR
	Pirfenidone (low dose)		87	75	68.0 (7.6)	NR	NR	NR	NR	31.0 %	66.0%	3%	NR
	Placebo		174	74	66.3 (7.5)	NR	NR	NR	NR	29%	66.0%	5%	NR
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]	Pirfenidone		171	72	66.8 (7.9)	NR	NR	NR	NR	35.0 %	65.0%	0.0%	NR
	Placebo		173	72	67.0 (7.8)	NR	NR	NR	NR	37%	58.0%	5.0%	NR
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Pirfenidone	High resolution CT	72	86	64 (7.1)	NR	NR	NR	NR	11.0 %	79.0%	10.0%	<1 year: 28% 1-3 years: 24% >3 years: 54%
	Placebo		72	94	64.3 (7.6)	NR	NR	NR	NR	6.0%	86.0%	9.0%	<1 year: 28% 1-3 years: 24% >3 years: 54%
Taniguchi et al. 2010* (Taniguchi, 2010[49]) (SP3)	Pirfenidone (high dose)	ATS/ ERS	108	78.7	65.4 (6.2)	NR	NR	NR	NR	20.4 %	75.0%	4.6%	<1 year: 35.2% 1-3 years: 26.9% >3 years: 38%
	Pirfenidone (low dose)		55	85.5	63.9 (7.5)	NR	NR	NR	NR	21.8 %	60.0%	18.2%	<1 year: 36.4% 1-3 years: 23.6% >3 years: 40%
	Placebo		104	77.9	64.7 (7.3)	NR	NR	NR	NR	20.2 %	67.3%	12.5%	<1 year: 39.4% 1-3 years: 24% >3 years: 36.5%
King et al. 2014 (King, 2014[50]) (ASCEND)	Pirfenidone	ATS/ ERS/ JRS/ ALAT statement	278	79.9	68.4 (6.7)	NR	NR	NR	NR	NR	66.2%	NR	1.7(1.1)
	Placebo		277	76.9	67.8 (7.3)	NR	NR	NR	NR	NR	61.1%	NR	1.7 (1.1)
RCTs for NAC													

The Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et al.*, 2012[51]) (PANTHER-IPF)	NAC triple therapy	ATS/ ERS/ JRS/ ALAT statement	77	77	68.8 (7.3)	97 %	1%	NR	2%	30%	66%	4%	0.9 (1.1)
	NAC		81	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo		78	73	67.9 (8.1)	96 %	0%	NR	4%	26%	69%	5%	1.1 (1.0)
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	NAC (overall)	ATS/ ERS/ JRS/ ALAT statement	133	80.5	68.3 (8.4)	94.7 %	0.0 %	3.0 %	2.3 %	27.3 %	70.5%	2.3%	0.99 (1.00)
	Placebo (overall)		131	74.8	67.2 (8.2)	96.2 %	1.5 %	1.5 %	0.8 %	25.2 %	71.0%	3.8%	1.13 (1.00)
Tomioka et al. 2005 (Tomioka, 2005[53])	NAC	ATS/ ERS/ JRS/ ALAT statement	10	NR	70.0 (4.9)	NR	NR	NR	NR	50%	40%	10%	NR
	Bromhexine hydrochloride		12	NR	70.0 (5.3)	NR	NR	NR	NR	30%	50%	40%	NR
Homma et al. 2012 (Homma, 2012[54])	NAC	ATS/ ERS	38	76.0	67.6 (6.4)	NR	NR	NR	NR	26.3 %	65.8%	7.9%	3.0 (3.4)
	Placebo		38	76.0	68.2 (7.7)	NR	NR	NR	NR	23.7 %	68.4%	7.9%	3.2 (2.5)

*Taniguchi 2011(Taniguchi, 2011[55]) and Azuma 2011 (Azuma *et al.*, 2011[56]) use the same data as Taniguchi 2010 (Taniguchi, 2010[49]) and are therefore not reported here.

Table 24: Patient characteristics of RCTs considered for inclusion in the NMA (2)

Author and year for primary records (Trial name)	Treatment	FVC, % pred (SD)	TLC, % pred (SD)	DLco, % pred (SD)	PaO2 at rest, mmHg (SD)	SpO2, % (SD)	6MWT, m (SD)	Surgical Lung Biopsy, %	Prior Treatment for IPF, %
RCTs for pirfenidone									
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Pirfenidone (high dose)	74.5 (14.5)	NR	46.4 (9.5)	NR	17	417.5 (112.8)	49	NR
	Pirfenidone (low dose)	76.4 (14.4)	NR	47.2 (8.2)	NR	17	410.0 (90.9)	37	NR
	Placebo	76.2 (15.5)	NR	46.1 (10.2)	NR	14	378.0 (82.2)	49	NR
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]	Pirfenidone	74.9 (13.2)	NR	47.8 (9.8)	NR	28	399.1 (89.7)	55	NR
	Placebo	73.1 (14.2)	NR	47.4 (9.2)	NR	28	417.5 (112.8)	54	NR
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Pirfenidone	NR	78.5 (17.9)	57.6 (17.2)	80.3 (7.7)	87.1 (3.9)	NR	21.0	14
	Placebo	NR	73.9 (16.4)	57.7 (13.8)	82.0 (7.6)	87.1 (4.2)	NR	23.0	14
Taniguchi et al. 2010* (Taniguchi, 2010[49]) (SP3)	Pirfenidone (high dose)	NR	75.2 (15.7)	52.1 (16.8)	79.8 (10.2)	NR	NR	24.1	8.3
	Pirfenidone (low dose)	NR	78.5 (17.9)	53.6 (19.1)	81.6 (8.4)	NR	NR	29.1	10.9
	Placebo	NR	73.9 (16.4)	55.2 (18.2)	81.0 (9.5)	NR	NR	26.9	5.8
King et al. 2014 (King, 2014[50]) (ASCEND)	Pirfenidone	67.8 (11.2)	43.7 (10.5)	NR	NR	NR	415.0 (98.5)	86	NR
	Placebo	68.6 (10.9)	44.2 (12.5)	NR	NR	NR	420.7 (98.1)	79	NR
RCTs for NAC									
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et*	NAC triple therapy	69.3 (15.1)	NR	42.1 (10.2)	79.6 (9.7)	NR	362 (113)	NR	NR
	NAC	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo	72.1	NR	45.3	78.8	NR	368.9	NR	NR

Author and year for primary records (Trial name)	Treatment	FVC, % pred (SD)	TLC, % pred (SD)	DLco, % pred (SD)	PaO2 at rest, mmHg (SD)	SpO2, % (SD)	6MWT, m (SD)	Surgical Lung Biopsy, %	Prior Treatment for IPF, %
al., 2012[51]) (PANTHER-IPF)		(14.4)		(12.4)	(12.6)		(117.3)		
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	NAC (overall)	72.2 (15.9)	NR	44.7 (10.8)	80.7 (10.5)	95.75 (2.45)	371 (116)	52.7	NR
	Placebo (overall)	73.4 (14.3)	NR	46.0 (12.2)	81.5 (11.8)	96.1 (2.30)	375 (105)		NR
Tomioka et al. 2005 (Tomioka, 2005[53])	NAC	67.6 (15.7)	NR	64.7 (21.7)	NR	90.1 (5.9)	385 (90)	NR	NR
	Bromhexine hydrochloride	76.6 (19.1)	NR	60.7 (16.7)	NR	91.1 (5.9)	390 (116)	NR	NR
Homma et al. 2012 (Homma, 2012[54])	NAC	89.2	82.5 (17.4)	72.3 (25.3)	NR	NR	NR	NR	0.0
	Placebo	88.7	81.2 (13.3)	64.4 (20.1)	NR	NR	NR	NR	0.0

*Taniguchi 2011 and Azuma 2011 use the same data as Taniguchi 2010 and are therefore not reported here.

Table 25: Efficacy outcomes of RCTs considered for inclusion in the NMA (1)

Author and year for primary records (Trial name)	Treatment	ITT, n	Time point	Change in FVC from baseline, L		Mean difference between treatments, L (p-value),	Change in FVC from baseline, % predicted		Mean difference between treatments, % pred (p-value)
				Description	Mean (SD)		Description	Mean (SD)	
RCTs for pirfenidone									
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Pirfenidone (high dose)	174	72 weeks	Categorical change in FVC ≥10%	35	14.4 (p=0.001)	NR	NR	NR
	Pirfenidone (low dose)	87			NR	NR	NR	NR	NR
	Placebo	174			60	-	NR	NR	NR
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]	Pirfenidone	171			39	3.8 (p=0.440)	NR	NR	NR
	Placebo	173			46		NR	NR	NR
Noble et al. 2011 (Noble, 2011[40]) Pooled analysis	Pirfenidone	345			74	9.1 (p=0.003)	NR	NR	NR
	Placebo	347			106		NR	NR	NR
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Pirfenidone	72	24 weeks	Absolute	-0.01 (0.21)	(p=0.0995)	NR	NR	NR
	Pirfenidone	35	36 weeks	Absolute	-0.03 (0.22)	(p=0.0366)	NR	NR	NR
	Placebo	72	24 weeks	Absolute	-0.08 (0.19)	-	NR	NR	NR
	Placebo	35	36 weeks	Absolute	-0.13 (0.19)	-	NR	NR	NR
Taniguchi et al. 2010 (Taniguchi,	Pirfenidone (high dose)	108	52 weeks	Absolute	-0.09	0.07 (0.0416)	NR	NR	NR

2010[49])*(SP3)	Pirfenidone (low dose)	55		Absolute	-0.08	0.09 (0.0394)	NR	NR	NR
	Placebo	104		Absolute	-0.16	-	NR	NR	NR
Azuma et al. 2011 (Azuma *et al.*, 2011[56]) (SP3)*	Pirfenidone (high dose)	35 (Subgroup:6MWT SpO2 ≥90, %VC ≥70)	52 weeks	Absolute	-0.072	(p=0.7035)	Absolute	-2.083	(p=0.6438)
		9 (Subgroup:6MWT SpO2≥90, %VC <70)	52 weeks	Absolute	-0.263	(p=0.7181)	Absolute	-9.084	(p=0.5902)
		36 (Subgroup:6MWT SpO2 <90, %VC ≥70)	52 weeks	Absolute	-0.05	(p=0.0359)	Absolute	-1.737	(p=0.0555)
		23 (Subgroup:6MWT SpO2<90, %VC <70)	52 weeks	Absolute	-0.148	(p=0.7768)	Absolute	-4.143	(p=0.5037)
	Pirfenidone (low dose)	16 (Subgroup:6MWT SpO2 ≥90, %VC ≥70)	52 weeks	Absolute	-0.004	(p=0.1595)	Absolute	-0.42	(p=0.2209)
		6 (Subgroup:6MWT SpO2≥90, %VC <70)	52 weeks	Absolute	-0.185	(p=0.735)	Absolute	-7.34	(p=0.9647)
		17 (Subgroup:6MWT SpO2 <90, %VC ≥70)	52 weeks	Absolute	-0.016	(p=0.0372)	Absolute	-0.59	(p=0.0493)
		15 (Subgroup:6MWT SpO2<90, %VC <70)	52 weeks	Absolute	-0.181	(p=0.8735)	Absolute	-5.09	(p=0.828)
	Placebo	32 (Subgroup:6MWT SpO2 ≥90, %VC ≥70)	52 weeks	Absolute	-0.09	-	Absolute	-2.801	-
		9 (Subgroup:6MWT SpO2≥90, %VC <70)	52 weeks	Absolute	-0.225	-	Absolute	-7.167	-
		36 (Subgroup:6MWT SpO2 <90, %VC ≥70)	52 weeks	Absolute	-0.199	-	Absolute	-6.08	-
		26 (Subgroup:6MWT SpO2<90, %VC <70)	52 weeks	Absolute	-0.168	-	Absolute	-5.669	-
King et al. 2014 (King, 2014[50]) (ASCEND)	Pirfenidone	278	52 weeks	Absolute	-0.123	-0.14 (p<0.0001)	Categorical change (% patients with ≥10% Decline or	16.50%	<0.000001
	Placebo	277	52 weeks	Absolute	-0.262			31.80%	

							Death)		
RCTs for NAC									
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et al.*, 2012[51]) (PANTHER-IPF)	NAC triple therapy	77	60 weeks	Absolute	-0.24	-0.01 (p=0.85)	NR	NR	NR
	NAC	81	NR	NR	NR	NR	NR	NR	NR
	Placebo	78	60 weeks	Absolute	-0.23	-	NR	NR	NR
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	NAC (overall)	133	60 weeks	Absolute	-0.18	0.01 (p=0.77)	Absolute	-3.9 (6.3)	(p=0.875)
	Placebo (overall)	131	60 weeks	Absolute	-0.19		Absolute	-3.92 (6.95)	
Tomioka et al. 2005 (Tomioka, 2005[53])	NAC	10	52 weeks	NR	NR	NR	Absolute	−7.2	NR
	Bromhexine hydrochloride	12	52 weeks	NR	NR		Absolute	−9.6	
Homma et al. 2012 (Homma, 2012[54])	NAC	38	48 weeks	Absolute	-0.09 (0.3)	(p=0.266)	NR	NR	NR
	Placebo	38	48 weeks	Absolute	-0.15 (0.2)		NR	NR	

*Taniguchi 2011(Taniguchi, 2011[55]) uses the same data as Taniguchi 2010 (Taniguchi, 2010[49]) and are therefore not reported here.

Table 26: Efficacy outcomes of RCTs considered for inclusion in the NMA (2)

Author and year for primary records (Trial name)	Treatment	ITT, n	Change in 6MWT from baseline, m Mean (SD)	Mean difference between treatments, m (p-value)	Change in worst SpO2 during 6MWT from baseline, % Mean (SD)	Mean difference between treatments, % (p-value)	Change in DLco from baseline, ml/min/mmHg Mean (SD)	Mean difference between treatments, ml/min/mmHg (p-value)
RCTs for pirfenidone								
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Pirfenidone (high dose)	174	-60.4 (NR)	16.4 (p=0.171)	-1.5	0.8 (p=0.087)	-7.9 (NR)	2.0 (p=0.145)
	Pirfenidone (low dose)	87	NR	NR	NR	NR	NR	NR
	Placebo	174	-76.8 (NR)	-	-2.3	-	-9.9 (NR)	-
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]	Pirfenidone	171	-45.1 (NR)	31.8 (p<0.001)	-1.9	-0.5 (0.893)	-9.8 (NR)	-0.5 (p=0.996)
	Placebo	173	-76.9 (NR)		-1.3		-9.2 (NR)	
Noble et al. 2011 (Noble, 2011[40]) Pooled analysis	Pirfenidone	345	-52.8 (NR)	24 (p<0.001)	-1.7	0.1 (p=0.261)	-8.8 (NR)	0.7 (0.301)
	Placebo	347	-76.8 (NR)		-1.8		-9.6 (NR)	
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Pirfenidone	72	NR	NR	0.6364 (3.5502)	0.1489	-0.5 (0.207)	(p=0.4894)
	Pirfenidone	35	NR	NR	0.4697 (3.8838)		-0.57 (2.15)	(p=0.2120)
	Placebo	72	NR	NR	-0.5484 (3.7933)	0.0722	-0.83 (2.16)	-
	Placebo	35	NR	NR	-0.9355 (3.3559)		-1.19 (2.3)	-
Taniguchi et al. 2010 (Taniguchi, 2010[49]) (SP3)	Pirfenidone (high dose)	108	NR	NR	-1.7 (p=0.35)	-0.17 (p=0.739)	NR	NR
	Pirfenidone (low dose)	55	NR	NR	-0.84 (0.48)	0.69	NR	NR
	Placebo	104	NR	NR	-1.53 (p=0.35)	-	NR	NR
Taniguchi et al. 2011 (Taniguchi,	Pirfenidone (high dose)	NR	NR	NR	NR	NR	NR	NR
	Pirfenidone	NR	NR	NR	NR	NR	NR	NR

2011[55]) (SP3)	(low dose)							
	Placebo	NR	NR	NR	NR	NR	NR	NR
Azuma et al. 2011 (Azuma *et al.*, 2011[56]) (SP3)	Pirfenidone		NR	NR	NR	NR	NR	NR
	Placebo		NR	NR	NR	NR	NR	NR
King et al. 2014 (King, 2014[50]) (ASCEND)	Pirfenidone	278	-33.2	26.7 (p=0.036)	NR	NR	NR	NR
	Placebo	277	-59.8		NR		NR	
RCTs for NAC								
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et al.*, 2012[51]) (PANTHER-IPF)	NAC triple therapy	77	-93	-19.5 (p=0.56)	NR	NR	NR	NR
	NAC	81	NR	NR	NR	NR	NR	NR
	Placebo	78	-73.6	-	NR	NR	NR	NR
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	NAC (overall)	133	NR	NR	NR	NR	NR	NR
	Placebo (overall)	131	NR		NR		NR	
Tomioka et al. 2005 (Tomioka, 2005[53])	NAC	10	14	NR	NR	NR	NR	NR
	Bromhexine hydrochloride	12	-52.4		NR		NR	
Homma et al. 2012 (Homma, 2012[54])	NAC	38	NR	NR	NR	NR	NR	NR
	Placebo	38	NR		NR		NR	

* Data also presented for a subgroup based on 6MWT, SpO2 and PaO2. Please see Data extraction file in Appendix C.

Table 27: Safety outcomes of RCTs considered for inclusion in the NMA

Author and year for primary records (Trial name)	Treatment	Safety population, n	Time point	Exacerbations			Mortality				
				Definition	Patients experiencing event, n	Patients experiencing event, %	Definition	Total number of patients experiencing event	% patients experiencing event	HR	P-value
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-2) [PIPF-004]	Pirfenidone (high dose)	174	72 weeks	NR	NR	NR	NR	NR	NR	NR	NR
	Pirfenidone (low dose)	87	72 weeks	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo	174	72 weeks	NR	NR	NR	NR	NR	NR	NR	NR
Noble et al. 2011 (Noble, 2011[40]) (CAPACITY-1) [PIPF-005]	Pirfenidone	171	72 weeks	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo	173	72 weeks	NR	NR	NR	NR	NR	NR	NR	NR
Noble et al. 2011 (Noble, 2011[40]) Pooled analysis	Pirfenidone	345	72 weeks	4	NR	1.16	All-cause	27	8	0.77	0.32
							IPF-related	18	5	0.62	0.12
	Placebo	347	72 weeks	4	NR	1.15	All-cause	34	10	0.65	0.14
							IPF-related	28	8	0.48	0.03
Azuma et al. 2005 (Azuma *et al.*, 2005[48]) (SP2)	Pirfenidone	73	32 weeks	Worsening, otherwise unexplained clinical features within 1 month: progression of dyspnoea over a few days to less than 5 weeks, new radiographic/HRCT parenchymal abnormalities without pneumothorax or pleural effusion (e.g., new, superimposed	0	0	NR	0	0	NR	NR
	Placebo	36	32 weeks		5	13.9	NR	1	2.78	NR	NR

				ground-glass opacities), a decrease in the PaO2 by 10 mm Hg or more, and exclusion of apparent infection based on absence of Aspergillus and pneumococcus antibodies in blood, urine for Legionella pneumophila, and sputum cultures.							
Taniguchi et al. 2010 (Taniguchi, 2010[49]) (SP3)	Pirfenidone (high dose)	109	52 weeks	Acute exacerbation of IPF was defined according to previous reports and revised criteria for acute exacerbation of IPF in Japan	6	5.6	Any	0	0%	NR	NR
	Pirfenidone (low dose)	55	52 weeks		3	5.5	Any	0	0%	NR	NR
	Placebo	107	52 weeks		5	4.8	Any	0	0%	NR	NR
Taniguchi et al. 2011 (Taniguchi, 2011[55]) (SP3)	Pirfenidone	267	52 weeks	Acute exacerbations	NR	13	NR	NR	NR	NR	NR
	Placebo		52 weeks		NR		NR	NR	NR	NR	NR
Azuma et al. 2011 (Azuma *et al.*, 2011[56]) (SP3)	Pirfenidone	108	52 weeks	NR	NR	NR	NR	NR	NR	NR	NR
	Placebo	104									
King et al. 2014 (King, 2014[50]) (ASCEND)	Pirfenidone	278	52 weeks	NR	NR	NR	All-cause	11	4.0%	0.55	0.105
	Pirfenidone	278	52 weeks		NR	NR	IPF-related	3	1.1%	0.44	0.23
	Placebo	277	52 weeks		NR	NR	All-cause	20	7.2%	-	-
	Placebo	277	52 weeks		NR	NR	IPF-related	7	2.5%	-	-
The	NAC triple	77	60	The definition of an	5.0	6%	All-cause	8.0	10%	NR	0.01

Idiopathic Pulmonary Fibrosis Clinical Research Network 2012 (Research *et al.*, 2012[51]) (PANTHER-IPF)	therapy		weeks	acute exacerbation was prespecified and was in accordance with criteria reported previously			Respiratory cause	7	9%	NR	0.02
	NAC	78	NR		NR	NR	NR	NR	NR	NR	NR
	Placebo	77	60 weeks		5.0	6%	All-cause	1	1%	-	-
							Respiratory cause	1	1%	-	-
The Idiopathic Pulmonary Fibrosis Clinical Research Network 2014 (Martinez *et al.*, 2014[52]) (PANTHER-IPF)	NAC (overall)	133	60 weeks		3	2.3%	All-cause	6	4.5%	NR	NR
	Placebo (overall)	131	60 weeks		3	2.3%	All-cause	3	2.3%	NR	
Tomioka et al. 2005 (Tomioka, 2005[53])	NAC	10	50 weeks	Acute exacerbations	NR	NR	NR	NR	NR	NR	NR
	Bromhexine hydrochloride	12	50 weeks		NR	NR	NR	NR	NR	NR	NR
Homma et al. 2012 (Homma, 2012[54])	NAC	44	48 weeks	Acute exacerbations	1	2.2%	NR	0	0.0%	NR	NR
	Placebo	46	48 weeks		4	8.7%	NR	0	0.0%	NR	NR

In addition, it came to light shortly before the submission date that an NMA comparing new treatments for IPF was published online on the 18th April 2015(Loveman *et al.*, 2015[57]). Given the short time frame between viewing the publication and submitting the NICE HTA, a brief review and comparision between the NMA carried out in this submission and the Loveman NMA was undertaken.

A top line comparison of the inputs and results between the Loveman 2015 NMA and the BI NMA was performed with a focus on the NDB vs. PFN comparison. The outcomes considered were mortality, acute exacerbations, and at least 10% decline in FVC.

Several data discrepancies were observed in the reported outcomes between Loveman 2015 and the BI NMA. For the mortality data inputs, Loveman 2015 excluded the Japanese PFN studies by Azuma 2005(Azuma *et al.*, 2005[48]) and Taniguxhi 2010(Taniguchi, 2010[49]) from the NMA, while the BI economic model used all available evidence (including the above studies). Loveman 2015 reported PFN mortality values for CAPACITY(Noble, 2011[40]) that were different from the BI analysis; however, it is likely that they have used 52 weeks mortality data rather than 72 weeks data that the BI analysis used. The point estimate for the NDB vs. PFN mortality OR was higher in the Loveman 2015 analysis than the BI NMA; however, neither result was statistically significant (Loveman NMA – 1.39 (95% CI 0.70-2.82), BI NMA – 1.03 (95% CI 0.55 – 1.91).

Differences in the NDB and PFN exacerbation data reported were also observed. The BI analysis used NDB exacerbations data from the phase II trial(Richeldi, 2011[29]) using IPF progression as a proxy for exacerbations. The Loveman 2015 analysis used Richeldi 2011 values that were calculated based on the incidence of acute exacerbations per 100 patient years; which resulted in different values than the BI analysis used. As for the PFN exacerbations data inputs, Loveman 2015 did not report any exacerbations for CAPACITY (Noble, 2011[40]), while the BI NMA included exacerbations from CAPACITY (reported in the NICE STA 2011). A comparison of the Loveman and BI NMA results for exacerbations was not possible since the Loveman 2015 analysis did not undertake the indirect comparison due to several data uncertainties.

As for FVC decline data inputs, it is not clear whether Loveman used the “at least 10% decline in FVC or death” or the “10% decline in FVC” definition for the NDB inputs of this outcome. For PFN FVC decline data inputs, Loveman 2015 assumed that VC is equivalent to FVC and as a result included 10% VC decline data from Azuma 2005. The BI NMA did not consider the VC evidence as equivalent to FVC decline, for this reason it was not included in the analysis. However, Taniguchi 2010 also reported VC decline, but data from Taniguchi were not included in the Loveman NMA. The point estimate for the NDB vs. PFN 10% FVC decline OR was higher in the Loveman 2015 analysis than the BI NMA; however, both results were not statistically significant (Loveman NMA – 1.21 (95% CI 0.86-1.72), BI NMA – 1.02 (05%CI 0.72 – 1.46).

Moreover, the Loveman 2015 NMA considered FVC decline data defined as “change in FVC%Pred or absolute change from baseline.” The indirect comparison result for this outcome definition was found to be statistically significant for NDB vs. PFN. The BI analysis did not consider this definition of FVC decline so a comparison is not possible for this outcome definition.

Table 28: Bias risks in clinical trials

Bias type	nintedanib		pirfenidone				N-acetylcysteine		
	Richeldi et al. 2011*(Richeldi *et al.*, 2011[1])	Richeldi et al. 2014*(Richeldi *et al.*, 2014[12])	Noble et al. 2011(Noble, 2011[40])	King et al. 2014 (King *et al.*, 2014[41])	Azuma et al. 2005 (Azuma *et al.*, 2005[48])	Taniguchi et al. 2010 (Taniguchi, 2010[49])	Martinez et al. 2014 (Martinez *et al.*, 2000[58])	Homma et al. 2012 (Homma, 2012[54])	Tomioka et al. 2005 (Tomioka, 2005[53])
Random sequence generation (selection bias)	Low	Low	Low	Low	Low	Low	Low	Low	Low
Allocation concealment (selection bias)	Low	Low	Low	Low	Low	Low	Low	Low	High: open prospective study
Double Blinding (performance bias)	Low	Low	Low	Low	Low	Low	Low	High: not a blinded study	High: not a blinded study
Blinding of outcomes assessment (detection bias)	Low	Low	Low	Low	Low	Low	Low	High: not a blinded study	High: not a blinded study
Incomplete outcome data (attrition bias)	Low	Low	Low	Low	Low	Low	Low	High: not reporting outcomes for patients who were excluded for a number of reasons	High: not reporting outcomes for patients who died, lost to follow-up, had other disorders
Selective reporting (reporting bias)	Low	Low	Low	Low	Low	Low	Low	Low	Moderate: does not report clear primary endpoints

*The availability of individual patient data was considered when assessing the risk of bias

The majority of studies were of good quality, with low risk of systematic errors. Moderate or high risk of bias was identified in two studies.

The study by Homma et al. (Homma, 2012[54]) was not blinded or placebo-controlled. Outcomes were not reported for patients who were excluded for reasons such as, acute exacerbation, progression of disease, pneumonia, death, and after patients' request. No deaths were observed during the trial.

The study by Tomioka et al. (Tomioka, 2005[53]) was an open-label, prospective study. The primary and secondary endpoints were not clearly defined; as a result the outcomes may have been selectively reported. Outcomes were not reported for patients who were excluded due to respiratory failure death, lost to follow-up, or developed another disease (lung cancer).

Sensitivity analysis tested the impact of the identified study distribution biases to the network by excluding Homma et al. (Homma, 2012[54]), where applicable.

Networks and key data

Overall Survival

Eight studies reported mortality. Homma et al (Homma, 2012[54]) did not report any deaths and therefore, it was omitted from the network as it would not contribute any evidence to estimate the treatment effect in the network. Figure 11 presents the network for OS. Table 29 presents the study data for overall survival. All-cause mortality in Azuma et al. (Azuma *et al.*, 2005[48]) was not reported clearly; the study only reported one death that followed an acute exacerbation event and no further information about overall survival.

Figure 11: Network for OS

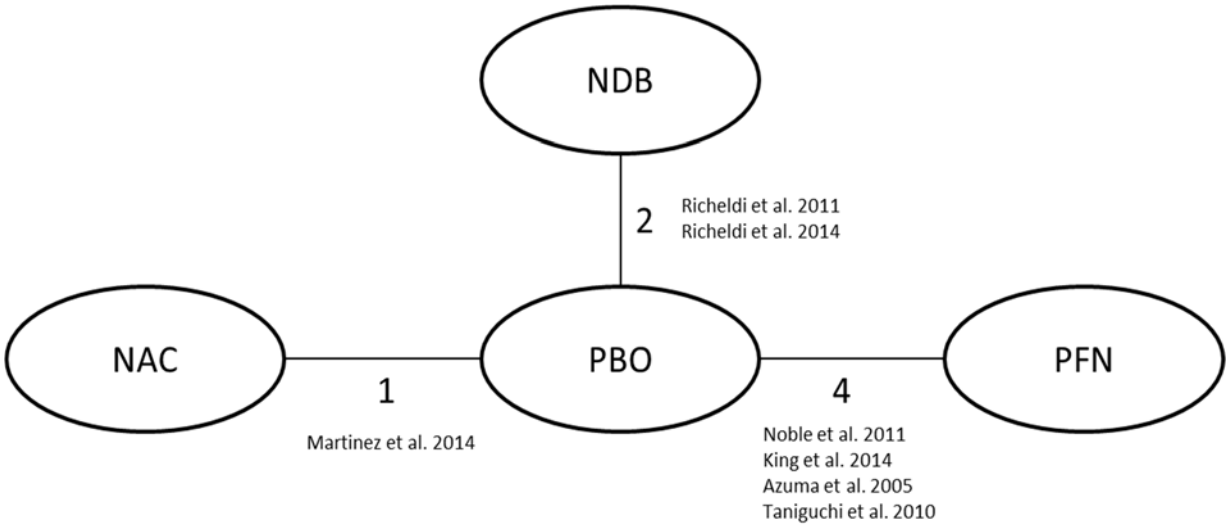


Table 29: Overall survival evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	7	85
Richeldi 2011	PBO	9	85
Richeldi 2014	NDB	35	638
Richeldi 2014	PBO	33	423
Noble 2011	PFN	27	345
Noble 2011	PBO	34	347
Azuma 2005	PFN	1*	73
Azuma 2005	PBO	0	36
Taniguchi 2010	PFN	3	108
Taniguchi 2010	PBO	4	104
King 2014	PFN	11	278
King 2014	PBO	20	277
Martinez 2014	NAC	6	133
Martinez 2014	PBO	3	131
Homma 2012	NAC	0	44
Homma 2012	PBO	0	46

*One death following an acute exacerbation event. Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Acute exacerbation events

Seven studies reported acute exacerbations. Figure 12 presents the network for acute exacerbations. Table 30 presents the study data for acute exacerbation events. For Richeldi et al. (Richeldi *et al.*, 2014[12]) acute exacerbation event data were available based on two definitions: investigator reported and adjudication committee evaluation. The analysis used only the investigator reported outcomes.

Figure 12: Network for acute exacerbations

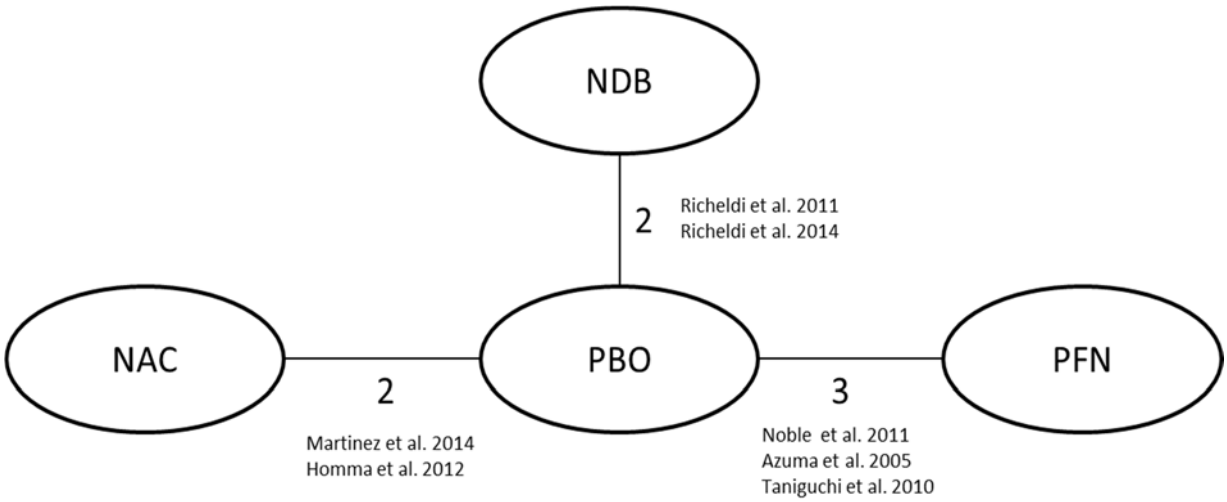


Table 30: Acute exacerbations evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	4	85
Richeldi 2011	PBO	11	85
Richeldi 2014	NDB	31	638
Richeldi 2014	PBO	32	423
Noble 2011	PFN	4	345
Noble 2011	PBO	4	347
Azuma 2005	PFN	0	72
Azuma 2005	PBO	5	35
Taniguchi 2010	PFN	6	108
Taniguchi 2010	PBO	5	104
Martinez 2014	NAC	3	133
Martinez 2014	PBO	3	131
Homma 2012	NAC	1	44
Homma 2012	PBO	4	46

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Pulmonary function parameters – loss of lung function

Forced vital capacity (FVC) is used consistently across clinical trials in IPF and is considered to be a marker for the progression of the disease. Moreover, it is often used in clinical practice to assess patient levels of pulmonary function and changes from previous visits. FVC percent predicted (FVC%Pred) is a standardization of the FVC outcome adjusted for factors that could affect the capacity of the individual: age, gender, and height. This makes FVC%Pred more useful for comparisons across studies as it removes some heterogeneity around the outcome. A number of studies have investigated the minimal clinically important difference in changes to FVC percent predicted values, ranging from 2-10% (du Bois *et al.*, 2011[7], Raghu *et al.*, 2011[13]). After consultation with clinical experts the analysis was based on a categorical change of FVC%Pred of 10 points.

Five studies reported values for loss of lung function in FVC%Pred. Figure 13 presents the network for this outcome. Table 31 presents the study data.

The data for TOMORROW (Richeldi, 2011[29]) and INPULSIS (Richeldi *et al.*, 2014[12]) were obtained from a post-hoc analysis. The analysis was based on observed data only; no imputation or other assumptions were made to account for the fact that some patients may not have had an FVC%Pred measure at every visit or may have dropped from the study before the end of the 52 weeks.

For the CAPACITY trial (Noble, 2011[40]) loss of lung function was available from a report to the European Medicines Agency (EMA) (EMA, 2013[59]). The patient figures were presented in a breakdown of a composite outcome of PFS.

In King et al. (King *et al.*, 2014[41]) progression was presented as the cumulative number of patients dead or experiencing a decline of 10%-points or more in FVC%Pred or death at 13, 26, 39, and 52 weeks compared to baseline. In line with study duration, the 52 week estimate was used in this analysis.

In Martinez et al. (Martinez *et al.*, 2014[52]) progression was defined as a reduction of more than 10% in FVC or death. Since FVC % predicted is estimated based on a standardisation of FVC for age, gender and height, it was reasonable to assume that the categorical outcome of the percentage of patients who experienced a drop in either FVC or FVC%Pred would be similar if not the same.

Figure 13: Network for loss of lung function

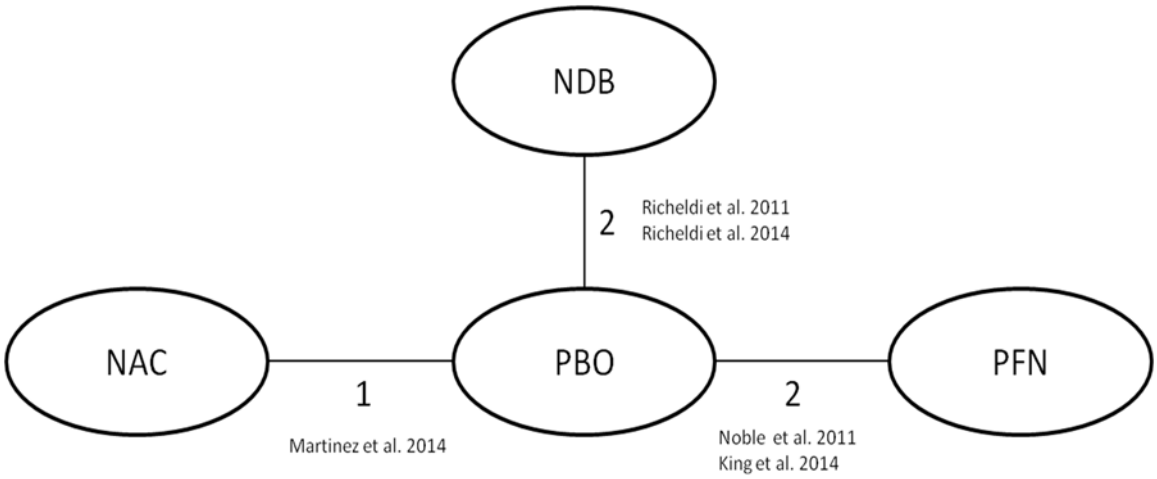


Table 31: Loss of lung function evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	19	85
Richeldi 2011	PBO	29	85
Richeldi 2011*	NDB	25	85
Richeldi 2011*	PBO	34	85
Richeldi 2014	NDB	148	638
Richeldi 2014	PBO	153	423
Richeldi 2014*	NDB	173	638
Richeldi 2014*	PBO	175	423
Noble 2011	PFN	59	345
Noble 2011	PBO	80	347
King 2014*	PFN	46	278
King 2014*	PBO	88	277

Study	Treatment	Observed events	Patients at risk
Martinez 2014*	NAC	36	133
Martinez 2014*	PBO	35	131

*10% drop in FVC%Pred or death

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Frequently reported pulmonary function parameters in clinical trials are carbon monoxide diffusing lung capacity (DLco) and peripheral capillary oxygen saturation (SpO_2). Both of these parameters require several criteria to be met concomitantly, requiring experienced staff and logistics, which make these outcomes not easily reproducible (BI data on file). Based on this these outcomes were considered unreliable for a comparison across different studies.

Other pulmonary function parameters are either raw values not correcting for demographic factors or are built on models that might differ across trials, making it difficult to synthesise them in an indirect comparison.

PFS

PFS is not consistently reported or defined in the same way across studies. Three pirfenidone (PFN) studies included data on PFS (Noble, 2011[40], King *et al.*, 2014[41], Taniguchi, 2010[49]). A pairwise comparison of nintedanib (NDB) and PFN on PFS using the definition of a composite endpoint presented in Noble et al (Noble, 2011[40]) was possible (Figure 14). In Noble et al. (Noble, 2011[40]) PFS was defined as time to confirmed ≥10% decline in FVC%Pred, or ≥15% decline in DLco%Pred, or death. The study presented PFS as a hazard ratio (HR) with 95% confidence intervals (CI). Individual patient data (IPD) from the Richeldi et al. (Richeldi *et al.*, 2014[12]) study were analysed post-hoc to replicate the PFS outcome definition from Noble et al. (Noble, 2011[40]). Table 32 presents the relevant data from each study.

Figure 14: Pairwise comparison for PFS

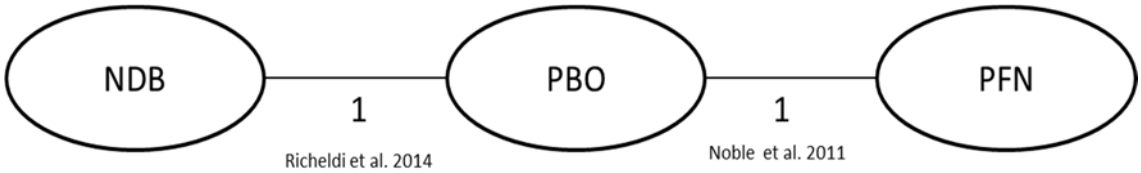


Table 32: PFS evidence (Noble et al. [2011] definition)

Study	HR vs. PBO	95% CI Lower limit	95% CI Upper limit
Richeldi 2014	0.74	0.61	0.91
Noble 2011	0.74	0.57	0.96

Physical function parameters

The change from baseline in a 6MWT was the only physical function outcome consistently reported across different trials. Although the 6MWT is widely used in clinical practice, the value of the test is limited in clinical practice due to the high degree of standardization required. For example, the test requires a number of criteria to be met: the test should be performed indoors, across a long, flat, straight, enclosed corridor with a hard surface that is seldom travelled; the walking course must be 30m in length; with marks every 3m; turnaround points marked with cones, measure SpO_2 and pulse rate. There is no clear consensus if desaturation, distance walked, and other variables measured during 6MWT in a population of patients with IPF are reproducible in the clinical setting (Alexeeff *et al.*, 2007[43], Griffiths *et al.*, 2012[60]). Therefore, any interpretation of the results of the relative efficacy of investigational drugs based on the 6MWT is very difficult.

Three studies reported complete data on the change from baseline of the distance in a 6MWT (Figure 15). Table 33 presents the relevant data from each study. Note that King et al. (King *et al.*, 2014[41]) and Tomioka et al. (Tomioka, 2005[53]) did not report any precision around the point estimates, and a comparison with the other studies was not possible.

Figure 15: Network for 6MWT

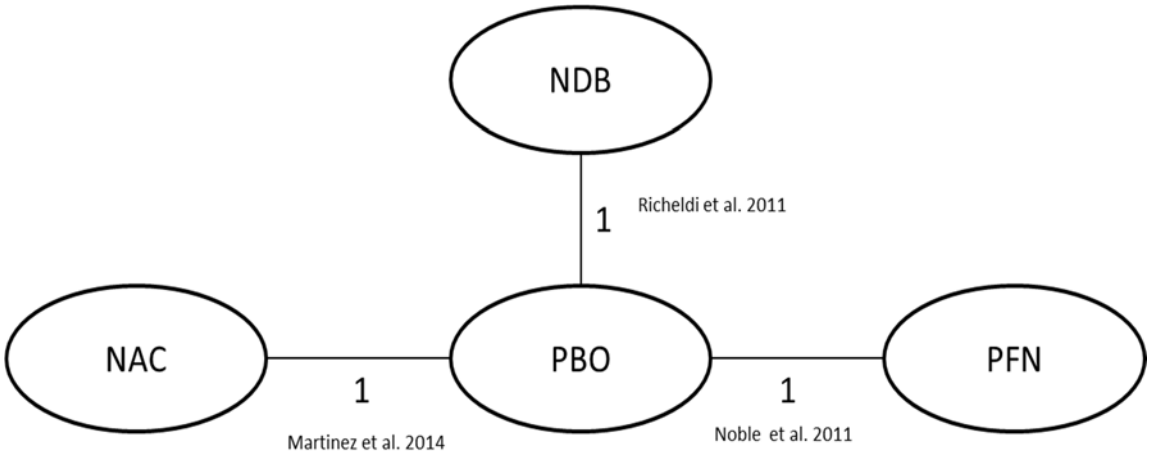


Table 33: 6MWT evidence

Study	Treatment	Distance at baseline (m)		Change from baseline (m)	
		Mean	SD	Change in distance	Absolute difference
Richeldi 2011	NDB	437	13.69	-25.15	6.32 (SE 16.98) 95% CI (027.08, 39.72), p=0.7101
Richeldi 2011	PBO	411.1	15.9	-26	
Noble 2011	PFN	394.6	88.6	-52.8	absolute difference (95% CI) 24.0 (4.3 to 43.7), p=0.0009
Noble 2011	PBO	404.6	90.4	-76.8	
Martinez 2014	NAC	371	116	-23.4	treatment difference 95% CI, 24.2 (-2.6 to 50.9), p=0.08
Martinez 2014	PBO	375	105	-47.5	

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Safety evidence

Safety was analysed by selecting events (individual or grouped in classes) that satisfied the following criteria in at least one of the experimental studies considered here:

- AEs with a likely significant impact on costs and quality adjusted life-years (QALYs): assumed severe or serious
- AEs with an incidence greater or equal to 5%
- AEs with an incidence 1.5 times greater between the two arms
- AEs of particular focus to clinicians:
    - For NDB: Liver enzyme and bilirubin elevations, gastrointestinal (GI) disorders, arterial thromboembolic events including myocardial infarction, gastrointestinal perforation
    - For PFN: photosensitivity and rash

Based on the systematic literature review and the above criteria two AEs were common across all three comparators; serious cardiac and serious GI events. There were no events that qualified the above criteria and were common across any two comparators.

Four studies present evidence on serious cardiac events (Figure 16). Table 34 includes the relevant data.

Figure 16: Network for serious cardiac events

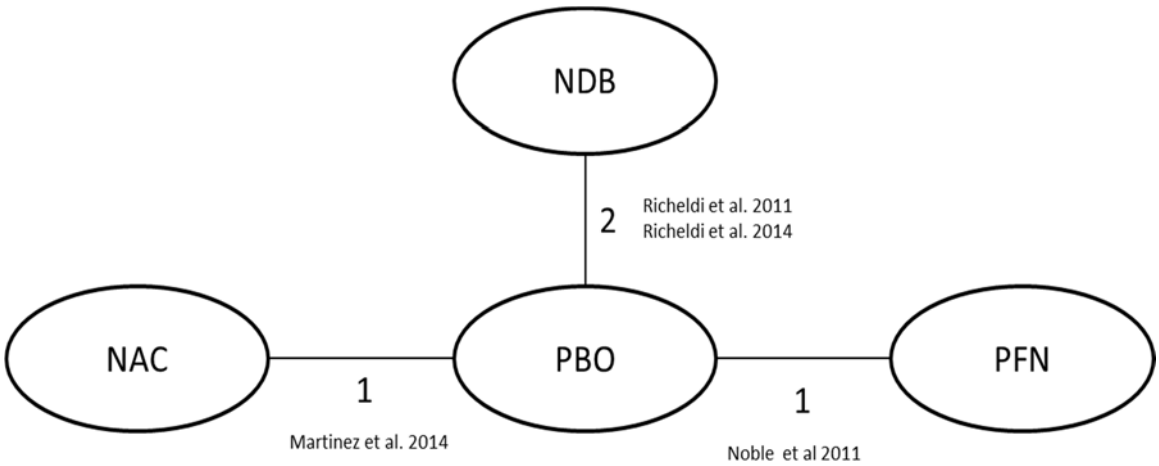


Table 34: Serious cardiac events evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	1	85
Richeldi 2011	PBO	7	85
Richeldi 2014	NDB	32	638
Richeldi 2014	PBO	23	423
Noble 2011	PFN	21	345
Noble 2011	PBO	17	347
Martinez 2014	NAC	9	133
Martinez 2014	PBO	2	131

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Four studies present evidence on serious GI events (Figure 17). Table 35 includes the relevant data.

Figure 17: Network for serious GI

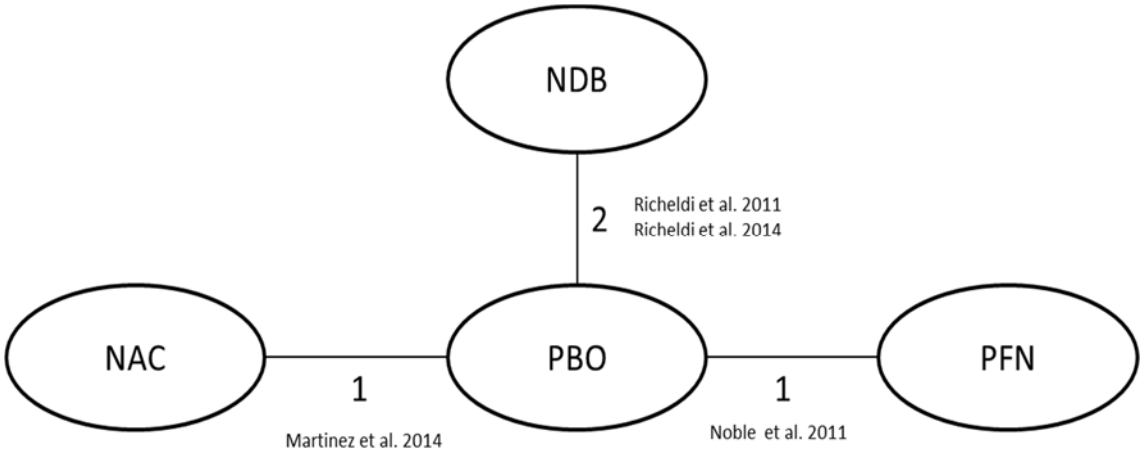


Table 35: Serious GI events evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	4	85
Richeldi 2011	PBO	0	85
Richeldi 2014	NDB	19	638
Richeldi 2014	PBO	7	423
Noble 2011	PFN	254	345
Noble 2011	PBO	173	347
Martinez 2014	NAC	0	133
Martinez 2014	PBO	6	131

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Treatment tolerability

Treatment tolerability was analysed in two ways: discontinuation due to AEs and any discontinuation (overall).

Seven studies included evidence on discontinuation due to AEs (Figure 18). Table 36 presents the relevant data.

Figure 18: Network for discontinuation due to AEs

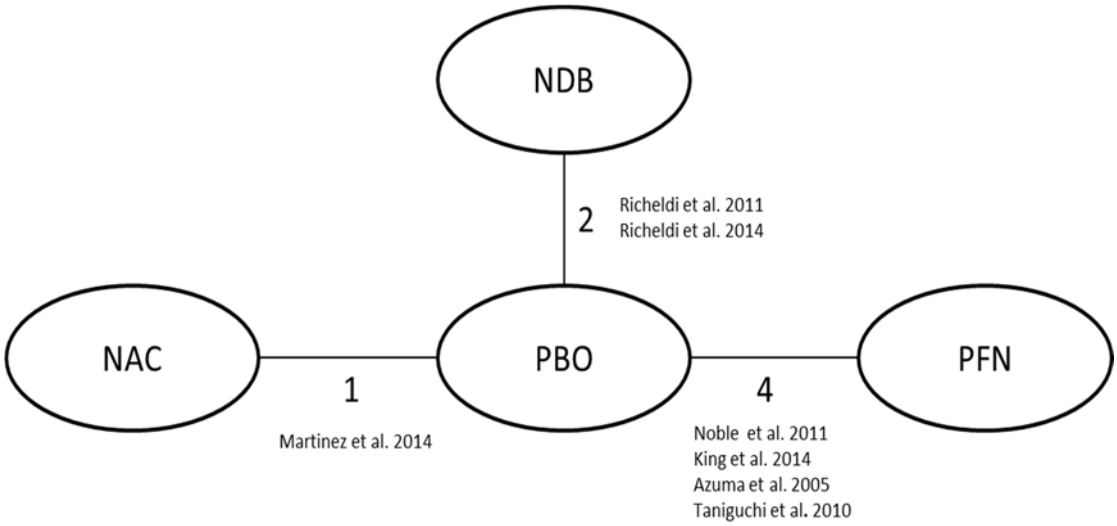


Table 36: Discontinuation due to AEs evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	26	85
Richeldi 2011	PBO	22	85
Richeldi 2014	NDB	123	638
Richeldi 2014	PBO	55	423
Noble 2011	PFN	51	345
Noble 2011	PBO	30	347

Study	Treatment	Observed events	Patients at risk
King 2014	PFN	40	278
King 2014	PBO	30	277
Azuma 2005	PFN	11	73
Azuma2005	PBO	2	36
Taniguchi 2010	PFN	15	109
Taniguchi 2010	PBO	7	107
Martinez 2014	NAC	2	133
Martinez 2014	PBO	3	131

Abbreviations: NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Six studies included evidence on overall discontinuation (Figure 19). Table 37 presents the relevant evidence.

Figure 19: Network for overall discontinuation

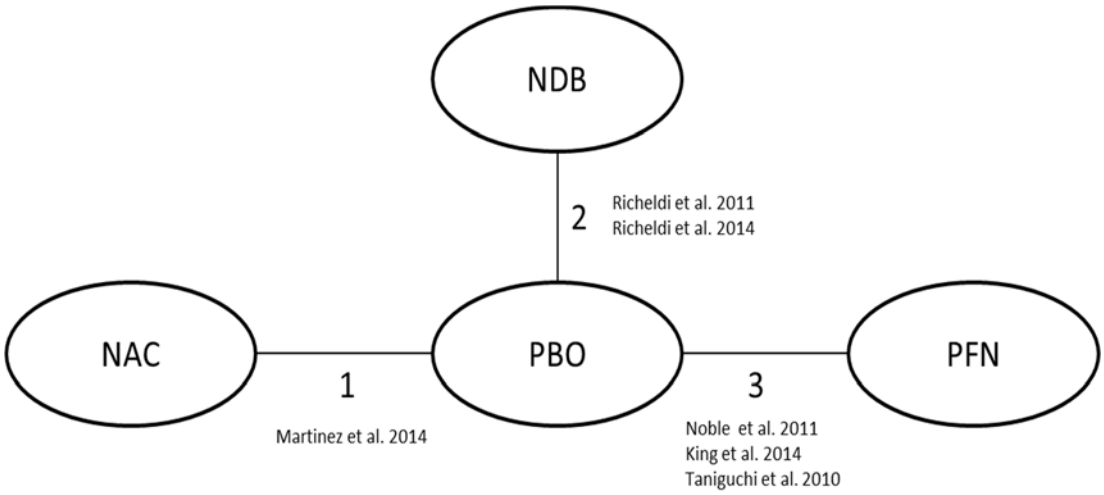


Table 37: Overall discontinuation evidence

Study	Treatment	Observed events	Patients at risk
Richeldi 2011	NDB	32	85
Richeldi 2011	PBO	24	85
Richeldi 2014	NDB	78	638
Richeldi 2014	PBO	36	423
Noble 2011	PFN	72	345
Noble 2011	PBO	62	347
King 2014	PFN	55	278
King 2014	PBO	39	277
Taniguchi 2010*	PFN	40	109
Taniguchi 2010*	PBO	31	107
Martinez 2014	NAC	34	133
Martinez 2014	PBO	29	131

*Values for patients who withdrew/discontinued the study; NDB: nintedanib, PBO: placebo, PFN: pirfenidone, NAC: N-acetylcysteine

Health related quality of life (HRQL)

A comparison was not possible across any intervention. HRQL data for NDB were reported as EQ-5D and SGRQ utility values in Richeldi et al (Richeldi *et al.*, 2014[12], Richeldi, 2011[29]), while data for NAC were reported in Tomioka et al (Tomioka, 2005[53]) only as SF-36 utility values.

Table 38: Summary of evidence considered

Studies	Mortality	Exacerbation rate	Pulmonary function parameters (FVC%Pred)	Progression-free survival (PFS)	Physical function parameters (6MWT)	Safety analysis (serious cardiac and serious GI events)	Treatment tolerability (discontinuation due to AEs and overall)	HRQL
Azuma 2005		✓					✓ only discontinuation due to AEs reported	
Homma 2012		✓						
King 2014	✓		✓	X a comparison was not made	X a comparison was not made		✓	
Martinez 2014		✓	✓		✓	✓	✓	
Noble 2011	✓	✓	✓	✓	✓	✓	✓	
Richeldi 2011	✓	✓	✓		✓	✓	✓	X a comparison was not possible
Richeldi 2014	✓	✓	✓	✓		✓	✓	X a comparison was not possible
Taniguchi 2010	✓	✓		X a comparison was not made			✓	
Tomioka 2005					X a comparison was not possible			X a comparison was not possible

Section 4.10 above provides details of the study design, patient characteristics and key outcomes across the studies identified for the NMA.

Feasibility assessment

After final study selection, eight clinical endpoints were extracted and were considered in the feasibility analysis. These endpoints are described in Table 39 below.

Table 39: Description of the 8 endpoints considered for the NMA

Endpoint	Definition
Overall Survival	All-cause mortality
Acute exacerbations	Acute exacerbations were defined as events meeting all of the INPULSIS criteria, based upon (Collard *et al.*, 2007[61]).
Loss of lung function	10 point decrease in FVC% Predicted by the end of the study follow up. This is considered a clinically important difference (Raghu *et al.*, 2011[13])
Progression free survival	Time to confirmed ≥10% decline in FVC% predicted, confirmed ≥15% decline in DLco% Predicted, or death
6 minute walk test	Distance walked
Serious cardiac and GI events	Serious Adverse events >5% incidence, and 1.5 times difference between two arms.
Discontinuation due to AEs	Stopping therapy due to adverse events of any kind
Overall discontinuation	Discontinuation of therapy for any reason
HRQL	EQ-5D, SF-36, or the St. George Respiratory Questionnaire

Results of the feasibility assessment

***Question 1:* Is there one network of interlinked RCTs to allow the comparisons of interest? If not, is it possible to group certain treatments in order to have one network?**

There was heterogeneity and variation between studies that provided challenges to building a feasible network and assumptions were defined in discussion with Boehringer Ingelheim's clinical team. The main variation across studies was with respect to:

- Lack of reporting in studies of the characteristics that could be potential treatment effect modifiers to enable assessment to be made across the set of studies.
- Variations in the definitions for the outcomes included across studies, for acute exacerbations, and loss of lung function.
- Variation in study duration which could lead to potential biases.

The evidence identified in the systematic literature review (SLR) was sufficient to develop feasible networks for eight of the nine outcomes considered. It was not deemed feasible to create a network for HRQL as no comparable data was available between therapies.

***Question 2:* Are there any differences in study and patient characteristics across comparisons that affect the treatment effects of the interventions of interest relative to a reference treatment?**

The potential treatment effect modifiers were identified in discussion with a clinical expert during the feasibility assessment stage. The feasibility assessment identified some discrepancies between the TOMORROW and INPULSIS trials and the other trials. (See Table 40 for the potential treatment modifiers).

Table 40: Potential treatment effect modifiers identified

Potential treatment effect	Potential treatment effect modifiers
Efficacy and safety	• Treatment dose and route of administration • Age • Smoking status • Duration of disease • Time since diagnosis • FVC% Predicted at baseline • DLco% Predicted at baseline • SpO2% at baseline • 6MWT at baseline • Study duration

Treatment dosing and route of administration

In the phase II TOMORROW study (Richeldi, 2011[29]) patients were assigned to receive one of the following oral doses of nintedanib: 50mg once daily, 50mg BID, 100mg BID, or 150mg BID. In both INPULSIS phase III trials (1199.32 and 1199.34) nintedanib was prescribed orally at 150mg BID (Richeldi *et al.*, 2014[12]).

The pirfenidone CAPACITY programmes included two phase III trials, studies 004 and 006 (Noble, 2011[40]). In CAPACITY study 004 patients were randomly assigned to receive oral pirfenidone 801mg TID or pirfenidone 399mg TID. In CAPACITY study 006 pirfenidone was prescribed orally at 801mg TID. In the ASCEND study (King *et al.*, 2014[41]) patients were assigned to receive oral pirfenidone at 801mg TID. In the SP2 study (Azuma *et al.*, 2005[48]), patients received oral pirfenidone tables at a dose of 200mg TID for the first 2 days, 400mg TID for the following 2 days, and 600mg TID (maximum dose) for the last 3 days. The maximum dose was maintained for patients tolerating it for the remaining duration of the study. Finally, in the SP3 study (Taniguchi, 2010[49]), patients were randomised to receive oral pirfenidone as such: pirfenidone 200mg TID for the first 2 weeks (high dose 600mg per day, low dose 600mg per day), then pirfenidone 400mg TID for the following 2 weeks (high dose 1200mg per day, low dose 600mg per day), and pirfenidone 600mg TID for the remaining 48 weeks (high dose 1800mg per day, low dose 1200mg per day).

In the PANTHER study (Martinez *et al.*, 2014[52]) N-acetylcysteine was prescribed at 600mg orally TID. Patients reported significant dose escalations half-way, and at the end of the study. Homma et al. (Homma, 2012[54]) investigated the use of N-acetylcysteine in a dose of 354mg by inhalation BID, diluted with saline to a total volume of 4mL, and using microair nebulizers and vibration mesh technology (NE-U22, Omron, Tokyo, Japan). Tomioka et al. (Tomioka, 2005[53]) used a similar total dose schedule (354mg inhaled BID) diluted with saline to a total volume of 5mL.

With reference to nintedanib and pirfenidone treatments, the analysis considered the licensed dose of each treatment: 150mg BID for nintedanib and 2,403mg daily for pirfenidone. The alternative treatment arms that investigated lower doses of nintedanib (50mg and 100mg) or pirfenidone (1,197mg and 1,200mg) were not used in the analysis. Sensitivity analysis with inclusion of the lower dose arms was not considered necessary given the available data from the high dose arms and the potential impact of the treatment dose to the relative effect.

The high maintenance pirfenidone dose in the Japanese studies (Azuma *et al.*, 2005[48], Taniguchi, 2010[49]) was reported to be 1,800mg per day, which is different to the European licensed dose (2,403mg per day) (2013[62]). This difference was attributed to the average weights of the Japanese patients. According to Noble et al. (Noble, 2011[40]) the high dose of 2,403mg per day was derived by normalisation of the 1,800mg per day from the Japanese studies to the predicted bodyweights of the predominantly US-based population in the CAPACITY study. As a result the doses considered in the multicentre studies (Noble, 2011[40], King *et al.*, 2014[41]) and the Japanese studies (Azuma *et al.*, 2005[48], Taniguchi, 2010[49]) were determined to be comparable.

N-acetylcysteine use in IPF is off-label and therefore all possible doses (in terms of drug concentration and dose escalation) could be relevant for the analysis. The inhalation route is not commonly used in the UK. To synthesise the trials in the all-evidence network it was necessary to assume that the route of administration did not act as a treatment effect modifier. This was tested in sensitivity analysis by excluding from the network the study that used N-acetylcysteine by inhalation (Homma, 2010[63]).

Patient Characteristics

The covariates considered as possible treatment effect modifiers equally relevant to all investigated outcomes are presented in Table 41 for the nintedanib studies, Table 42 for the pirfenidone studies, and Table 43 for the N-acetylcysteine studies.

Table 41: Population and disease characteristics for the nintedanib clinical trials

Baseline characteristics	TOMORROW (Richeldi *et al.*, 2011[1])					INPULSIS (Richeldi *et al.*, 2014[12])			
						INPULSIS I		INPULSIS II	
	nintedanib 50mg per day	nintedanib 50mg BID	nintedanib 100mg BID	nintedanib 150mg BID	placebo	nintedanib 150mg BID	placebo	nintedanib 150mg BID	placebo
Patients	86	86	86	85	85	309	204	329	219
Age, mean (SD)	65.3 (9.4)	64.9 (8.5)	65.1 (8.6)	65.4 (7.8)	64.8 (8.6)	66.9 (8.4)	66.9 (8.2)	66.4 (7.9)	67.1 (7.5)
Smoking, N (%)									
Non-smoker	27 (31.4)	29 (33.7)	32 (37.2)	25 (29.4)	28 (32.9)	71 (23.0)	51 (25.0)	103 (31.3)	71 (32.4)
Ex-smoker	55 (64.0)	53 (61.6)	52 (60.5)	58 (68.2)	51 (60.0)	217 (70.2)	144 (70.6)	218 (66.3)	139 (63.5)
Current smoker	4 (4.7)	4 (4.7)	2 (2.3)	2 (2.4)	6 (7.1)	21 (6.8)	9 (4.4)	8 (2.4)	9 (4.1)
Duration of disease, N (%)									
<1 year	48 (55.8)	55 (64.0)	50 (58.1)	57 (67.1)	48 (56.5)	133 (43.0	92 (45.1)	141 (42.9)	101 (46.1)
1-3 years	25 (29.1)	23 (26.7)	25 (29.1)	21 (24.7)	21 (24.7)	115 (37.2	75 (36.8)	128 (38.9)	79 (36.1)
>3 years	13 (15.2)	8 (9.3)	11 (12.8)	7 (8.2)	16 (18.8)	61 (19.8)	37 (18.1)	60 (18.2)	39 (17.8)
Time since diagnosis, mean (SD)	1.4 (1.3)	1.1 (1.2)	1.2 (1.2)	1.0 (1.2)	1.4 (1.5)	1.7 (1.4)	1.6 (1.4)	1.6 (1.3)	1.6 (1.3)
FVC%Pred, mean (SD)	80.4 (17.8)	79.8 (15.8)	85.5 (19.2)	79.1 (18.5)	81.7 (17.6)	79.5 (17.0)	80.5 (17.3)	80.0 (18.1)	78.1 (19.0)
DL_{CO}%Pred, mean (SD)	46.2 (13.6)	46.6 (12.7)	48.7 (12.8)	47.5 (11.0)	48.4 (12.9)	47.8 (12.3)	47.5 (11.7)	47.0 (14.5)	46.6 (14.8)
SpO_2%, mean (SD)	95.0 (2.7)	95.4 (2.2)	95.3 (2.0)	95.6 (1.7)	95.3 (2.2)	95.9 (2.0)	95.9 (1.9)	95.8 (2.6)	95.7 (2.1)
6MWT, meters (SD)	422.5 (12.75)	409.3 (11.63)	429.8 (13.25)	437.0 (13.69)	411.1 (15.90)	N/R	N/R	N/R	N/R

Table 42: Population and disease characteristics for the pirfenidone clinical trials

Baseline characteristics	CAPACITY (Noble, 2011[40])					ASCEND (King *et al.*, 2014[41])		SP2 (Azuma *et al.*, 2005[48])		SP3 (Taniguchi, 2011[55])		
	Study 004			Study 006								
	pirfenidone 1197mg per day	pirfenidone 2403mg per day	placebo	pirfenidone 2403mg per day	placebo	pirfenidone 2403mg per day	placebo	pirfenidone 1800mg per day	placebo	pirfenidone 1800mg per day	pirfenidone 1200mg per day	placebo
Patients	87	174	174	171	173	278	277	72	35	108	55	104
Age, mean (SD)	68.0 (7.6)	65.7 (8.2)	66.3 (7.5)	66.8 (7.9)	67.0 (7.8)	68.4 (6.7)	67.8 (7.3)	64.0 (7.1)	64.3 (7.6)	65.4 (6.2)	63.9 (7.5)	64.7 (7.3)
Smoking, N (%)												
Non-smoker	27 (31)	56 (32)	51 (29)	59 (35)	64 (37)	N/R	N/R	8 (11)	2 (6)	22 (20.4)	12 (21.8	21 (20.2)
Ex-smoker	57 (66)	110 (63)	114 (66)	112 (65)	101 (58)	184 (66)	169 (61)	57 (79)	30 (86)	81 (75.0)	33 (60.0)	70 (67.3)
Current smoker	3 (3)	8 (5)	9 (5)	0 (0)	8 (5)	N/R	N/R	7 (10)	3 (9)	5 (4.6)	10 (18.2)	13 (12.5)
Duration of disease, N (%)												
<1 year	46 (53)	83 (48)	81 (47)	100 (58)	107 (62)	N/R	N/R	20 (28)	6 (17)	38 (35.2)	20 (36.4)	41 (39.4)
1-3 years	N/R	N/R	N/R	N/R	N/R	N/R	N/R	17 (24)	10 (29)	29 (26.9)	13 (23.6)	25 (24.0)
>3 years	N/R	N/R	N/R	N/R	N/R	N/R	N/R	35 (49)	19 (54)	41 (38.0)	22 (40.0)	38 (36.5)
Time since diagnosis, mean (SD)	N/R	N/R	N/R	N/R	N/R	1.7 (1.1)	1.7 (1.1)	N/R	N/R	N/R	N/R	N/R
FVC%Pred, mean (SD)	76.4 (14.4)	74.5 (14.5)	76.2 (15.5)	74.9 (13.2)	73.1 (14.2)	67.8 (11.2)	68.6 (10.9)	81.6 (20.3)[i]	78.4 (17.2)[i]	77.3 (16.8)[i]	76.2 (18.7)[i]	79.1 (17.4)[i]
DL_{CO}%Pred, mean (SD)	47.2 (8.2)	46.4 (9.5)	46.1 (10.2)	47.8 (9.8)	47.4 (9.2)	43.7 (10.5)	44.2 (12.5)	57.6 (17.2)	57.7 (13.8)	52.1 (16.8)	53.6 (19.1)	55.2 (18.2)
SpO_2%, mean (SD)	N/R	N/R	N/R	N/R	N/R	N/R	N/R	87.1 (3.9)[ii]	87.1 (4.2)[ii]	89.0 (2.3)[iii]	88.8 (2.4)[iii]	89.0 (2.0)[iii]
6MWT, meters (SD)	417 (112.8)	411.1 (91.8)	410.0 (90.9)	378 (82.2)	399.1 (89.7)	415.0 (98.5)	420.7 (98.1)	N/R	N/R	N/R	N/R	N/R

[i] Note this outcome is VC%Pred rather than FVC%Pred; [ii] Lowest SpO_2% during 6MET; [iii] Lowest SpO_2%

Table 43: Population and disease characteristics for the N-acetylcysteine clinical trials

Baseline characteristics	PANTHER (Martinez *et al.*, 2014[52])		(Homma, 2012[54])		(Tomioka, 2005[53])	
	N-acetylcysteine 1800mg per day	placebo	N-acetylcysteine 352mg per day	placebo	N-acetylcysteine 352mg per day	Bromhexine 4mg per day
Patients	133	131	38	38	10	12
Age, mean (SD)	68.3 (8.4)	67.2 (8.2)	67.6 (6.4)	68.2 (7.7)	70 (4.9)	70 (5.3)
Smoking, N (%)						
Non-smoker	36 (27.3)	33 (25.2)	10 (26.3)	9 (23.7)	5 (50)	3 (25)
Ex-smoker	N/R	N/R	25 (65.8)	26 (68.4)	4 (40)	5 (41.7)
Current smoker	N/R	N/R	10 (26.3)	3 (7.9)	1 (10)	4 (33.3)
Duration of disease, N (%)						
<1 year	N/R	N/R	9 (23.7)	5 (13.2)	N/R	N/R
1-3 years	N/R	N/R	17 (44.7)	14 (36.8)	N/R	N/R
>3 years	N/R	N/R	12 (31.6)	19 (50)	N/R	N/R
Time since diagnosis, mean (SD)	1.0 (1.0)	1.1 (1.0)	3.0 (3.4)	3.2 (2.5)	N/R	N/R
FVC%Pred, mean (SD)	72.2 (15.9)	73.4 (14.3)	89.2 (17.8)	88.7 (15.5)	67.6 (15.7)[iii]	76.6 (19.1)[iii]
DL_{CO}%Pred, mean (SD)	44.7 (10.8)	46.0 (12.2)	72.3 (25.3)	64.4 (20.1)	64.7 (21.7)	60.7 (16.7)
SpO_2%, mean (SD)	95.75 (2.45)[i]	96.12 (2.30)[i]	93.1 (2.1)[ii]	92.4 (2.0)[ii]	90.1 (5.9)[iv]	91.1 (5.9)[iv]
6MWT, meters (SD)	371 (116)	375 (105)	N/R	N/R	385 (90)	390 (116)

[i] Resting SpO_2%; [ii] Lowest SpO_2 during a 6MWT; [iii] VC%Pred rather than FVC%Pred; [iv] Lowest SaO_2 during 6-min walking test (%)

.

The ASCEND (King *et al.*, 2014[41]) cohort had a lower average FVC%Pred than the other studies. The authors stated that the ASCEND selection criteria were chosen to "[…] allow enrolment of patients with an increased risk of disease progression" (King *et al.*, 2014[41]). Among other inclusion criteria they reported: a) IPF diagnosed at least 0.5 years before visit 2, b) FVC 50-90% predicted and c) FEV1/FVC ≥ 0.8.

There were four studies that reported the results of trials in Japanese patients. Three studies (SP2 (Azuma *et al.*, 2005[48]), SP3 (Taniguchi, 2011[55]), and Homma, 2005 (Homma, 2005[64])) reported a longer disease duration when compared to the other studies. One study, Tomioka et al (Tomioka, 2005[53]) did not report disease duration. The balance between smokers and non-smokers in SP2 was also different compared to the other studies, with the majority of patients being former smokers. Homma et al (Homma, 2005[64]) also reported the highest FVC%Pred scores of any study. The discrepancies in the average disease duration in the Japanese studies could be an important treatment effect modifier in IPF where patients can have a median survival of 3-5 years (Raghu *et al.*, 2011[13]).

As ASCEND, SP2, SP3, and Homma et al had potentially significant differences in terms of potential treatment modifiers, this was tested by excluding these studies in sensitivity analyses.

Outcomes assessed in the NMA

Overall Survival

A HR or relative risk (RR) and confidence interval was available for nintedanib (Richeldi *et al.*, 2011[1], Richeldi *et al.*, 2014[12]) and pirfenidone (Noble, 2011[40], King *et al.*, 2014[41]). Azuma et al. (Azuma *et al.*, 2005[48]) reported patient mortality after the onset of an acute exacerbation event. Taniguchi et al. (Taniguchi, 2010[49]) and Martinez et al. (Martinez *et al.*, 2014[52]) reported only the total number of patients dead at the end of the study; without any further measures of uncertainty. Tomioka et al. (Tomioka, 2005[53]) excluded from the study patients who had respiratory failure death.

The available evidence did not allow the assessment of overall survival as a continuous outcome. In order to perform an NMA between the three interventions of interest a dichotomous outcome was used for mortality.

To synthesise the above trial in the all-evidence analysis we assumed that the mortality reported in Azuma et al. (Azuma *et al.*, 2005[48]) reflected all-cause mortality. We tested this assumption in sensitivity analysis.

Acute exacerbation

The definition of acute exacerbation was not consistently reported in all publications, and when reported the definition varied across studies.

In TOMORROW (Richeldi *et al.*, 2011[1]) the definition of acute exacerbation was available from the clinical study report. Exacerbation was defined as otherwise unexplained clinical features within one month including all of the following:

- Progression of dyspnoea over several days to 4 weeks
- New diffuse pulmonary infiltrates on chest X-ray, and/or new HRCT parenchymal abnormalities with no pneumothorax or pleural effusion (new ground-glass opacities) since last visit
- Decrease in PaO2 ≥ 10 mmHg or PaO2/FiO2 <225 since last visit
- Exclusion of infection as per routine clinical practice and microbiological studies
- Absence of congestive heart failure, pulmonary embolism, etc.

In INPULSIS (Richeldi *et al.*, 2014[12]) acute exacerbations were defined as events meeting all of the following criteria: unexplained worsening or development of dyspnoea within the previous 30 days; new diffuse pulmonary infiltrates visualised on chest radiography, HRCT, or both, or the development of parenchymal abnormalities with no pneumothorax or pleural effusion (new ground-glass opacities) since the preceding visit; and exclusion of any known causes of acute worsening, including infection, left heart failure, pulmonary embolism, and any identifiable cause of acute lung injury, in accordance with routine clinical practice and microbiologic studies. All acute exacerbations reported by the site investigators were categorised by the adjudication committee, whose members were unaware of the study group assignments, as confirmed or suspected or were not considered to be an acute exacerbation according to pre-specified criteria.

In CAPACITY (Noble, 2011[40]) acute IPF exacerbation was defined as an event within a 4-week period where a patient developed evidence of all of the following criteria:

1. Worsening of PaO2 (≥8mmHg drop from most recent pre-worsening value)

2. Clinically significant worsening of dyspnoea

3. New, superimposed ground-glass opacities of HRCT in ≥1 lobe

4. All other causes such as cardiac, thromboembolic, aspiration or infectious processes have been ruled out.

In the ASCEND study (King *et al.*, 2014[41]), acute exacerbations were not reported.

In the Azuma et al. (Azuma *et al.*, 2005[48]) study, acute exacerbation of IPF was defined as a manifestation of all of the following: worsening, otherwise unexplained clinical features within 1 month, progression of dyspnoea over a few days to less than 5 weeks, new radiographic/HRCT parenchymal abnormalities without pneumothorax or pleural effusion (e.g., new, superimposed ground-glass opacities), a decrease in the PaO2 by 10mmHg or more, and exclusion of apparent infection based on absence of Aspergillus and pneumococcus antibodies in blood, urine for Legionella pneumophila, and sputum cultures.

In the Taniguchi et al. (Taniguchi, 2010[49]) study, the diagnostic criteria of acute exacerbation of IPF were:

- All of the following three conditions had to be satisfied during the disease progression of IPF within a month:
    1. Dyspnoea increases
    2. New ground-glass opacities appear on HRCT in addition to previous honeycomb lesions
    3. Oxygen partial pressure in resting arterial blood (PaO2) is lower by more than 10Torr than previous one
- Exclude obvious causes, such as infection, pneumothorax, cancer, pulmonary embolism, or congestive heart failure
- The serum levels of CRP, LDH are usually elevated as well as serum markers of interstitial pneumonias, such as KL-6, Sp-A or Sp-D

In the PANTHER study (Martinez *et al.*, 2014[52]), the pre-specified definition of acute exacerbation was in accordance with published criteria:

- Aggravation of dyspnoea within one month
- Hypoxemia with an arterial oxygen tension/inspired oxygen tension ratio of <225
- Newly developing pulmonary infiltrates on chest radiography
- Absence of apparent infection or heart disease

Homma et al. (Homma, 2012[54]) and Tomioka et al. (Tomioka, 2005[53]) did not provide any definitions for the acute exacerbation events reported.

The differences in the definition of acute exacerbation are likely to introduce bias to the analysis; results of this comparison should be treated with caution.

Loss of lung function

Loss of lung function was defined as a 10-points decrease in FVC%Pred by the end of the study follow-up. FVC%Pred is an indicator of lung function, a widely used measure of disease status, and a common endpoint in clinical trials. A 10% reduction in

FVC%Pred is reported in the literature as a clinically important difference (Raghu *et al.*, 2011[13]).

Richeldi et al. (Richeldi *et al.*, 2011[1]), Richeldi et al. (Richeldi *et al.*, 2014[12]), and Noble et al. (Noble, 2011[40]) provided evidence on loss of lung function.

In King et al. (King *et al.*, 2014[41]) progression was presented as the cumulative number of patients dead or experiencing a decline of 10%-points or more in FVC%Pred or death at 13, 26, 39, and 52 weeks compared to baseline.

In Martinez et al. (Martinez *et al.*, 2014[52]) progression was defined as a reduction of more than 10% in FVC or death. Since FVC%Pred is estimated based on a standardisation of FVC for age, gender and height, it was reasonable to assume that the categorical outcome of the percentage of patients who experienced a drop in either FVC or FVC%Pred would be similar if not the same.

Loss of lung function defined at least as a 10-points decrease in FVC%Pred or FVC was not reported in Azuma et al. (Azuma *et al.*, 2005[48]), Taniguchi et al. (Taniguchi, 2010[49]) Homma et al. (Homma, 2012[54]), and Tomioka et al. (Tomioka, 2005[53]).

Unfortunately, not all interventions reported loss of lung function as defined above. The significance of the inclusion of death as a composite outcome for loss of lung function was tested in sensitivity analysis.

Progression free survival

In order to perform a comparison on this outcome, individual patient data from the Richeldi et al. (Richeldi *et al.*, 2014[12]) study were analysed to replicate the PFS outcome definition from Noble et al. (Noble, 2011[40]). Therefore, by design, these outcomes are comparable across Richeldi et al. (Richeldi *et al.*, 2014[12]) and Noble et al. (Noble, 2011[40]).

6MWT

No differences in the definition of the outcome were observed across the studies reporting 6MWT (Richeldi et al. 2011 (Richeldi *et al.*, 2011[1]), Noble et al. 2011 (Noble, 2011[40]), Martinez et al. (Martinez *et al.*, 2014[52])). Note that although King et al. (King *et al.*, 2014[41]) and Tomioka et al. (Tomioka, 2005[53]) reported point estimates for this outcome, no precision data were available, and a comparison with the other studies was not possible.

Serious cardiac and GI events

For the purposes of the NMA, we only looked at serious adverse events (SAEs) that met the criteria of over 5% incidence, and 1.5 times difference between the two arms.

Serious cardiac events and serious GI events were reported in Richeldi et al. (Richeldi *et al.*, 2011[1]), Richeldi et al. (Richeldi *et al.*, 2014[12]), Noble et al. (Noble, 2011[40]), and Martinez et al. (Martinez *et al.*, 2014[52]). The SAE data for pirfenidone were extracted from the Manufacturer's Submission (MS) report (NICE, 2013[17]) for the CAPACITY trial. The serious events were grouped within a larger, more encompassing system organ class (SOC) for cardiac and GI. Across studies, each class may contain different events. This analysis is not addressing the heterogeneity in the composition of the SAE classes.

Discontinuation due to AEs

No differences in the definition of the outcome were observed across the studies reporting discontinuation due to AEs (Richeldi et al. (Richeldi *et al.*, 2011[1]), Richeldi et al. (Richeldi *et al.*, 2014[12]), Noble et al. (Noble, 2011[40]), King et al. (King *et al.*, 2014[41]), Azuma et al. (Azuma *et al.*, 2005[48]), Taniguchi et al. (Taniguchi, 2010[49]), and Martinez et al. (Martinez *et al.*, 2014[52])).

Overall discontinuation

Five studies report discontinuation from treatment or discontinuation from study medication (Richeldi et al. (Richeldi *et al.*, 2011[1]), Richeldi et al. (Richeldi *et al.*, 2014[12]), Noble et al. (Noble, 2011[40]), King et al. (King *et al.*, 2014[41]), and Martinez et al. (Martinez *et al.*, 2014[52])). Taniguchi et al. (Taniguchi, 2010[49]) defined this outcome as withdrawal or discontinuation from the study. For the purposes of the NMA the outcomes were considered similar in the all evidence analysis. Taniguchi et al. (Taniguchi, 2010[49]) was excluded in a scenario analysis to test this assumption.

Study follow-up differences

The duration of follow-up varied across most studies, as shown in Table 44 below.

Table 44: Study follow-up duration

Study	Duration of follow-up
Richeldi 2011 (Richeldi, 2011[29])	52 weeks
Richeldi 2014 (Richeldi *et al.*, 2014[12])	52 weeks
Noble 2011 (Noble, 2011[40])	72 weeks
King 2014 (King *et al.*, 2014[41])	52 weeks
Azuma 2005 (Azuma *et al.*, 2005[48])	36 weeks
Taniguchi 2010 (Taniguchi, 2010[49])	52 weeks
Martinez 2014 (Martinez *et al.*, 2014[52])	60 weeks
Homma 2012 (Homma, 2012[54])	48 weeks
Tomioka 2005 (Tomioka, 2005[53])	12 months

The majority of the studies had a follow-up duration of approximately 1 year. Noble *et al.* (Noble, 2011[40]) and Martinez *et al.* (Martinez *et al.*, 2014[52]) had a longer follow-up duration, while Azuma *et al.* (Azuma *et al.*, 2005[48]) had a shorter duration.

The discrepancy in the study follow-up duration may have introduced bias in the analysis. Nevertheless, the NMA compared the relative efficacy of each active arm vs. the study control. No analysis was undertaken to explore the bias due to study follow-up duration.

Critical appraisal of NMA studies

Study characteristics were assessed for systematic error (biases) based on the Cochrane Collaboration Bias Methods (Cochrane Bias Methods Group, [Accessed October 2014][65]). The risk of bias (low, moderate, high) is presented in Table 28.

A NMA in the form of an ITC was performed for these outcomes:

- Overall survival
- Acute exacerbations
- Loss of lung function
- Progression free survival
- 6MWT
- Serious Cardiac events
- Serious GI events

The results presented below include all studies as presented in the network diagrams above. For scenarios where studies were excluded, please see Appendix B.

Models

The network analysis was implemented in a Bayesian framework, using both fixed-effect and random-effect models.

Homogeneity of the variances is assumed in the random-effect approach, meaning that the parameter σ2 is the same for all treatment comparisons. The variance parameter was given a vague prior, with the standard deviation following a uniform distribution. Vague non-informative normal priors were given to all other parameters.

Implementation

The models were implemented in WinBugs, version 1.4.3 using two chains with different sets of initial values. For each analysis, an initial 20,000 iterations was run. If convergence has not been achieved after this, the burn-in period was extended until convergence was achieved. Auto-correlations between results from successive simulations were checked. In case of the presence of such correlations, a thinning process was applied. Results were based on a further 50,000 iterations.

Both fixed-effect and random-effect models were run for each outcome and the most appropriate model in each case was selected based on the Deviance Information Criterion (DIC).

Reporting and interpretation of results

For each outcome, results were reported for:

- OR of each treatment vs. placebo: median and associated 95% credible interval
- OR of nintedanib vs. each treatment: median and associated 95% credible interval
- OR of each treatment vs. nintedanib: median and associated 95% credible interval
- Ranking matrix

A 95% credible interval can be interpreted intuitively as the range of values within which the parameter has a 95% probability of falling. For example, in this analysis, if the 95% credible interval does not include 1, it can be interpreted as there being less than 5% chance that there is no difference between those two treatments. In other words, if the credible interval does not include 1 then the two treatments can be considered as significantly different. Conversely, if the credible interval includes 1 then it cannot be concluded that the difference between the two treatments is statistically significant. The ranking matrix is populated by the proportion of simulations in which each treatment is ranked 'best', 'second best' etc. in terms of relative efficacy.

Results of the NMA on overall survival

Table 45 presents the scenarios considered after the feasibility assessment for overall survival. Please note that only the all evidence scenario is presented in the text below. All other scenarios are presented in Appendix B.

Table 45: Scenarios considered for overall survival

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding King, Azuma and Taniguchi	Excluded studies due to differences in patient characteristics
3	Excluding Richeldi 2011, King, Azuma and Taniguchi	Similar to scenario 2, excluded Richeldi 2011 (phase II) as well to consider only phase III data for nintedanib
4	Excluding Azuma	Excluded study due to bias in outcome reporting
5	Excluding Azuma and Taniguchi	Excluded Azuma 2005 and Taniguchi 2010 (Japanese populations)

Scenario 1: all evidence

Table 46, Table 47, and Table 48 present the results of the pairwise comparisons. Table 49 presents the NMA results. Table 50 shows the NMA ranking matrix for the

fixed effects and random effects models. Figure 20 and Figure 21 present the results of the NMA in a forest plot for the fixed effects and the random effects model respectively.

No heterogeneity was observed in the pairwise meta-analyses.

In the NMA, the total residual deviance approximated the available number of data points (14), indicating a good fit for both models. Based on the DIC, the fixed effects performed better than the random effects model.

None of the results were statistically significant. In terms of point estimates, the results indicated that nintedanib and pirfenidone have similar efficacy, better than placebo and N-acetylcysteine. N-acetylcysteine was estimated to have worse OS than placebo; however, the credible interval was very wide.

Table 46: Nintedanib vs. placebo; scenario 1 OS

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	0.76 (0.27, 2.14)	18.0	18.4
Richeldi 2014	0.69 (0.42, 1.12)	82.0	81.6
Pooled OR (95% CI)		0.70 (0.45, 1.09)	0.70 (0.45, 1.09)
p-value (test OR=1)		p=0.115	p=0.114
Heterogeneity	I^2=0%, chi-square p-value=0.865		

Table 47: Pirfenidone vs. placebo, scenario 1 OS

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Noble 2011	0.78 (0.46, 1.33)	56.7	61.1
Azuma 2005	1.51 (0.06, 38.00)	1.2	1.6
Taniguchi 2010	0.71 (0.16, 3.27)	7.2	7.4
King 2014	0.53 (0.25, 1.13)	34.9	29.9
Pooled OR (95% CI)		0.70 (0.46, 1.05)	0.70 (0.46, 1.06)
p-value (test OR=1)		p=0.085	p=0.089
Heterogeneity	I^2=0%, chi-square p-value=0.823		

Table 48: N-acetylcysteine vs. placebo; scenario 1 OS

Study	OR (95% CI)	p-value
Martinez 2014	2.02 (0.49, 8.24)	0.329

Table 49: NMA; scenario 1 OS

	FIXED-EFFECTS (DIC=74.2, tot res dev 11.4)	RANDOM-EFFECTS (DIC=76.1, tot res dev 12.1)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	0.70 (0.45, 1.10)	0.70 (0.25, 2.02)
pirfenidone vs. placebo	0.70 (0.46, 1.05)	0.70 (0.32, 1.87)
N-acetylcysteine vs. placebo	2.15 (0.53, 11.06)	2.15 (0.33, 17.54)
All vs. nintedanib		
placebo vs. nintedanib	1.43 (0.91, 2.24)	1.42 (0.50, 4.06)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	1.00 (0.54, 1.83)	1.00 (0.28, 4.32)
N-acetylcysteine vs. nintedanib	3.08 (0.71, 16.74)	3.07 (0.36, 32.32)
Nintedanib vs. all		
nintedanib vs. placebo	0.7 (0.45, 1.1)	0.70 (0.25, 2.02)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	1.00 (0.55, 1.85)	1.00 (0.23, 3.59)
nintedanib vs. N-acetylcysteine	0.32 (0.06, 1.41)	0.33 (0.03, 2.79)

Figure 20: Forest plot; scenario 1 OS fixed effects model

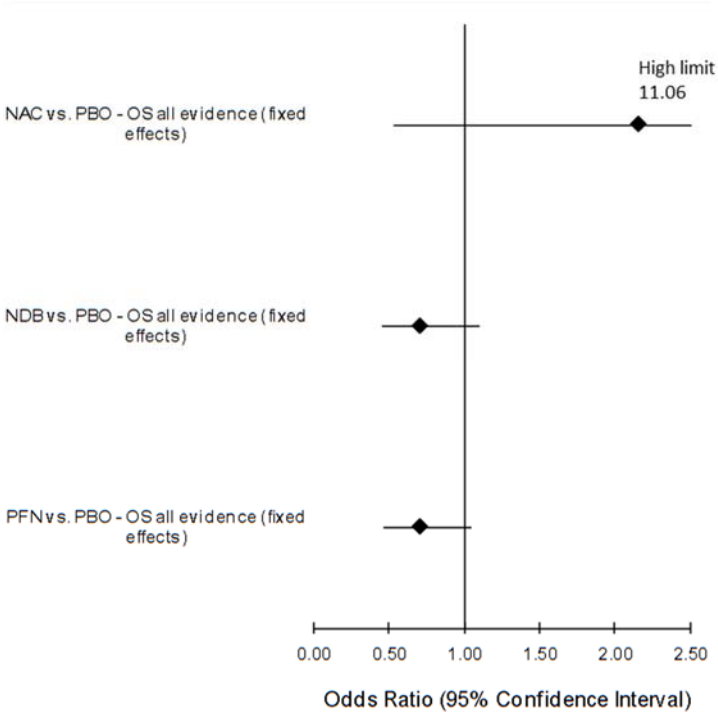


Figure 21: Forest plot; scenario 1 OS random effects model

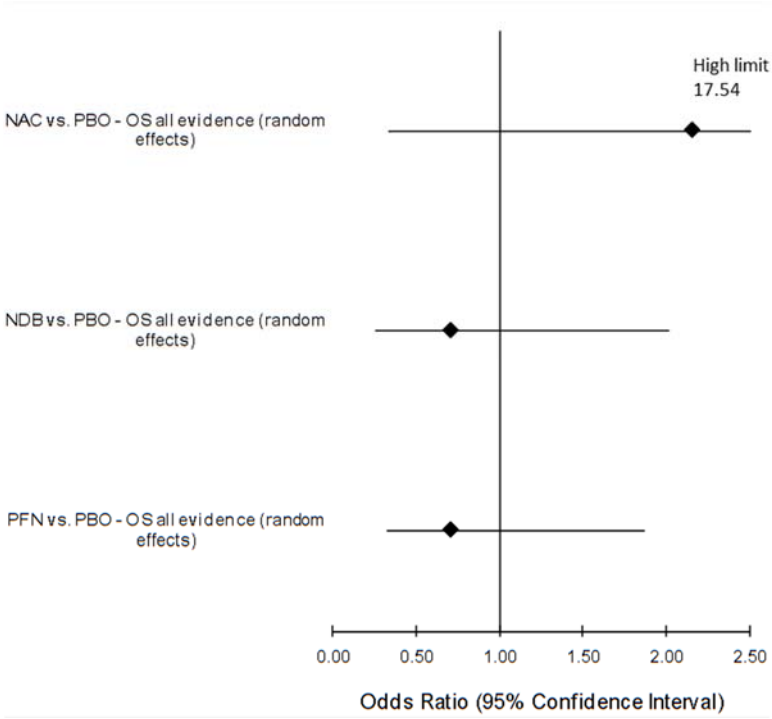


Table 50: Ranking matrix; scenario 1 OS

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	0.2%	8.3%	78.3%	13.2%	2.7%	19.9%	64.9%	12.6%
nintedanib	47.6%	43.1%	8.3%	1.0%	45.1%	35.5%	14.4%	5.0%
pirfenidone	47.6%	44.6%	7.1%	0.7%	44.7%	38.2%	13.0%	4.1%
N-acetylcysteine	4.5%	4.0%	6.4%	85.1%	7.6%	6.4%	7.7%	78.3%

*1: best, 4: worst

Results of NMA on acute exacerbations

Table 51 presents the scenarios considered for acute exacerbations. Please note that only the all evidence scenario is presented in the text below. All other scenarios are presented in Appendix B.

Table 51: Scenarios considered for acute exacerbations

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding Homma	Excluded study due to different route of drug administration and high study bias
3	Excluding Azuma, Taniguchi and Homma	Excluded studies due to differences in patient characteristics – Japanese populations
4	Excluding Richeldi 2011, Azuma, Taniguchi and Homma	Similar to scenario 3, excluded Richeldi 2011 (phase II) to consider only phase III data for NDB
5	Excluding Azuma	Excluded study because it introduced heterogeneity in the results

Scenario	Description	Reasons of choice
6	Excluding Azuma and Homma	Similar to scenario 5, excluded Homma as well because of high study bias

Scenario 1: all evidence

Heterogeneity was observed in the pairwise meta-analysis of the pirfenidone studies (Table 53). We attributed this to the study by Azuma et al. (Azuma *et al.*, 2005[48]). This study was excluded from the all-evidence analysis in a separate scenario (see Appendix B).

In the NMA, the total residual deviance of the fixed effects model did not approximate the available number of data points (14), indicating a poor fit. Based on the DIC, the random effects performed better than the fixed effects model. Nevertheless, in the random effects model the credible interval for all comparators indicated a high level of uncertainty around the point estimates making any interpretation difficult.

None of the results were statistically significant.

Table 52: Nintedanib vs. placebo; scenario 1 acute exacerbations

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	0.33 (0.10, 1.09)	22.3	15.6
Richeldi 2014	0.62 (0.38, 1.04)	77.7	84.4
Pooled OR (95% CI)		0.56 (0.35, 0.89)	0.57 (0.35, 0.89)
p-value (test OR=1)		p=0.014	p=0.017
Heterogeneity	I^2=0%, chi-square p-value=0.338		

Table 53: Pirfenidone vs. placebo; scenario 1 acute exacerbations

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Noble 2011	1.01 (0.25, 4.06)	24.5	39.1
Azuma 2005	0.04 (0.00, 0.71)	45.5	18.6
Taniguchi 2010	1.17 (0.34, 3.94)	29.9	42.3
Pooled OR (95% CI)		0.61 (0.28, 1.34)	0.58 (0.123, 2.69)
p-value (test OR=1)		p=0.220	p=0.490
Heterogeneity	I^2=60.0%, chi-square p-value=0.082		

Table 54: N-acetylcysteine vs. placebo; scenario 1 acute exacerbations

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Martinez 2014	0.99 (0.20, 4.97)	43.6	65.5
Homma 2012	0.24 (0.03, 2.28)	56.4	34.5
Pooled OR (95% CI)		0.57 (0.16, 1.98)	0.61 (0.16, 2.26)
p-value (test OR=1)		p=0.373	p=0.458
Heterogeneity	I^2=0%, chi-square p-value=0.319		

Table 55: NMA; scenario 1 acute exacerbations

	FIXED-EFFECTS (DIC=78.0, tot res dev 21.8)	RANDOM-EFFECTS (DIC=75.50, tot res dev 14.6)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	0.56 (0.35, 0.89)	0.47 (0.01, 15.96)
pirfenidone vs. placebo	0.59 (0.24, 1.35)	0.37 (0.01, 4.81)
N-acetylcysteine vs. placebo	0.54 (0.13, 1.91)	0.47 (0.01, 17.32)
All vs. nintedanib		
placebo vs. nintedanib	1.79 (1.12, 2.85)	2.13 (0.06, 78.2)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	1.05 (0.39, 2.74)	0.82 (0.00, 52.03)
N-acetylcysteine vs. nintedanib	0.97 (0.22, 3.70)	1.02 (0.01, 158.4)
nintedanib vs. all		
nintedanib vs. placebo	0.56 (0.35, 0.89)	0.47 (0.01, 15.96)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	0.96 (0.36, 2.58)	1.22 (0.02, 257)
nintedanib vs. N-acetylcysteine	1.04 (0.27, 4.51)	0.98 (0.01, 189.3)

Figure 22: Forest plot; scenario 1 acute exacerbations fixed effects

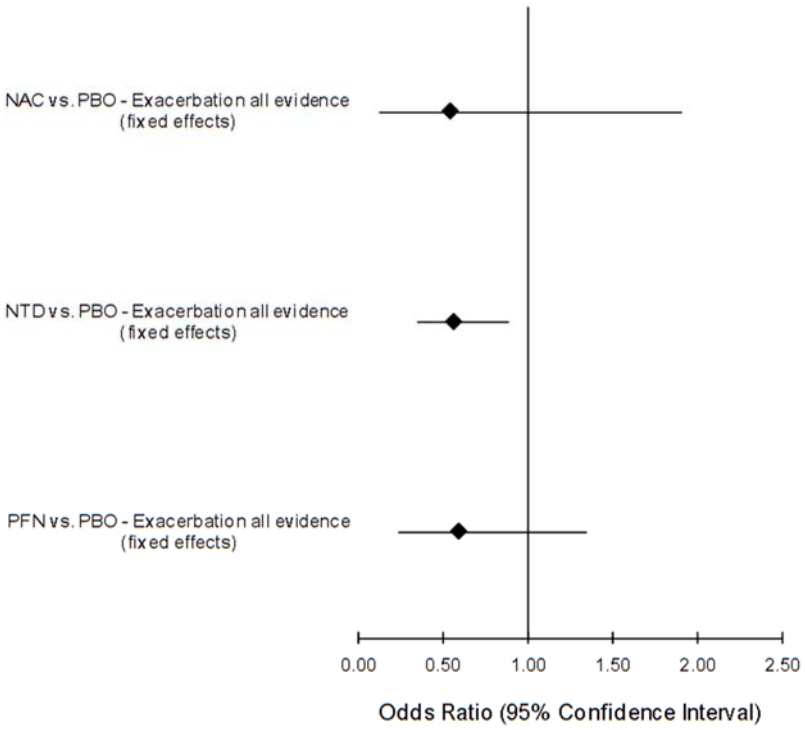


Figure 23: Forest plot; scenario 1 acute exacerbations random effects

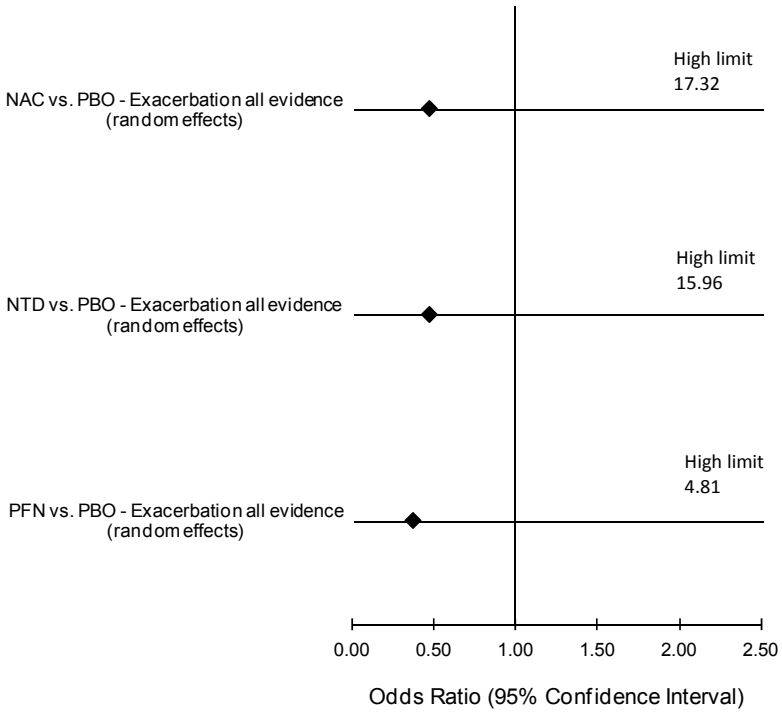


Table 56: Ranking matrix; scenario 1 acute exacerbations

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	0.0%	1.9%	24.4%	73.7%	1.5%	13.0%	39.7%	45.8%
nintedanib	28.3%	44.9%	26.2%	0.6%	29.7%	30.1%	21.3%	18.9%
pirfenidone	29.3%	32.6%	28.7%	9.4%	37.1%	31.9%	19.4%	11.7%
N-acetylcysteine	42.4%	20.6%	20.7%	16.4%	31.8%	25.0%	19.6%	23.6%

*1: best, 4: worst

Results of NMA on loss of lung function

Table 57 presents the scenarios for loss of lung function. Please note that only scenario 1 is presented in the text below. All other analyses are presented in Appendix B.

Table 57: Scenarios considered for loss of lung function

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding King	Excluded study because it introduced heterogeneity in the results (see scenario 1 results)
3	Excluding Richeldi 2011 and King	Excluded studies due to differences in patient characteristics
4	Including death (without Noble)	Included death as a composite outcome for loss of lung function
5	Including death (with Noble)	Similar to scenario 4, included Noble 2011 assuming that it reports the same outcome

Scenario 1: all evidence

Heterogeneity was observed in the pairwise meta-analysis of the pirfenidone studies (Table 59). In scenario 2, King et al. (King *et al.*, 2014[41]) was excluded from the all evidence analysis.

In the NMA the total residual deviance approximates the available number of data points (10). Based on the DIC, the fixed effects performed better than the random effects model. The credible interval was narrower in the fixed effects model than the random effects model.

The ORs for nintedanib and pirfenidone were statistically significant in the fixed effects model. N-acetylcysteine has similar efficacy to placebo. The results indicated that both nintedanib and pirfenidone have better efficacy than placebo and N-acetylcysteine.

Table 58: Nintedanib vs. placebo; scenario 1 loss of lung function

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	0.56 (0.28, 1.10)	13.7	13.7
Richeldi 2014	0.53 (0.41, 0.70)	86.3	86.3
Pooled OR (95% CI)		0.54 (0.42, 0.69)	0.54 (0.42, 0.69)
p-value (test OR=1)		p<0.001	p<0.001
Heterogeneity	I^2=0%, chi-square p-value=0.910		

Table 59: Pirfenidone vs. placebo; scenario 1 loss of lung function

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Noble 2011	0.69 (0.47, 1.00)	47.3	51.3
King 2014	0.43 (0.28, 0.64)	52.7	48.7
Pooled OR (95% CI)		0.55 (0.42, 0.72)	0.55 (0.34, 0.87)
p-value (test OR=1)		p<0.001	p=0.011
Heterogeneity	I^2=65.6%, chi-square p-value=0.088		

Table 60: N-acetylcysteine vs. placebo; scenario 1 loss of lung function

Study	OR (95% CI)	p-value
Martinez 2014	1.02 (0.59, 1.75)	0.949

Table 61: NMA; scenario 1 loss of lung function

	FIXED-EFFECTS (DIC=74.7, tot res dev 10.3)	RANDOM-EFFECTS (DIC=75.2, tot res dev 9.3)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	0.54 (0.42, 0.69)	0.54 (0.11, 2.70)
pirfenidone vs. placebo	0.55 (0.41, 0.72)	0.54 (0.11, 2.69)
N-acetylcysteine vs. placebo	1.02 (0.59, 1.76)	1.02 (0.10, 9.73)
All vs. nintedanib		
placebo vs. nintedanib	1.87 (1.45, 2.41)	1.86 (0.37, 9.39)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	1.02 (0.7, 1.49)	1.01 (0.10, 9.92)
N-acetylcysteine vs. nintedanib	1.9 (1.04, 3.47)	1.89 (0.11, 31.32)
nintedanib vs. all		
nintedanib vs. placebo	0.54 (0.42, 0.69)	0.54 (0.11, 2.70)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	0.98 (0.67, 1.42)	0.99 (0.10, 9.98)
nintedanib vs. N-acetylcysteine	0.53 (0.29, 0.96)	0.53 (0.03, 8.90)

Figure 24: Forest plot; scenario 1 loss of lung function fixed effects

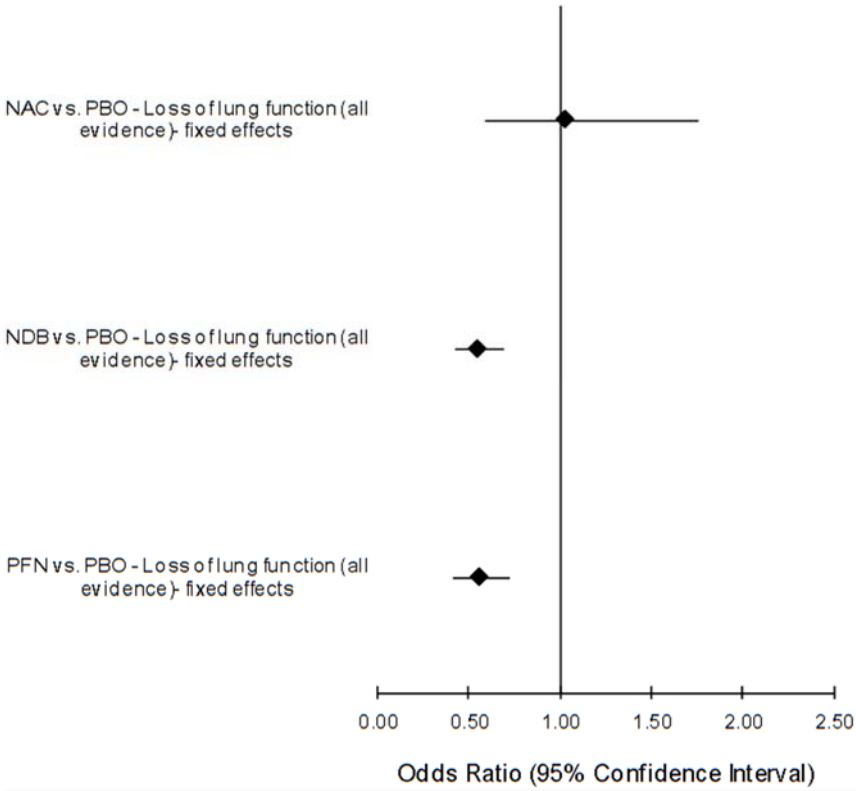


Figure 25: Forest plot; scenario 1 loss of lung function random effects

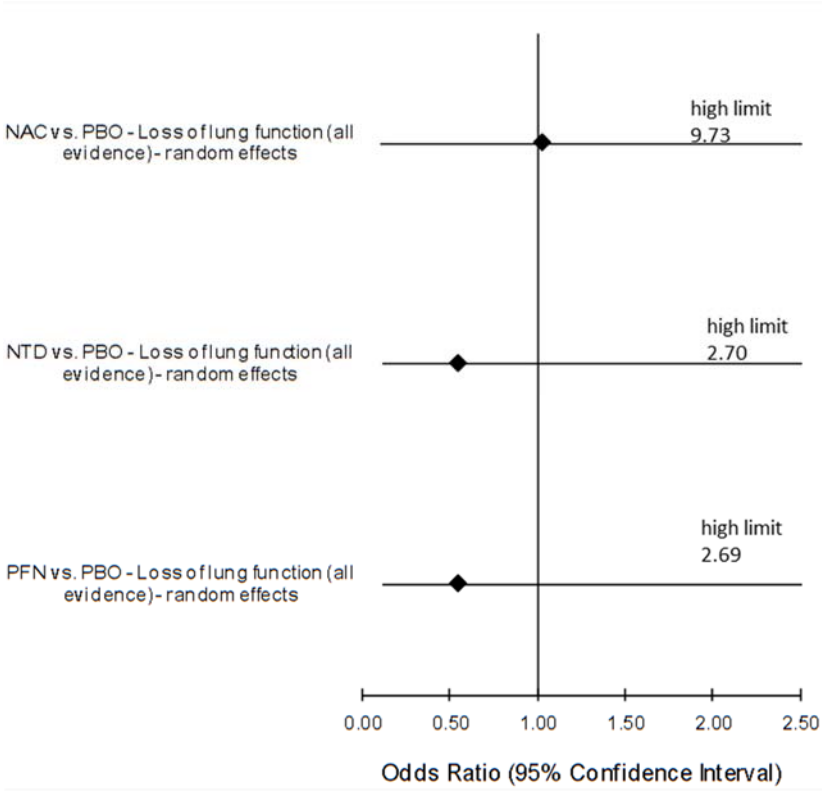


Table 62: Ranking matrix; scenario 1 loss of lung function

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	0.0%	0.0%	52.9%	47.1%	1.2%	8.3%	50.1%	40.4%
nintedanib	54.4%	44.3%	1.3%	0.0%	45.5%	38.6%	10.4%	5.5%
pirfenidone	44.7%	53.4%	1.9%	0.0%	43.2%	41.0%	10.4%	5.4%
N-acetylcysteine	1.0%	2.3%	43.9%	52.9%	10.2%	12.0%	29.1%	48.7%

*1: best, 4: worst

Results of pairwise comparison on PFS

The estimated HR of nintedanib vs. pirfenidone is 1.00 (0.71, 1.39). The estimated p-value is 0.982. These results indicate no difference in PFS between nintedanib and pirfenidone.

Results of NMA on 6MWT

In this analysis one study per active intervention is included in the NMA. Table 63, Table 64 and Table 65 present the results of the pairwise comparisons of the 6MWT, reported as changes from baseline. Table 66 presents the NMA results. Table 67 shows the NMA ranking matrix for the fixed effects models.

Figure 26 presents the forest plot for the fixed effects model.

In the NMA, the total residual deviance of the fixed effects model did not approximate the available number of data points (n=6), indicating a poor fit.

The mean change from baseline value was statistically significant for the PFN result, all other results were not. However, the credible intervals were very large for all comparisons. Regarding the point estimates, the NDB result reported the smallest change from baseline, while PFN and NAC reported higher and similar values.

Table 63: NDB vs. PBO; 6MWT change from baseline

Study	WMD (95% CI)	p-value
Richeldi 2011	6.3 (-27.0, 39.6)	0.710

Table 64: PFN vs. PBO; 6MWT change from baseline

Study	WMD (95% CI)	p-value
Noble 2011	24.0 (4.3, 43.7)	0.017

Table 65: NAC vs. PBO; 6MWT change from baseline

Study	WMD (95% CI)	p-value
Martinez 2014	24.2 (-2.6, 50.9)	0.076

Table 66: NMA; 6MWT change from baseline

	FIXED-EFFECTS (DIC=26.9, tot res dev 2.3)
Comparison	**Median WMD (95% CrI)**
All vs. PBO	
PBO vs. PBO	0.00 (-)
NDB vs. PBO	6.2 (-26.5, 38.8)
PFN vs. PBO	23.7 (4.1, 43.4)
NAC vs. PBO	23.8 (-2.7, 50.4)

Figure 26: Forest plot; change from baseline 6MWT fixed effects

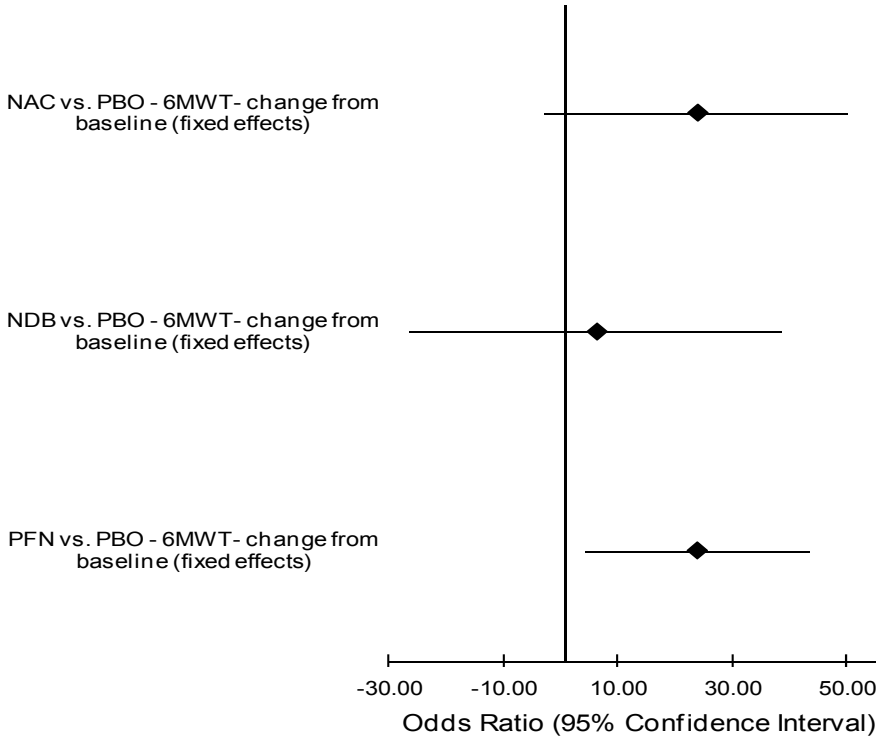


Table 67: Ranking matrix; 6MWT change from baseline (fixed effects)

	Fixed effects			
Ranking *	**1**	**2**	**3**	**4**
PBO	0.0%	1.7%	37.0%	61.3%
NDB	10.4%	17.8%	36.7%	35.1%
PFN	44.0%	43.3%	12.1%	0.6%
NAC	45.6%	37.3%	14.2%	2.9%

*1: best, 4: worst

Results of NMA on serious cardiac events

Table 68 presents the scenarios considered for serious cardiac events. Only the all evidence scenario is presented below. Scenario 2 can be found in Appendix B.

Table 68: Scenarios considered for serious cardiac events

Scenario	**Description**	**Reasons of choice**
1	All evidence	N/A
2	Excluding Richeldi 2011	Excluded Richeldi 2011 (phase II) to consider only phase III data for NDB + and to remove heterogeneity of the results (observed in scenario 1)

Scenario 1: all evidence

Table 69, Table 70, and Table 71 present the results of the pairwise comparisons. Table 72 presents the NMA results. Table 73 shows the NMA ranking matrix for the fixed effects and random effects models. Figure 27 and Figure 28 present the results of the NMA in a forest plot for the fixed effects and the random effects model respectively.

Heterogeneity was observed in the pairwise meta-analysis of the nintedanib studies (Table 53). Richeldi et al. (Richeldi *et al.*, 2011[1]) was excluded from the all-evidence analysis in a separate scenario (see Appendix B).

In the NMA, the total residual deviance of the fixed effects model did not approximate the available number of data points (8), indicating a poor fit. Based on the DIC, the random effects performed better than the fixed effects model. Nevertheless, in the random effects model the credible interval for all comparators indicated a high level of uncertainty around the point estimates and is therefore difficult to interpret the clinical relevance of this analysis.

None of the results were statistically significant, except for N-acetylcysteine vs. placebo in the fixed effects model. In terms of point estimates, the results indicated that nintedanib had a better safety profile than placebo, while N-acetylcysteine was estimated to be worse than placebo; however, the credible interval was very wide.

Table 69: Nintedanib vs. placebo; scenario 1 serious cardiac events

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	0.13 (0.02, 1.10)	20.8	35.8
Richeldi 2014	0.92 (0.53, 1.59)	79.2	64.2
Pooled OR (95% CI)		0.76 (0.45, 1.26)	0.46 (0.07, 2.89)
p-value (test OR=1)		0.282	0.407
Heterogeneity	I2=67.5%, chi-square p-value=0.079		

Table 70: Pirfenidone vs. placebo; scenario 1 serious cardiac events

Study	OR (95% CI)	p-value
Noble 2011	1.26 (0.65, 2.43)	0.494

Table 71: N-acetylcysteine vs. placebo; scenario 1 serious cardiac events

Study	OR (95% CI)	p-value
Martinez 2014	4.68 (0.99, 22.10)	0.051

Table 72: NMA; scenario 1 serious cardiac events

	FIXED-EFFECTS (DIC=50.3, tot res dev 10.7)	**RANDOM-EFFECTS (DIC=48.1, tot res dev 7.6)**
Comparison	Median OR (95% CrI)	Median OR (95% CrI)
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	0.76 (0.45, 1.27)	0.42 (0, 21.16)
pirfenidone vs. placebo	1.26 (0.65, 2.49)	1.26 (0, 459.98)
N-acetylcysteine vs. placebo	5.40 (1.27, 41.00)	5.64 (0.01, 2610.28)
All vs. nintedanib		
placebo vs. nintedanib	1.32 (0.79, 2.2)	2.41 (0.05, 234.9)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	1.67 (0.72, 3.9)	2.98 (0, 5868.54)
N-acetylcysteine vs. nintedanib	7.17 (1.53, 57.24)	13.96 (0.01, 35549.24)
Nintedanib vs. all		
nintedanib vs. placebo	0.76 (0.45, 1.27)	0.42 (0, 21.16)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	0.60 (0.26, 1.39)	0.34 (0, 340.6)
nintedanib vs. N-acetylcysteine	0.14 (0.02, 0.65)	0.07 (0, 75.98)

Figure 27: Forest plot; scenario 1 serious cardiac events fixed effects

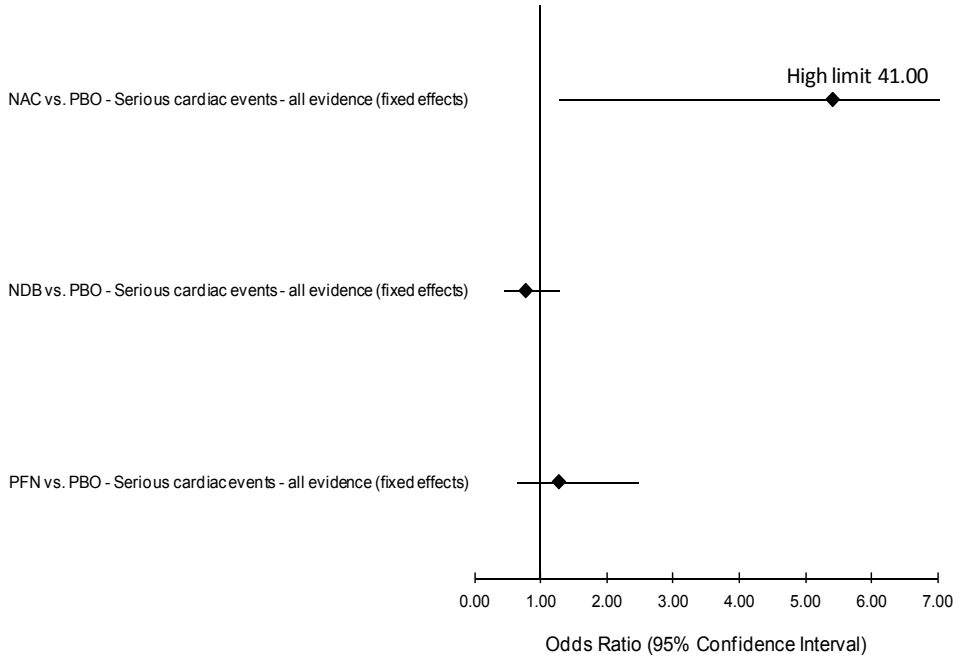


Figure 28: Forest plot; scenario 1 serious cardiac events random effects

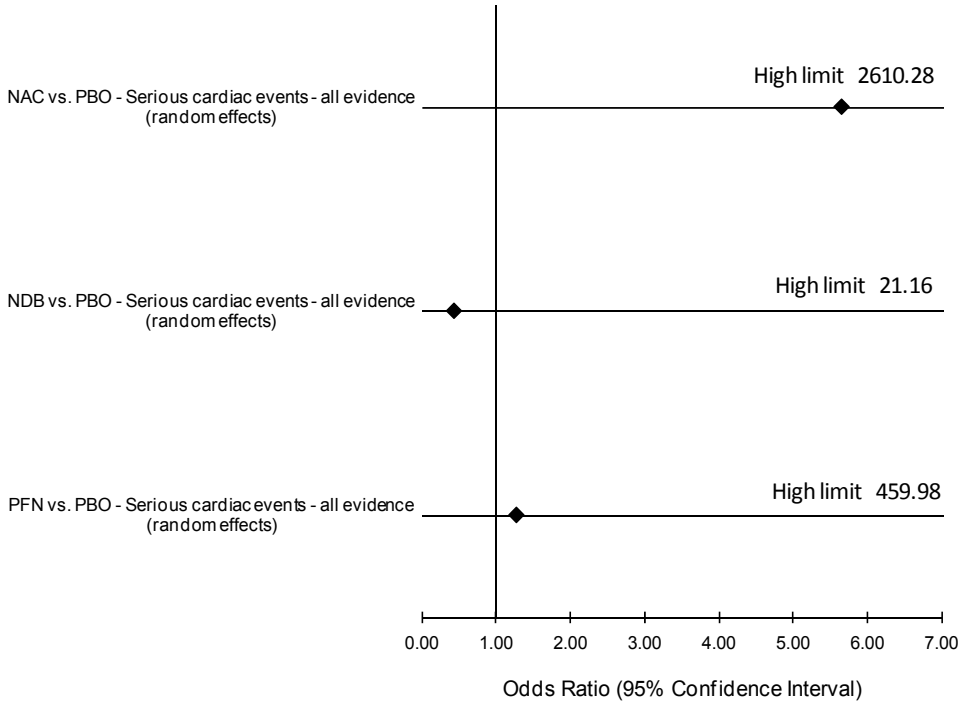


Table 73: Ranking matrix; scenario 1 serious cardiac events

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	10.7%	67.5%	21.6%	0.2%	10.5%	45.6%	36.8%	7.2%
nintedanib	79.5%	14.6%	5.8%	0.1%	56.2%	23.3%	13.9%	6.6%
pirfenidone	9.5%	17.2%	69.6%	3.6%	23.8%	20.8%	32.1%	23.3%
N-acetylcysteine	0.4%	0.6%	3.0%	96.0%	9.6%	10.3%	17.2%	62.9%

*1: best, 4: worst

Results of NMA on serious GI events

Table 74 presents the scenarios considered for serious GI events. Note that only the all evidence scenario is presented below. Scenario 2 can be found in Appendix B.

Table 74: Scenarios considered for serious GI events

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding Richeldi 2011	Excluded Richeldi 2011 (phase II) to consider only phase III data for NDB

Scenario 1: all evidence

No significant heterogeneity was observed in the pairwise meta-analyses (Table 75). Richeldi 2011 was excluded from the analysis in a separate scenario (see Appendix A).

In the NMA the total residual deviance approximates the available number of data points (8), indicating a good fit for both models. Based on the DIC, the fixed effects performed marginally better than the random effects model. The credible interval was narrower in the fixed effects model than the random effects model.

The OR for NBD vs. placebo and N-acetylcysteine vs. placebo were statistically significant in the fixed effects model. The results indicated that nintedanib had a worse profile than placebo, while N-acetylcysteine was estimated to be better than placebo. Pirfenidone was similar to placebo in this outcome.

Table 75: Nintedanib vs. placebo; scenario 1 serious GI events

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	9.44 (0.50, 178.14)	5.5	13.1
Richeldi 2014	1.82 (0.76, 4.38)	94.5	86.9
Pooled OR (95% CI)		2.24 (0.98, 5.11)	2.26 (0.75, 6.80)
p-value (test OR=1)		0.055	0.146
Heterogeneity	I^2=11.8%, chi-square p-value=0.287		

Table 76: Pirfenidone vs. placebo; scenario 1 serious GI events

Study	OR (95% CI)	p-value
Noble 2011	0.61 (0.25, 1.49)	0.278

Table 77. N-acetylcysteine vs. placebo; scenario 1 serious GI events

Study	OR (95% CI)	p-value
Martinez 2014	0.07 (0.004, 1.30)	0.075

Table 78: NMA; scenario 1 serious GI events

	FIXED-EFFECTS (DIC=42.4, tot res dev 8.2)	RANDOM-EFFECTS (DIC=42.5, tot res dev 7.7)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	2.35 (1.05, 5.88)	3.52 (0.08, 429.92)
pirfenidone vs. placebo	0.60 (0.23, 1.45)	0.59 (0, 178.99)
N-acetylcysteine vs. placebo	0.03 (0.00, 0.46)	0.03 (0, 14.94)

	FIXED-EFFECTS (DIC=42.4, tot res dev 8.2)	RANDOM-EFFECTS (DIC=42.5, tot res dev 7.7)
All vs. nintedanib		
placebo vs. nintedanib	0.43 (0.17, 0.95)	0.28 (0, 12.09)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	0.25 (0.07, 0.85)	0.17 (0, 121.88)
N-acetylcysteine vs. nintedanib	0.01 (0.00, 0.23)	0.01 (0, 9.25)
Nintedanib vs. all		
nintedanib vs. placebo	2.35 (1.05, 5.88)	3.52 (0.08, 429.92)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	3.96 (1.18, 14.51)	5.94 (0.01, 12570)
nintedanib vs. N-acetylcysteine	70.77 (4.39, 32340)	134.2 (0.11, 1466000)

Figure 29: Forest plot; scenario 1 serious GI events fixed effects

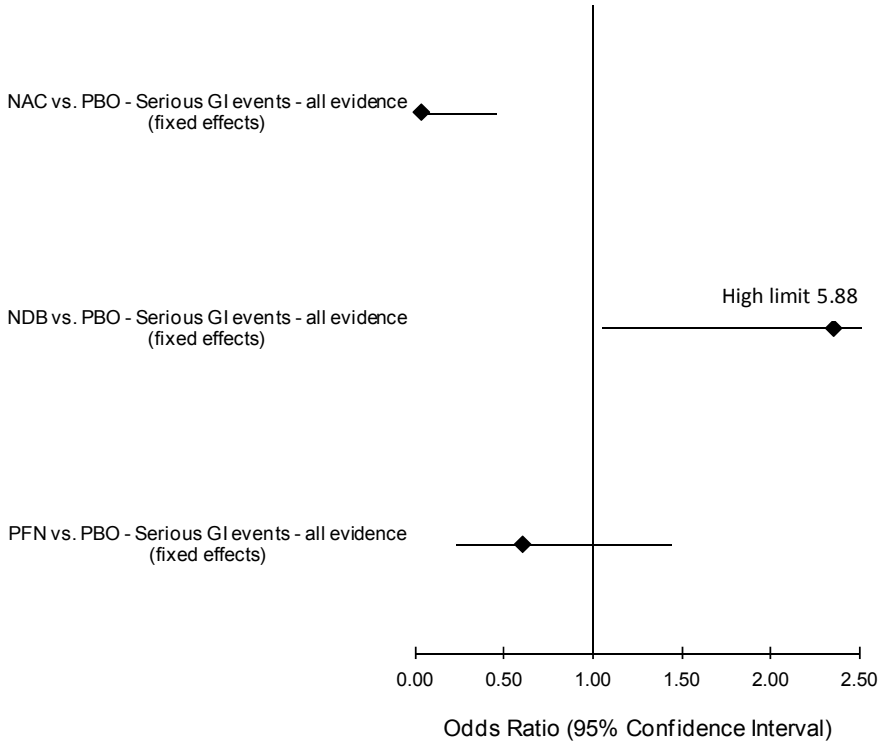


Figure 30: Forest plot; scenario 1 serious GI events random effects

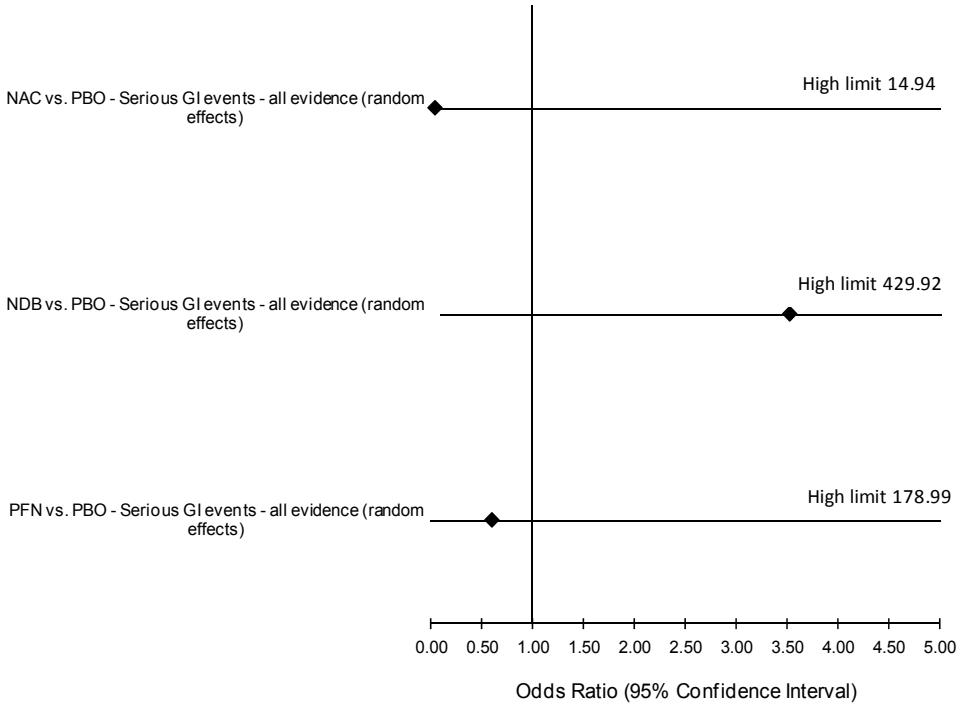


Table 79: Ranking matrix; scenario 1 serious GI events

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	0.0%	13.1%	85.3%	1.6%	3.2%	31.8%	56.5%	8.6%
nintedanib	0.0%	0.4%	2.3%	97.3%	2.5%	8.9%	17.5%	71.1%
pirfenidone	2.2%	84.6%	12.1%	1.1%	15.3%	47.4%	20.8%	16.5%
N-acetylcysteine	97.8%	1.9%	0.3%	0.0%	79.0%	11.9%	5.2%	3.8%

*1: best, 4: worst

Results of NMA on discontinuation due to AEs

Table 80 presents the scenarios considered for discontinuation due to adverse events. Only scenario 1 is considered below. All other scenarios are presented in Appendix B.

Table 80: Scenarios considered for discontinuation due to AEs

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding King, Azuma, and Taniguchi	Excluded studies due to differences in patient characteristics (See appendix B)
3	Excluding King, Azuma, Taniguchi, and Richeldi 2011	Similar to scenario 2, excluded Richeldi 2011 (phase II) as well to include only phase III data for NDB
4	Excluding Azuma and Taniguchi (Japanese studies)	Excluded Azuma 2005 and Taniguchi 2010 (Japanese populations)

Scenario1: all evidence

Table 81, Table 82, and Table 83 present the results of the pairwise comparisons. Table 84 presents the NMA results. Table 85 shows the NMA ranking matrix for the fixed effects and random effects models. Figure 31 and Figure 32 present the results of the NMA in a forest plot for the fixed effects and the random effects model respectively.

There was no heterogeneity observed in the pairwise meta-analyses of the nintedanib and pirfenidone studies.

In the NMA, the total residual deviance of the fixed effects model did not approximate the available number of data points (14), indicating a poor fit. Based on the DIC, the fixed effects performed better than the random effects model. In the random effects model the slightly larger credible interval for all comparators indicated a higher level of uncertainty around the point estimates.

The ORs for nintedanib and pirfenidone vs. placebo were statistically significant in the fixed effects model, while only the pirfenidone OR was statistically significant in the random effects model. In terms of point estimates, the results indicated that nintedanib and pirfenidone had a worse profile than placebo for this outcome, while N-acetylcysteine was estimated to be nearly equivalent to placebo. Comparing only the active treatments, nintedanib reported less discontinuation than pirfenidone, but more than N-acetylcysteine.

Table 81: Nintedanib vs. placebo; scenario 1 discontinuation due to AEs

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	1.26 (0.65, 2.47)	22.2	21.0
Richeldi 2014	1.60 (1.13, 2.26)	77.8	79.0
Pooled OR (95% CI)		1.52 (1.12, 2.07)	1.52 (1.12, 2.07)
p-value (test OR=1)		0.007	0.007
Heterogeneity	I^2=0%, chi-square p-value=0.539		

Table 82: Pirfenidone vs. placebo; scenario 1 discontinuation due to AEs

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Noble 2011	1.83 (1.14, 2.96)	42.8	44.5
King 2014	1.38 (0.83, 2.30)	43.2	39.8
Azuma 2005	3.02 (0.63, 14.41)	3.8	4.2
Taniguchi 2010	2.28 (0.89, 5.84)	10.2	11.5
Pooled OR (95% CI)		1.73 (1.26, 2.38)	1.72 (1.25, 2.36)
p-value (test OR=1)		0.001	0.001
Heterogeneity	I^2=0%, chi-square p-value=0.655		

Table 83: N-acetylcysteine vs. placebo; scenario 1 discontinuation due to AEs

Study	OR (95% CI)	p-value
Martinez 2014	1.02 (0.59, 1.75)	0.949

Table 84: NMA; scenario 1 discontinuation due to AEs

	FIXED-EFFECTS (DIC=89.9, tot res dev 11.4)	RANDOM-EFFECTS (DIC=91.7, tot res dev 11.8)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	1.52 (1.12, 2.08)	1.50 (0.72, 2.92)
pirfenidone vs. placebo	1.73 (1.27, 2.39)	1.78 (1.09, 3.35)
N-acetylcysteine vs. placebo	1.02 (0.59, 1.76)	1.02 (0.37, 2.73)
All vs. nintedanib		
placebo vs. nintedanib	0.66 (0.48, 0.89)	0.67 (0.34, 1.38)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	1.14 (0.73, 1.77)	1.19 (0.54, 3.24)
N-acetylcysteine vs. nintedanib	0.67 (0.36, 1.25)	0.68 (0.20, 2.34)
nintedanib vs. all		
nintedanib vs. placebo	1.52 (1.12, 2.08)	1.5 (0.72, 2.92)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	0.88 (0.56, 1.37)	0.84 (0.31, 1.85)
nintedanib vs. N-acetylcysteine	1.50 (0.80, 2.80)	1.47 (0.43, 4.89)

Figure 31: Forest plot; scenario 1 discontinuation due to AEs fixed effects

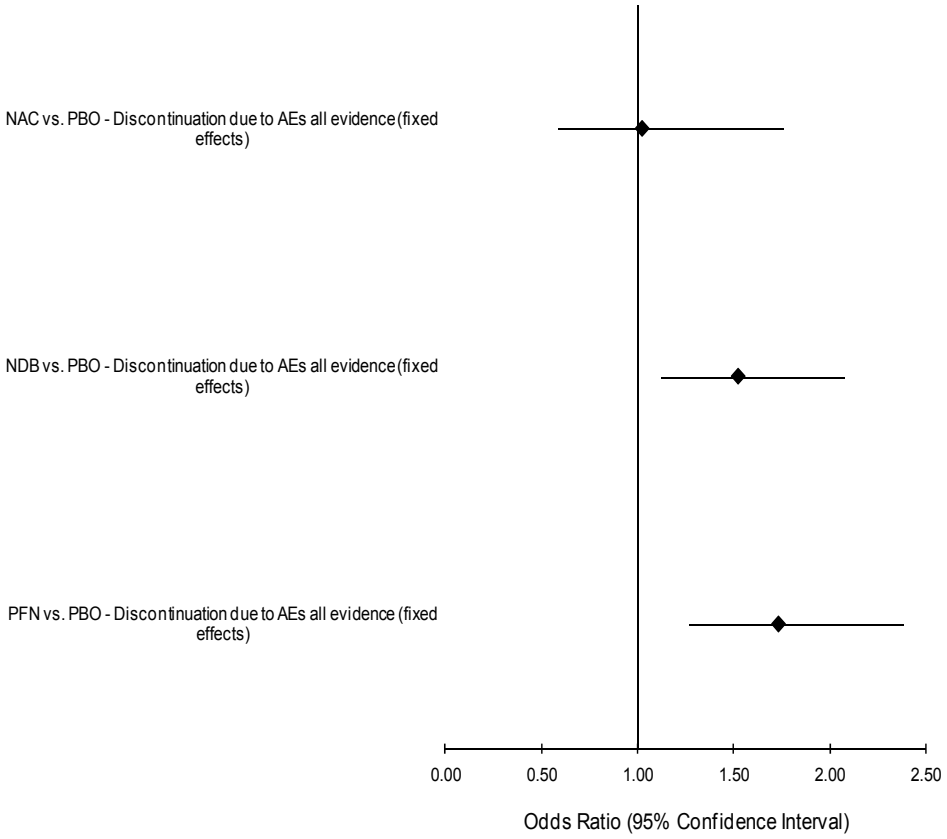


Figure 32: Forest plot; scenario 1 discontinuation due to AEs random effects

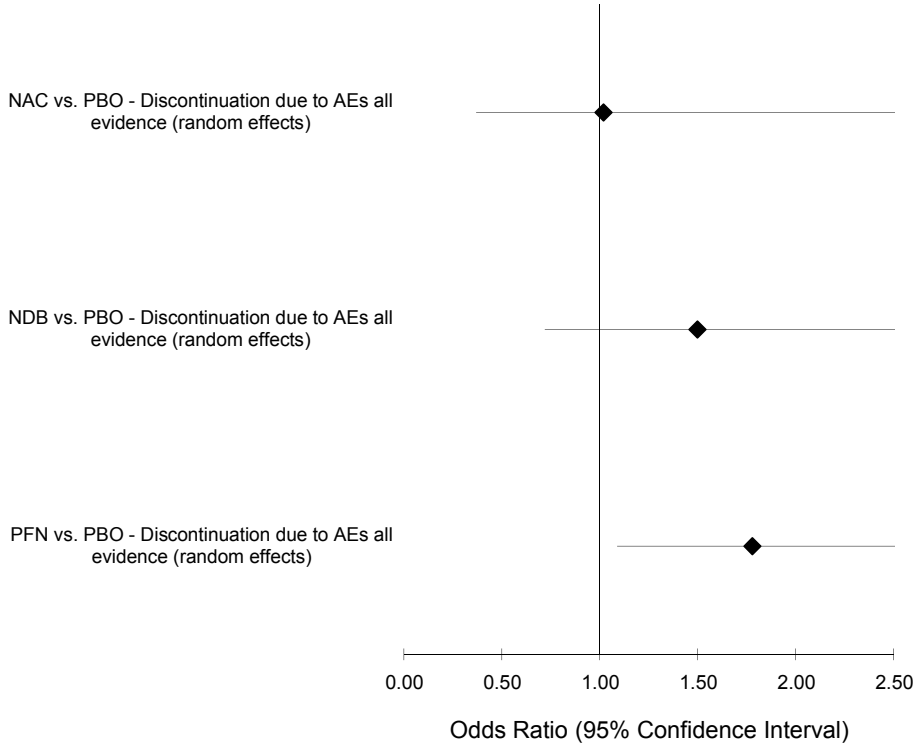


Table 85: Ranking matrix; scenario 1 discontinuation due to AEs

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	52.2%	47.6%	0.2%	0.0%	47.0%	48.2%	4.6%	0.2%
nintedanib	0.2%	9.3%	63.4%	27.2%	5.2%	15.5%	52.8%	26.5%
pirfenidone	0.0%	2.8%	27.7%	69.5%	0.8%	5.3%	28.6%	65.4%
N-acetylcysteine	47.6%	40.4%	8.8%	3.3%	47.1%	31.0%	14.0%	7.9%

*1: best, 4: worst

Results of NMA on overall discontinuation

Table 86 presents the scenarios considered for overall discontinuation. Note that only the all evidence scenario is presented below. All other scenarios are presented in Appendix B.

Table 86: Scenarios considered for overall discontinuation

Scenario	Description	Reasons of choice
1	All evidence	N/A
2	Excluding King and Taniguchi	Excluded studies due to differences in patient characteristics in King (see Appendix B) and because Taniguchi is a Japanese population study
3	Excluding King, Taniguchi, and Richeldi 2011	Similar to scenario 2, excluded Richeldi 2011 (phase II) as well to include only phase III data for NDB
4	Excluding King	Excluded study due to differences in patient characteristics in King (see Appendix B)
5	Excluding Taniguchi	Excluded Taniguchi (Japanese population)

Scenario 1: all evidence

Table 87, Table 88, and Table 89 present the results of the pairwise comparisons. Table 90 presents the NMA results. Table 91 shows the NMA ranking matrix for the fixed effects and random effects models. Figure 33 and Figure 34 present the results of the NMA in a forest plot for the fixed effects and the random effects model respectively.

There was no heterogeneity observed in the pairwise meta-analyses of the nintedanib and pirfenidone studies.

In the NMA, the total residual deviance of the fixed effects model did not approximate the available number of data points (12), indicating a poor fit. Based on the DIC, the fixed effects performed marginally better than the random effects model. In the random effects model the slightly larger credible interval for all comparators indicated a high level of uncertainty around the point estimates.

The OR for nintedanib vs. placebo was statistically significant in the fixed effects model, while the OR values for pirfenidone and N-acetylcysteine were not. Comparing

the point estimates of the active treatments, nintedanib reported the highest overall discontinuation, while N-acetylcysteine reported the least.

Table 87: Nintedanib vs. placebo; scenario 1 discontinuation overall

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Richeldi 2011	1.54 (0.81, 2.92)	28.3	29.4
Richeldi 2014	1.50 (0.99, 2.27)	71.8	70.6
Pooled OR (95% CI)		1.51 (1.06, 2.14)	1.51 (1.06, 2.14)
p-value (test OR=1)		0.021	0.021
Heterogeneity	I^2=0%, chi-square p-value=0.950		

Table 88: Pirfenidone vs. placebo; scenario 1 discontinuation overall

Study	OR (95% CI)	Weights FE (%)	Weights RE (%)
Noble 2011	1.04 (0.64, 1.70)	37.9	34.1
King 2014	1.51 (0.96, 2.36)	38.0	40.7
Taniguchi 2010	1.42 (0.80, 2.52)	24.0	25.2
Pooled OR (95% CI)		1.31 (0.98, 1.74)	1.31 (0.98, 1.74)
p-value (test OR=1)		0.066	0.067
Heterogeneity	I^2=0%, chi-square p-value=0.521		

Table 89: N-acetylcysteine vs. placebo; scenario 1 discontinuation overall

Study	OR (95% CI)	p-value
Martinez 2014	1.21 (0.69, 2.13)	0.514

Table 90: NMA; scenario 1 discontinuation overall

	FIXED-EFFECTS (DIC=81.9, tot res dev 9.6)	RANDOM-EFFECTS (DIC=83.6, tot res dev 10.1)
Comparison	**Median OR (95% CrI)**	**Median OR (95% CrI)**
All vs. placebo		
placebo vs. placebo	1.00 (-)	1.00 (-)
nintedanib vs. placebo	1.51 (1.07, 2.16)	1.52 (0.75, 3.09)
pirfenidone vs. placebo	1.30 (0.98, 1.73)	1.30 (0.73, 2.29)
N-acetylcysteine vs. placebo	1.21 (0.68, 2.15)	1.21 (0.44, 3.37)
All vs. nintedanib		
placebo vs. nintedanib	0.66 (0.46, 0.93)	0.66 (0.32, 1.34)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
pirfenidone vs. nintedanib	0.86 (0.54, 1.35)	0.85 (0.34, 2.12)
N-acetylcysteine vs. nintedanib	0.80 (0.41, 1.57)	0.80 (0.23, 2.78)
Nintedanib vs. all		
nintedanib vs. placebo	1.51 (1.07, 2.16)	1.52 (0.75, 3.09)
nintedanib vs. nintedanib	1.00 (-)	1.00 (-)
nintedanib vs. pirfenidone	1.16 (0.74, 1.84)	1.17 (0.47, 2.97)

	FIXED-EFFECTS (DIC=81.9, tot res dev 9.6)	RANDOM-EFFECTS (DIC=83.6, tot res dev 10.1)
nintedanib vs. N-acetylcysteine	1.25 (0.64, 2.45)	1.25 (0.36, 4.38)

Figure 33. Forest plot; scenario 1 overall discontinuation fixed effects

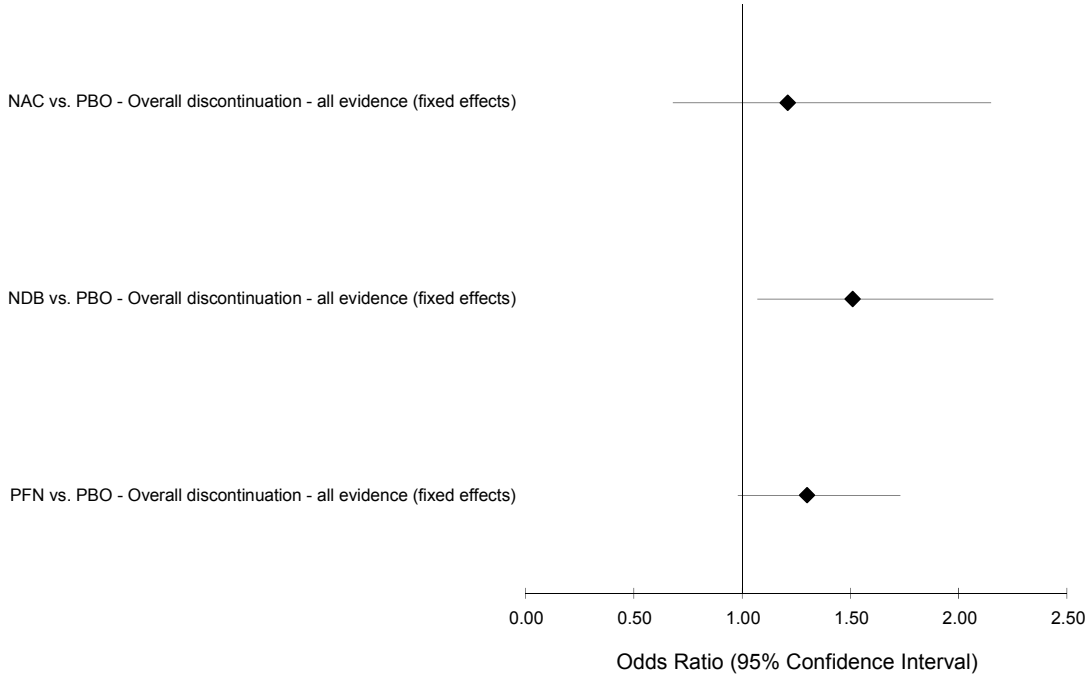


Figure 34: Forest plot; scenario 1 overall discontinuation random effects

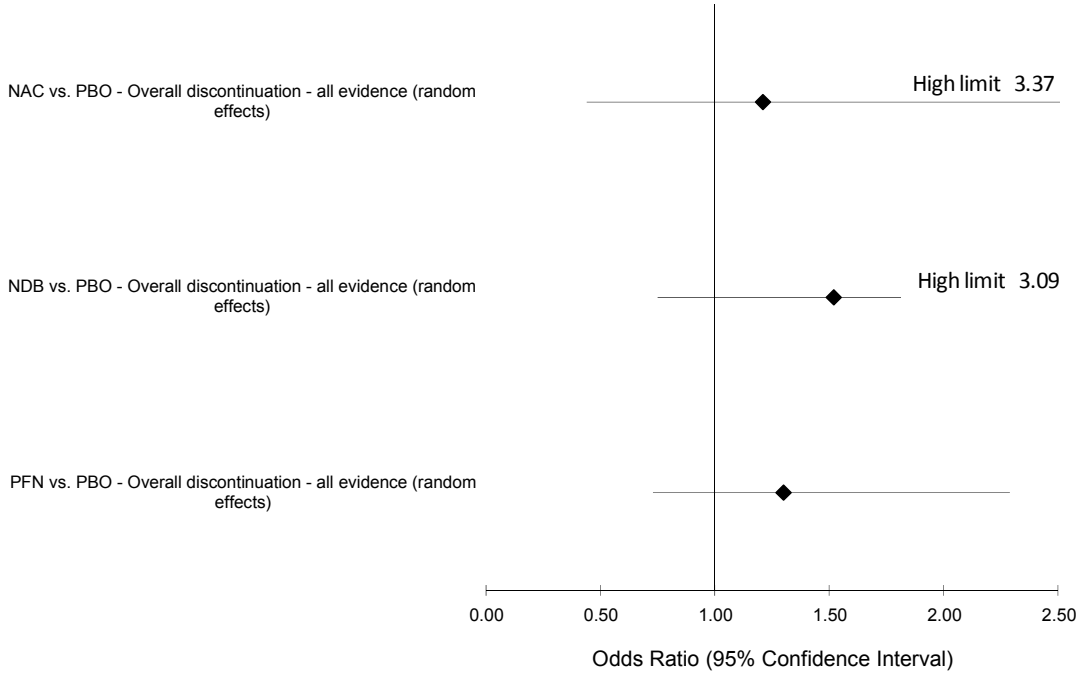


Table 91: Ranking matrix; scenario 1 discontinuation overall

	Fixed effects				Random effects			
Ranking*	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
placebo	71.2%	27.7%	1.1%	0.0%	58.0%	34.8%	6.6%	0.7%
nintedanib	0.8%	10.6%	28.9%	59.8%	4.8%	13.0%	29.4%	52.8%
pirfenidone	2.8%	32.1%	46.4%	18.7%	8.2%	29.2%	41.4%	21.2%
N-acetylcysteine	25.3%	29.7%	23.5%	21.5%	29.0%	23.0%	22.7%	25.3%

*1: best, 4: worst

Summary of the NMA Results

Indirect comparison is an important statistical methodology when head-to-head randomised controlled trial evidence is not available. There are limitations with respect to study similarity and potential causes of bias from the studies included. The impact of these studies was tested by including and excluding studies from the network. Furthermore, scenario analyses were performed when heterogeneity was identified by excluding certain studies.

With reference to overall survival, no intervention was statistically significantly different to placebo. However, nintedanib consistently performed better than placebo and N-acetylcysteine, and was comparable to pirfenidone. These results indicate that nintedanib could provide a much needed additional treatment option for IPF patients. The fixed effects model performed better than the random effects. N-acetylcysteine was estimated to have worse efficacy than placebo. All scenarios tested were consistent with the all evidence analysis.

No intervention was found to be statistically different to placebo for acute exacerbations, and there was large uncertainty around the point estimates. However, heterogeneity was identified in the pirfenidone pairwise meta-analysis. This was attributed to the study by Azuma et al (Azuma *et al.*, 2005[48]). When Azuma et al (Azuma *et al.*, 2005[48]) was included in the network all interventions had a similar efficacy comparing to placebo. The random effects model performed better than the fixed effect.

When Azuma was excluded from the network the results changed. When both clinical trials for nintedanib were considered, the results indicated a statistically significant difference in acute exacerbations in favour of nintedanib. The other two interventions were not statistically significantly different than placebo. In terms of the point estimates, the results on the efficacy of pirfenidone were dependent on the available studies; when Taniguchi et al (Taniguchi, 2010[49]) was included pirfenidone showed worse efficacy than placebo, albeit not statistically significant. Regarding N-acetylcysteine, when Homma et al (Homma, 2012[54]) was included the results favoured the intervention over placebo.

Loss of lung function was considered using two definitions (one with and one without death). The results were similar in both cases and indicated that both nintedanib and pirfenidone were statistically significant different to placebo, both indicating a greater clinical benefit. N-acetylcysteine had similar efficacy to placebo. In the all evidence scenario without death we identified heterogeneity in the pirfenidone studies. When King et al (King *et al.*, 2014[41]) was excluded from the network, the outcomes for nintedanib did not change; however, the results of pirfenidone became borderline statistically significant.

In the pairwise comparison in PFS the results showed no difference between nintedanib and pirfenidone.

In the serious cardiac events analysis, nintedanib and pirfenidone were similar to placebo, whereas N-acetylcysteine was significantly worse than placebo. A high level of heterogeneity was observed in the pairwise comparisons of the nintedanib studies. The wide confidence intervals seen in the random effects model in the NMA makes interpretation of these results difficult.

In serious GI events nintedanib was statistically worse than placebo (more than double the odds). Pirfenidone and N-acetylcysteine reported better safety profiles than nintedanib without showing any significant difference to placebo. When Richeldi et al (Richeldi, 2011[29]) was excluded from the analysis the point estimate for nintedanib was reduced and the uncertainty around the point estimate increased; the results were no longer statistically significant.

In the discontinuation due to AEs the results were significant for nintedanib and pirfenidone in the all evidence scenario; indicating that nintedanib and pirfenidone had a worse profile than placebo for this outcome. N-acetylcysteine was estimated to be nearly equivalent to placebo. Regardless of the inclusion or exclusion of King et al (King *et al.*, 2014[41]) and the Japanese studies (Azuma et al. 2005 (Azuma *et al.*, 2005[48]), Taniguchi et al (Taniguchi, 2010[49])), nintedanib had always lower discontinuation than pirfenidone.

In the overall discontinuation outcome all interventions have higher discontinuation to placebo. Statistical significance was only observed for the nintedanib results in scenario 1. Excluding King et al (King *et al.*, 2014[41]) and Taniguchi et al (Taniguchi, 2010[49]) from the network (scenario 2) reduced the pirfenidone results to near-equivalence with placebo. The other scenario analyses had no considerable impact on the results.

The overall survival estimates did not vary when different scenarios were considered; no intervention was statistically significant different to placebo. In acute exacerbations, the results changed when we considered homogeneity; indicating statistical significance in favour of nintedanib. Nevertheless, the results of this should be treated

with caution, considering that the input for acute exacerbations was patchy and the definition of the events varied across studies. For the outcome of loss of lung function the results were more stable across the different scenarios; both nintedanib and pirfenidone were shown to have better efficacy than placebo. When King et al (King *et al.*, 2014[41]) was excluded from the network the results for pirfenidone were borderline significant.

With regards to the 6MWT outcome, the mean change from baseline value was statistically significant for the pirfenidone result, while the rest of the results were not. The nintedanib result reported the smallest change from baseline.

The serious cardiac event estimates did not vary between scenarios; only the N-acetylcysteine intervention was statistically significant different to placebo. In serious GI events and in the all evidence scenario, nintedanib was significantly worse than placebo; while N-acetylcysteine was significantly better. The difference in serious GI events between nintedanib and the other interventions is in line with some of the concerns of the clinical community for this outcome.

With regards to discontinuation, for any reason or reasons related to AEs, both nintedanib and pirfenidone were significantly worse than placebo in most cases, although nintedanib had a better profile than pirfenidone. N-acetylcysteine was similar to placebo. When discontinuation was related to AEs, nintedanib had a better profile than pirfenidone. When we considered all evidence and discontinuation due to any cause, pirfenidone had better outcomes than nintedanib. The total residual deviance did not approximate the number of endpoints in most of the comparisons for the 6MWT and discontinuation; which indicated a poor fit of the model. This was probably due to the fact that the model had very few data points, and different from each other. Unfortunately, given the available evidence further analysis of this, by controlling for several covariates, was not possible. The results of the indirect analysis were considered reliable given that the individual pairwise comparisons also found similar OR values for all interventions.

Another limitation of the comparisons presented here is the absence of direct evidence with which to assess consistency (i.e. how well direct and indirect estimates of effect agree). An assessment of consistency between direct and indirect estimates of treatment effect can help to confirm or refute that the aforementioned assumptions have been satisfied. In the absence of direct evidence to enable this, we must simply accept that inconsistency may be present, but it is not yet quantifiable.

4.11 *Non-randomised and non-controlled evidence*

No non-randomised or non-controlled studies were identified in the systematic literature review.

4.12 *Adverse reactions*

Safety results were favourable and were consistent across trials. Total adverse events leading to discontinuation in both the TOMORROW trial and the replicate INPULSIS trials were similar between nintedanib and placebo. The most common adverse events were GI in nature and not severe. Though over 60% of patients on nintedanib reported diarrhoea within the INPULSIS trials, this was predominantly mild or moderate in severity and led to discontinuation in fewer than 5% of patients.

Table 92 displays the prevalence data for key adverse events in the TOMORROW study. Table 93 displays the prevalence data for key adverse events in the INPULSIS studies.

The mean duration of exposure to the study drug in the nintedanib and placebo groups was similar (approximately 45 weeks in each trial), but a higher proportion of patients in the nintedanib group than in the placebo group had dose reductions or interruptions.

The most frequent AE in the nintedanib groups in both trials was diarrhoea. Among the patients in the nintedanib groups who had diarrhoea, most reported events that were of mild or moderate intensity (93.7% in INPULSIS-1 and 95.2% in INPULSIS-2). Diarrhoea led to premature discontinuation of the study drug in fewer than 5% of patients in both of the INPULSIS studies. In both trials, the proportion of patients with serious adverse events was similar in the nintedanib and placebo groups. In both trials, a higher proportion of patients in the nintedanib groups than in the placebo groups had elevated levels of liver enzymes. Elevations of liver enzymes were reversible and not associated with clinically manifest liver disease.

Arterial thromboembolic events were infrequently reported (0.7% of patients in the placebo group, and 2.5% in the nintedanib treated group). While adverse events reflecting ischaemic heart disease were balanced between the nintedanib and placebo groups, a higher percentage of patients experienced myocardial infarctions in the nintedanib group (1.6%) compared to the placebo group (0.5%). Detailed analysis of the baseline characteristics of the patients who experienced MI revealed the presence of multiple co-morbidities known to be risk factors for MI including hypertension, diabetes mellitus, pre-existing heart ischaemia and dyslipidaemia. AEs related to disorders classified as “Other ischaemic heart disease” were balanced in favour of nintedanib. Overall, there was no imbalance in disfavour of nintedanib for deaths due to AEs reflecting cardiac disorders or ischaemic heart disease.

Table 92: Adverse Events of the TOMORROW Study

	Placebo	Nintedanib				Total
Event		50 mg Once daily	50 mg Twice daily	100 mg Twice daily	150 mg Twice daily	
	Number of patients (%)					
Any AE	77 (90.6)	78 (90.7)	78 (90.7)	82 (95.3)	80 (94.1)	395 (92.3)
Most frequent AEs*						
Diarrhoea	13 (15.3)	9 (10.5)	17 (19.8)	32 (37.2)	47 (55.3)	118 (27.6)
Cough	17 (20.0)	11 (12.8)	17 (19.8)	20 (23.3)	8 (9.4)	73 (17.1)
Nausea	8 (9.4)	9 (10.5)	8 (9.3)	17 (19.8)	20 (23.5)	62 (14.5)
Bronchitis	11 (12.9)	11 (12.8)	16 (18.6)	7 (8.1)	9 (10.6)	54 (12.6)
Dyspnoea	11 (12.9)	7 (8.1)	14(16.3)	13 (15.1)	6 (7.1)	51 (11.9)
Vomiting	4 (4.7)	1 (1.2)	6 (7.0)	11 (12.8)	11 (12.9)	33 (7.7)
Upper abdominal pain	3 (3.5)	6 (7.0)	10 (11.6)	2 (2.3)	10 (11.8)	31 (7.2)
SAEs	20 (23.5)	21 (24.4)	17 (19.8)	19 (22.1)	19 (22.4)	96 (22.4)
Fatal AEs	12 (14.1)	10 (11.6)	4 (4.7)	5 (5.8)	1 (1.2)	32 (7.5)
AEs leading to discontinuation	22 (25.9)	20 (23.3)	14 (16.3)	12 (14.0)	26 (30.6)	94 (22.0)
Respiratory, thoracic and mediastinal disorders	10 (11.8)	8 (9.3)	2 (2.3)	3 (3.5)	4 (4.7)	27 (6.3)
Gastrointestinal disorders	2 (2.4)	2 (2.3)	2 (2.3)	2 (2.3)	14 (16.5)	22 (5.1)
Infections and infestations	6 (7.1)	2 (2.3)	3 (3.5)	2 (2.3)	0	13 (3.0)
Cardiac disorders	6 (7.1)	2 (2.3)	2 (2.3)	1 (1.2)	0	11 (2.6)

AE: adverse event, SAEs: serious adverse events
*adverse events defined as most frequent were those having an incidence of more than 10% in any study group.

Table 93: Adverse Events of the INPULSIS Studies

	INPULSIS-1		INPULSIS-2	
Event	Nintedanib (N=309)	Placebo (N = 204)	Nintedanib (N=329)	Placebo (N = 219)
Any adverse event	298 (96.4)	181 (88.7)	311 (94.5)	198 (90.4)
Any adverse event, excluding progression of IPF*	296 (95.8)	179 (87.7)	311 (94.5)	197 (90.0)
Most frequent adverse events†				
Diarrhoea	190 (61.5)	38 (18.6)	208 (63.2)	40 (18.3)
Nausea	70 (22.7)	12 (5.9)	86 (26.1)	16 (7.3)
Nasopharyngitis	39 (12.6)	34 (16.7)	48 (14.6)	34 (15.5)
Cough	47 (15.2)	26 (12.7)	38 (11.6)	31 (14.2)
Progression of IPF	31 (10.0)	21 (10.3)	33 (10.0)	17 (7.8)
Bronchitis	36 (11.7)	18 (8.8)	30 (9.1)	24 (11.0)
Upper respiratory tract infection	28 (9.1)	18 (8.8)	30 (9.1)	24 (11.0)
Dyspnoea	22 (7.1)	23 (11.3)	27 (8.2)	25 (11.4)
Decreased appetite	26 (8.4)	14 (6.9)	42 (12.8)	10 (4.6)
Vomiting	40 (12.9)	4 (2.0)	34 (10.3)	7 (3.2)
Weight loss	25 (8.1)	13 (6.4)	37 (11.2)	2 (0.9)
Severe adverse events‡	81 (26.2)	37 (18.1)	93 (28.3)	62 (28.3)
Serious adverse events‡	96 (31.1)	55 (27.0)	98 (29.8)	72 (32.9)
Fatal adverse events	12 (3.9)	10 (4.9)	25 (7.6)	21 (9.6)
Adverse events leading to discontinuation§	65 (21.0)	22 (10.8)	58 (17.6)	33 (15.1)
Gastrointestinal disorders	26 (8.4)	3 (1.5)	21 (6.4)	2 (0.9)
Respiratory, thoracic, and mediastinal disorders	12 (3.9)	10 (4.9)	8 (2.4)	18 (8.2)
Investigation results	10 (3.2)	1 (0.5)	8 (2.4)	1 (0.5)
Cardiac disorders	5 (1.6)	4 (2.0)	2 (0.6)	3 (1.4)
General disorders and conditions involving the site of study-drug administration‖	8 (2.6)	3 (1.5)	2 (0.6)	1 (0.5)

*Progression of idiopathic pulmonary fibrosis was defined in accordance with the definition of idiopathic pulmonary fibrosis in the *Medical Dictionary for Regulatory Activities,* version 16.1, which includes disease worsening and exacerbations of idiopathic pulmonary fibrosis.

† The most frequent adverse events were defined as those with an incidence of more than 10% in any study group.

‡ A severe adverse event was related to intensity and was defined as an event that was incapacitating or that caused an inability to work or to perform usual activities. A serious adverse event was defined as any adverse event that resulted in death, was immediately life-threatening, resulted in persistent or clinically significant disability or incapacity, required or prolonged hospitalization, was related to a congenital anomaly or birth defect, or was deemed serious for any other reason.

§ Adverse events leading to study-drug discontinuation were reported when they occurred in 2% or more of patients in any study group and are listed according to system organ class. The analysis included adverse events with an onset after administration of the first dose of study medication and up to 28 days after administration of the last dose. Investigation results refer to the results of clinical laboratory tests, radiologic tests, physical examination, and physiologic tests.

‖ These events include disorders or conditions that involve several body systems or sites, including chest pain, fatigue, asthenia, and general deterioration of physical health.

4.13 *Interpretation of clinical effectiveness and safety evidence*

- Nintedanib is the first targeted treatment for IPF to consistently demonstrate efficacy in two identically designed Phase III trials and reach its primary endpoint in both studies. Indeed the phase III evidence for nintedanib confirmed the trend of efficacy observed in the phase II trial, which was conducted between 2007 and 2010, further supporting the validity of the results reported here. Pirfenidone, studied in the concurrent CAPACITY studies, failed to reach its primary endpoint in one of these studies.
- There is strong evidence from the two replicate phase III RCT's, INPULSIS-1 and INPULSIS-2, to support that nintedanib provides a clinically meaningful benefit for IPF patients, as confirmed by the annual rate of change in FVC compared to placebo. Considering that the average IPF patient has a lung function loss, measured by forced vital capacity (FVC), of 150–200mL per year (Ley et al. 2011). Nintedanib consistently reduced the decline in FVC by 50%, when compared to placebo over 52 weeks:
    - INPULSIS-1: -114.7mL (nintedanib) vs. -239.9mL (placebo)
    - INPULSIS-2: -113.6mL (nintedanib) vs. -207.3mL (placebo)

  This demonstrates its efficacy in limiting the progression of IPF. These results indicate that nintedanib provides a benefit to IPF patients who currently have limited treatment options. A reduction in the rate of decline in FVC is considered consistent with a slowing of disease progression (Wells, 2013[42]).

  The robustness of this finding was supported by the results of the sensitivity analyses for both trials, which were consistent with the results of the primary analysis. These included sensitivity analyses which assessed different ways of handling missing data.

- There were significant differences favouring nintedanib for the time to the first acute exacerbation only in INPULSIS-2 (hazard ratio, 0.38; 95% confidence interval, 0.19 - 0.77; P = .005). There was also less deterioration in quality of life as demonstrated by the change from baseline in the total SGRQ score in INPULSIS-2. There was no statistically significant difference between the nintedanib and placebo groups in the key secondary endpoints in INPULSIS-1. This discrepancy is likely related to the fact that exacerbations are rare in idiopathic pulmonary fibrosis and the trials were not powered to detect differences in this outcome (or the other secondary outcomes). Clinical interpretation may also be a reason for the difference between INPULSIS 1 and 2, as the adjudicated exacerbation analysis from the pooled results indicated a significant benefit for nintedanib vs. placebo (hazard ratio, 0.32; 95% confidence interval, 0.16 – 0.65; P = 0.001).
- The safety of nintedanib has been evaluated in clinical trials of over 1,000 IPF patients. The mean duration of exposure to trial medication was 10.28 months

for patients in the nintedanib group and 10.88 months for patients in the placebo group (INPULSIS-1) and 10.29 months for patients in the nintedanib group and 10.78 months for patients in the placebo group (INPULSIS-2), confirming that nintedanib has acceptable tolerability.

- The most common adverse events reported in both INPULSIS trials were gastrointestinal in nature, of mild to moderate severity and rarely led to treatment discontinuation.
- The most frequent adverse event in the nintedanib groups was diarrhoea, which was reported in 62% vs. 19% and 63% vs. 18% in INPULSIS-1 and INPULSIS-2, respectively). Less than 5% of patients in the nintedanib groups of INPULSIS-1 and INPULSIS-2 discontinued treatment due to this event.
- The proportion of patients with serious adverse events was similar in all groups for both INPULSIS trials. However, a higher percentage of patients in the nintedanib groups had myocardial infarctions; 1.3% vs. 0.5% (INPULSIS-1) and 0.9% vs. 0.5% (INPULSIS-2). This may be a possible class effect, since evidence that a number of tyrosine kinase inhibitors are associated with cardiotoxicity (Orphanos *et al.*, 2009[66]) and warrants further investigation of this observed effect of nintedanib in larger cohorts. The difference may also be due to the use of MedDRA codes, as the total 'ischaemic myocardial events' were balanced, and AEs related to disorders classified as "Other ischaemic heart disease" were balanced in favour of nintedanib.
- Nintedanib consistently performed better than placebo and N-acetylcysteine and was comparable to pirfenidone. These results indicate that nintedanib could provide a much needed additional treatment option for IPF patients

Strengths of the evidence base

Marketing authorisation status

EMA Review

In recognition that IPF is a rare disease, affecting no more than 3 per 10, 000 people in the EU, nintedanib was granted an orphan designation (EU/3/13/1123) by the European Commission on 26th April 2013 for the treatment of IPF (EMA, 2013[47]). Marketing authorisation was granted by the European Commission on the 15th January 2015 having undergone accelerated assessment, further confirming the urgent need to make effective treatments available to this patient group.

FDA Approval

Following Breakthrough Therapy designation granted by the FDA, nintedanib was approved for the treatment of IPF in October 2014. Approval was based on the INPULSIS phase III data presented in this submission dossier, confirming the strong evidence of efficacy of nintedanib in this population.

Phase III Evidence

Study Design

In their guideline on "clinical trials in small populations", the EMA confirm that "high levels of evidence in drug development come from well-planned and well executed controlled clinical trials, particularly trials that have minimised bias through appropriate blinding and randomisation" (EMA, 2006[48]).

These recommendations are endorsed in the INPULSIS studies which are replicate, randomised, double-blind phase III trials in which the safety and efficacy of nintedanib was evaluated in IPF patients. These studies provided consistent results demonstrating the efficacy and safety of nintedanib, the consistency of the results obtained (from phase II and phase III) further confirming that the evidence base is valid.

The evidence from the INPULSIS studies is based on a total of 1,066 patients, the size exceeding the CAPACITY studies for pirfenidone which included ~ 700 patients in the pooled analyses and for which NICE granted approval in April 2013. The INPULSIS studies were performed at >200 sites in 24 countries in the Americas, Europe (including the UK), Asia and Australia, thus including IPF patients from a diverse pool, further enhancing the applicability of the evidence base to IPF patients in clinical practice across the world. Missing data is a recognised limitation which may impact the evidence from clinical trials leading to variation depending on the assumptions selected for the analyses. The proportion of patients with missing FVC data at week 52 was approximately 15%. The proportion of patients with missing data did not differ significantly between the nintedanib and placebo groups. Here the consistency between the multiple imputation analyses and the primary analysis of the INPULSIS studies confirms the robustness of the primary analyses.

The choice of FVC as a valid endpoint is supported by the study reported by du Bois *et al.* (du Bois et al., 2011[49]), which describes FVC to be a reliable, valid, and responsive measure of clinical status in patients with IPF, and has been the most commonly used endpoint in trials of IPF.

Baseline characteristics

In each of the INPULSIS trials, patients were eligible to enter the study, upon satisfying pre-specified inclusion criteria which included an IPF diagnosis, based upon the most recent American Thoracic Society (ATS) / European Respiratory Society (ERS) /Japanese Respiratory Society (JRS) / Latin American Thoracic Society (ALAT) guidelines. In abiding to a consistent definition for IPF diagnosis across all trial sites with that applied in practice, INPULSIS study participants are representative of IPF patients in clinical practice.

The baseline characteristics were well balanced between the study arms at baseline. This similarity among the study arms indicates that the randomisation procedure was

effective and that the observed differences in outcome between the nintedanib and placebo groups are an indication of the true treatment effect and benefit IPF patients will receive from nintedanib. Both INPULSIS studies favoured nintedanib for the primary outcome, the annual rate of change in FVC was -114.7 ml with nintedanib vs. -239.9 ml with placebo (difference 125.3 ml; 95% confidence interval [CI], 77.7 to 172.8; P<0.001) in INPULSIS-1 and -113.6 ml with nintedanib vs. -207.3 ml with placebo (difference, 93.7 ml; 95% CI, 44.8 to 142.7; P<0.001) in INPULSIS-2.

Analyses

In both trials the results of the pre-specified sensitivity analyses were consistent with the primary analyses. In particular, in the multiple imputation analyses, which were based on the conservative assumption that missing data were informative rather than random, the estimates of treatment effect, including confidence intervals, were consistent with the results of the primary analysis in each study. This consistency supports that the primary results are robust and not influenced by alternative assumptions about missing data.

The pre-specified pooled analysis of data from the two trials showed that a significantly greater proportion of patients in the nintedanib group than placebo group had an FVC response when both definitions of response were considered (a decline in the percentage of predicted FVC that was not more than 5 percentage points and a decline that was not more than 10 percentage points at week 52).

Extension Trial

The on-going extension trial following patients studied in the INPULSIS trials (NCT01619085) is evidence that given the choice, patients opt to continue receiving the benefits from nintedanib treatment. The purpose of the extension trial is to assess the long-term safety of nintedanib treatment in patients with IPF who have completed one year treatment and the follow up period in the double-blind phase III placebo controlled (INPULSIS) parent trials, who wish to continue treatment with nintedanib.

Limitations of the evidence base

Placebo controlled

There are no head-to-head trials investigating nintedanib against the comparators specified in the decision problem. The INPULSIS studies, initiated in April 2011, were placebo-controlled. It should be noted that patients were allowed to use certain background therapies in case of acute exacerbations or disease decline after 6 months on therapy, which is similar to current BSC.

NICE's guidance for "the diagnosis and management of IPF", published in June 2013, confirms the limited options available for IPF patients and recommends best supportive care, lung transplantation and the option of pirfenidone, following its positive approval in April 2013. Pirfenidone was not a viable choice for a comparator in the INPULSIS trials, which were completed in October 2013, while pirfenidone did not receive positive

approval from the FDA until October 2014, nor was it commercially available in most of Europe at the time the INPULSIS trials were initiated, making it an inappropriate choice across all study centres at the time of study initiation. In recognition of the limited treatment options available for IPF patients, international guidelines recommend that high-quality prospective, controlled, clinical trials of new therapies for IPF are required and make no reference to the preferred choice of comparator which reflects the content of the guideline (Raghu *et al.*, 2011[13]).

Missing data

The proportion of patients with missing FVC data at week 52 was approximately 15%, and did not differ significantly between the nintedanib and placebo groups. Though missing data is a recognised limitation which may impact the evidence from clinical trials, leading to variation depending on the assumptions founding the analyses, the consistency between the multiple imputation analyses and the primary analysis of the INPULSIS studies confirms the robustness of the primary analyses.

Long-term Efficacy and Safety

The INPULSIS studies confirm the efficacy and safety of nintedanib over a 52 week period and although the trend suggests that the improvement imparted to IPF patients will be sustained beyond one year, evidence from the on-going extension trial will better inform this question.

Weakness in the evidence base for the NMA

There were limitations with respect to study similarity and potential causes of bias from the studies included. In particular, a large body of evidence for pirfenidone which was considered in the NMA was derived from Japanese patient cohorts. The impact of these studies was tested by including and excluding studies from the network. Furthermore, scenario analyses were performed when heterogeneity was identified by excluding certain studies.

Relevance to decision problem

Unmet need

There are over 5,000 new cases of IPF diagnosed each year in the UK (Navaratnam *et al.*, 2011[20]) and the incidence of IPF more than doubled between 1990 and 2003 (Gribbin *et al.*, 2006[67]) yet treatment options for these patients remain limited. The progressive nature of IPF means patients deteriorate rapidly; median survival is approximately 3 years from time of diagnosis (NICE, 2013[18]). Hospital admissions have increased 5% year on year between 1980 and 2010 (Navaratnam *et al.*, 2011[20]).

In the UK, more people will die each year from IPF than from ovarian cancer, lymphoma, leukaemia, mesothelioma or kidney cancer (Navaratnam *et al.*, 2011[20]). There is no cure and until April 2013 there were no licensed therapies to treat IPF in the UK. In their guidance for the diagnosis and management of suspected IPF, NICE

recognise that there is no conclusive evidence to support the use of any drugs for extending the survival of people with IPF. NICE acknowledge the limitations of available treatments and in separate guidance recommend pirfenidone as an option for treating IPF when specific criteria are satisfied.

Nintedanib has demonstrated efficacy in the treatment of IPF patients, and thus addresses this important unmet need.

Clinical trial endpoints

Nintedanib has demonstrated that it provides clinically relevant benefits in patients for which available treatment options are limited. The primary endpoint in the INPULSIS studies was the annual rate of decline in forced vital capacity (FVC). Indeed FVC has been widely used as the primary endpoint in Phase III clinical trials for IPF (Richeldi *et al.*, 2011[1], Noble, 2011[40], King *et al.*, 2014[41], Taniguchi, 2010[49], Tomioka, 2005[53], Homma, 2012[54], Richeldi, 2014[68]). Decline of FVC over 1 year has been shown to be associated with overall prognosis in patients with IPF and is correlated with survival time as demonstrated in multiple large cohorts (du Bois *et al.*, 2011[7], Collard *et al.*, 2003[69], Latsi *et al.*, 2003[70], Flaherty *et al.*, 2003[71]). In the recent study by du Bois *et al.* (du Bois *et al.*, 2011[7]), the authors conclude that FVC is a reliable, valid, and responsive measure of clinical status in patients with IPF, and a decline of 2–6%, although small, represents a clinically important difference. In both INPULSIS studies nintedanib delayed disease progression by reducing annual decline in lung function by approximately 50% when compared to placebo (~4% and ~8% annual decline in nintedanib and placebo groups, respectively for both INPULSIS studies). In the NICE guideline it is recognised that over time the symptoms of IPF are associated with decline in lung function. By impeding this decline, nintedanib is addressing an important characteristic of this disease and offering IPF patients a treatment that targets the course of their disease.

Acute exacerbations of IPF, unpredictable and catastrophic events where many patients present with severe hypoxemia, and respiratory failure requiring mechanical ventilation, are events of clinical significance associated with high mortality (up to 50% 30 day mortality) (Collard *et al.*, 2007[61], Mallick, 2008[72]). In the INPULSIS studies a benefit for nintedanib was observed only in INPULSIS-2, there was a significant increase in the time to first acute exacerbation in the nintedanib group, compared to placebo and the proportion of patients with at least one investigator reported acute exacerbation was lower in the nintedanib group compared to the placebo group (3.6% vs. 9.6%). Although the INPULSIS studies were not powered to detect the effect of nintedanib on acute exacerbations the difference observed in INPULSIS-2 favouring nintedanib may be representative of the true effect of nintedanib treatment requiring confirmation in a larger cohort. The average duration on mechanical ventilation for acute respiratory failure events in IPF patients is 8.3 days and typically unsuccessful (Mallick, 2008[72]); any reduction in acute exacerbations nintedanib offers would

therefore be significant in terms of reducing costs (attached to mechanical ventilation), and in reducing the risk of this fatal event in IPF patients.

There was no significant clinical difference between nintedanib and placebo groups in the SGRQ scores in both INPULSIS-1 and INPULSIS-2 studies although the pooled analysis showed a trend of reduced deterioration for the nintedanib patients. The SGRQ, developed for patients with obstructive lung diseases, is increasingly being used to evaluate HRQL in patients with IPF. A recent literature review investigating the appropriateness of the SGRQ for measuring HRQL in patients with IPF confirmed this tool as a useful measure of HRQL in this patient population. However, several questions remain unaddressed, and further research is needed to confirm the SGRQ's utility in IPF.

Survival analysis

Survival is a major endpoint in IPF. Time to death, including vital status follow-up for those patients who prematurely discontinued trial treatment, time to on-treatment death, and time to death due to respiratory cause were evaluated as secondary endpoints in this trial and were additionally analysed on data pooled from both trials (INPULSIS-1 and INPULSIS-2) in order to improve the accuracy and statistical power. There was a numerical advantage in favour of nintedanib for all time-to-event survival endpoints (time to death, time to death due to respiratory cause, time to on-treatment death, time to death or lung transplant, time to death or lung transplant or qualifying for a lung transplant) with a lower proportion of patients experiencing an event in the nintedanib group than in the placebo group. This positive trend toward a reduction in all-cause and respiratory mortality in patients with IPF treated with nintedanib would be of great benefit in clinical practice.

The proportion of patients who died from any cause over the 52 week treatment period was numerically lower in the nintedanib group (5.5%) compared to the placebo group (7.8%). The hazard ratio for time to death over 52 weeks in the nintedanib group was 0.70 (95% CI: 0.43, 1.12; $p = 0.14$).

Safety outcomes

The safety of nintedanib has been investigated in ~1,500 IPF patients. The most frequent adverse event in the nintedanib groups was diarrhoea with less than 5% of patients in the nintedanib groups of INPULSIS-1 and INPULSIS-2 discontinuing treatment due to this event. Nintedanib therefore has a tolerability profile which is manageable making it more likely that IPF patients will adhere to their treatment and in turn maximise the benefit they receive from nintedanib treatment. In this regard nintedanib has a more favourable dosing schedule and pill burden compared to pirfenidone, which may also contribute to ensuring treatment adherence. Nintedanib treatment consisted of 1 capsule of 150mg twice daily whereas the recommended dosage of pirfenidone is 3×267 mg capsules 3 times a day (2,403 mg/day) (NICE, 2013[17]).

4.14 *Ongoing studies*

Details of the TOMORROW and INPULSIS studies can be found in Table 10 and ongoing/extension studies are presented in Table 11.

5 Cost effectiveness

SECTION SUMMARY

- Nintedanib at list price and at discounted price dominated pirfenidone (more QALYs and lower costs) at its list price, and this result was consistent in the majority of sensitivity analyses.
- Scenario analyses investigating theoretical discounts for pirfenidone, demonstrated nintedanib with PAS implemented dominated pirfenidone up to a xxx discount of pirfenidone; at a xxx pirfenidone discount, nintedanib's ICER vs pirfenidone was xxx.
- Nintedanib and pirfenidone were largely equivalent in estimated costs and QALYs, and the small incremental QALY gain leads the results to be sensitive to a change in incremental costs.
- In the pairwise comparison with BSC, nintedanib's (discounted price) ICER was xxx.
- In the full incremental cost-effectiveness comparison, PFN was again dominated by NDB.

5.1 *Identification of studies*

A systematic literature review was conducted to identify existing economic evaluations for nintedanib vs. its comparators in adults with IPF. The format of this systematic review has been reported to be in line with NICE's STA requirements as outlined in the NICE STA template. To date, NICE's requirements are the most stringent with respect to systematic review requirements.

The following computerised bibliographic databases advised by the CRD (Dissemination, 2009[24]) were searched on 26/08/2014:

- MEDLINE;
- MEDLINE In-Process;
- Embase;
- EconLit;
- NHS Economic Evaluation Database (NHS EED).

MEDLINE, Embase and EconLit were searched via the Ovid® search platform, and the NHSEED was searched via the CRD website.

Nintedanib and its alternative names were included in a search strategy that incorporated disease and patient population terms consistent with the clinical searches in section 4.1. In addition to those terms, database-specific NHS QIS (Macpherson *et al.*, 2009[73]) search filters were added. The search was not limited by country,

language, study design or date but studies published in languages other than English were not reviewed in full. The search strategies by databases are provided in Appendix B.

Conference proceedings were searched from 2012 to 2014 to identify publications in adults with IPF, treated with nintedanib or its comparators for the following conference sites:

- International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (www.ispor.org);
- International Health Economics Association (iHEA) (www.healtheconomics.org/).

The process of study selection in terms of abstract and full-text review was the same as is detailed in Section 4.1. However, different PICOS criteria were developed to capture economic evaluations (see Table 94).

Table 94: Eligibility criteria for economic evaluations

Criteria	Inclusion criteria	Exclusion criteria
Population	• Adult patients with IPF (including subgroups of patients e.g. subgroups based on 6MWT SpO2 and %VC)	• Paediatric and younger than 18; • Healthy participants; • Non-human.
Intervention/Comparators	Nintedanib in comparison with: • Pirfenidone • N-acetylcysteine (monotherapy) • Best supportive care (placebo)	• Any studies not containing nintedanib
Outcomes	• ICER (cost per QALY, cost per clinical event); • Total and incremental LYs, QALYs, clinical events; • Total and incremental costs of health states and adverse events; • Budget impact.	• No outcomes
Study design	• Economic evaluations: CUA, CBA, CEA, CCA, CMA.	• Reviews and systematic literature reviews* of economic evaluations; • Case reports; • Critical appraisals; • Update or commentaries on data published elsewhere; • Notes, letters or editorials.

Abbreviations: CUA, Cost-utility analysis; CBA, Cost-benefit analysis; CEA, Cost-effectiveness analysis; CCA, Cost-consequence analysis, CMA, Cost-minimisation analysis

*Systematic literature reviews will be excluded from the main database search but their reference lists will be searched for relevant studies not captured by the main database search and to be included as hand search in the review.

Quality assessment

For critically appraising economic studies, as recommended by NICE, the Drummond checklist (Drummond *et al.*, 2005[74]) was used. The Drummond criteria were created to support the generation of high-quality, rigorous economic evaluations. They involve using a total of 36 questions to assess three broad areas of the studies, namely: study

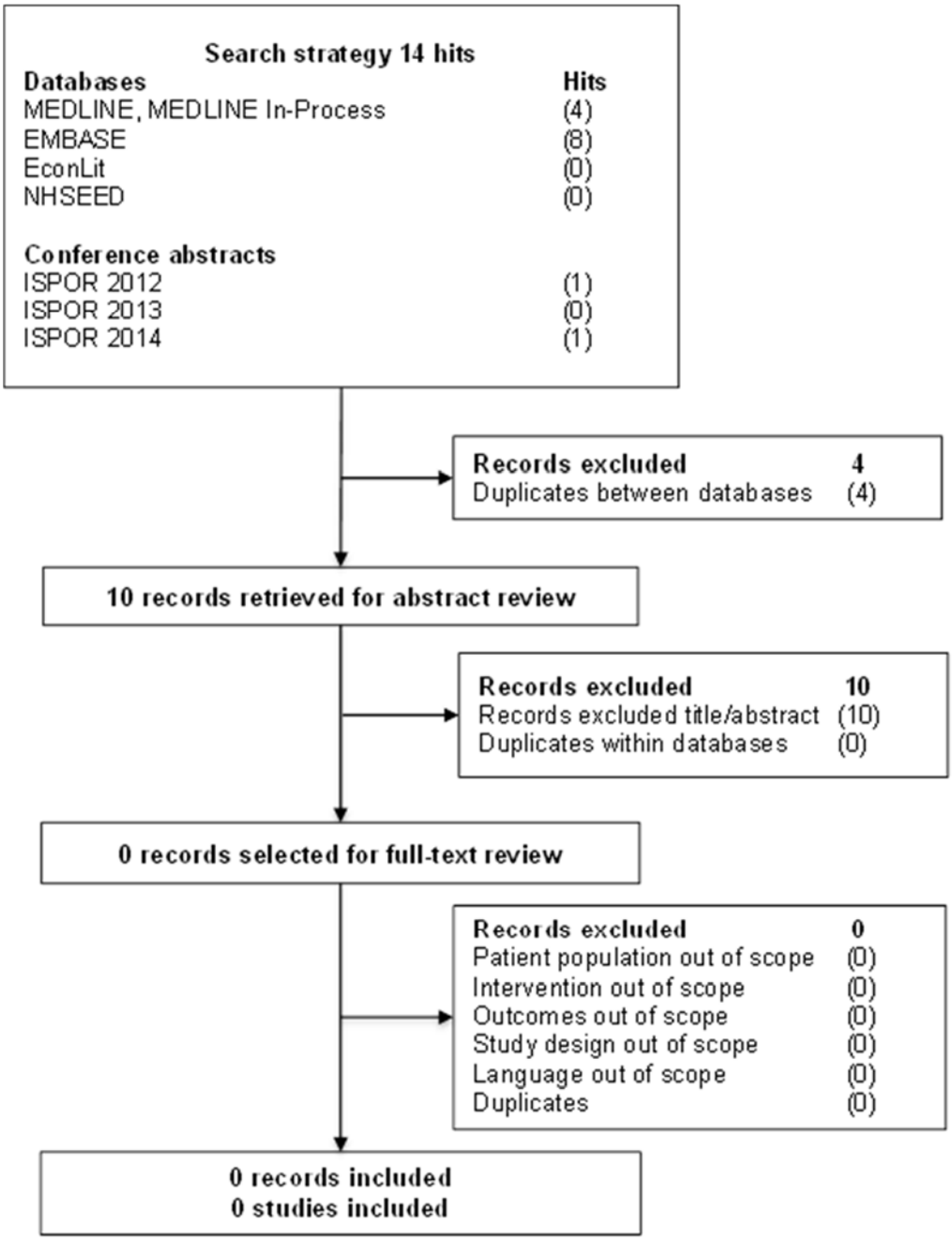
design; data collection; and analysis and interpretation of results. The result of the assessment process is a summary quality score on which models are judged to be either: high (++), moderate (+) or low quality (-). High-quality studies are considered to report clearly on almost all of the Drummond quality criteria questions, while studies of low quality do not report on most items. In this review, only studies in full-text form underwent a quality assessment because of the lack of details available for assessment in abstracts and posters.

Results

The process of data extraction for economic evaluation studies was the same as is detailed in section 4.1.

A total of 14 records were identified in the systematic literature review for economic evaluations of nintedanib in adults with idiopathic pulmonary fibrosis. Following duplicate removal, ten abstracts were retained for assessment against the predefined eligibility criteria (as described in Table 94). No relevant studies were found. Figure 35 presents the study selection process.

Figure 35: Study flow diagram for economic evaluation study selection (PRISMA diagram)



In addition to the SLR carried out for cost-effectiveness studies, a non-systematic search was performed to update the SLR to ensure all relevant cost-effectiveness studies had been considered (see Appendix A for methodology) This search resulted in the identification of one study (Loveman *et al.*, 2014[75]) which met the PICOS criteria described in Table 94.

Description of identified studies

Loveman *et al.* (Loveman *et al.*, 2014[75]), is a paper describing the methodology and results of a systematic review, network meta-analysis and health economic evaluation

investigating the effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis.

Methodology

Systematic reviews of clinical effectiveness, quality of life and cost effectiveness were undertaken. Eleven bibliographic databases were searched from inception to July 2013 and studies were assessed for eligibility against a set of pre-defined criteria. Two reviewers screened references, extracted data from included studies and appraised their quality. An advisory group was consulted about the choice of interventions. A narrative review was undertaken and where feasible fixed effect and random effects meta-analyses were undertaken including a network meta-analysis (NMA).

A decision-analytic Markov model was developed to estimate cost-effectiveness of pharmacological treatments for IPF. Following best practice recommendations, the model perspective was of the NHS and Personal Social Services (PSS), a discount rate of 3.5% for costs and health benefits was applied and outcomes were expressed as cost per QALY gained. Parameter values were obtained from the NMA and systematic reviews. Sensitivity analyses were undertaken.

Results

Fourteen studies were included in the review of clinical effectiveness, of which one evaluated azathioprine, three N-acetylcysteine [NAC] (alone or in combination), four pirfenidone, one nintedanib, one sildenafil, one thalidomide, two pulmonary rehabilitation, and one a disease management programme. Study quality was generally good. Evidence suggests that some effective treatments are available. In NMA only nintedanib and pirfenidone showed statistically significant improvements. The model results showed increased survival for five pharmacological treatments (NAC triple therapy, inhaled NAC, nintedanib, pirfenidone, and sildenafil) compared with best supportive care, at increased cost. The study reported only inhaled NAC was cost-effective at current willingness to pay thresholds but it may not be clinically effective.

Table 95: Summary of published cost-effectiveness studies identified

Study	Year	Summary of model	Patient population	QALYS	Costs(£)	ICER
Loveman et al 2014 (Loveman *et al.*, 2014[75])	2014	Decision analytic Markov model	Patients with initially unprogressed IPF	BSC: 2.98; Azathioprine & prednisolone: 2.66; NAC triple therapy: 3.03; Inhaled NAC: 3.37; Sildenafil: 3.11; Pirfenidone: 3.34; Nintedanib: 4.01	BSC: 3,084; Azathioprine & prednisolone: 4,313; NAC triple therapy: 5,021; Inhaled NAC: 5,029; Sildenafil: 12,008; Pirfenidone: 70,118; Nintedanib: 139,613	Azathioprine & prednisolone: Dominated; NAC triple therapy: Ext dominance; Inhaled NAC: 5,037; Sildenafil: Dominated; Pirfenidone: Dominated; Nintedanib: 209,246

NAC,n-acetylcysteine; BSC, Best supportive care; IPF, Idiopathic pulmonary fibrosis

Loveman et al (2014) (Loveman *et al.*, 2014[75]) included a literature review conducted in 2013, and as a result the NMA and cost effectiveness model did not consider the recently published INPULSIS, ASCEND and PANTHER-IPF NAC monotherapy (Richeldi, 2014[4], King *et al.*, 2014[41], Martinez *et al.*, 2014[52]) trials. For this reason, the information included within the paper relating to nintedanib is based on phase II data alone. The conclusions relating to NAC inhaled therapy must also be called into question as the PANTHER-IPF trial (Martinez *et al.*, 2014[52]) demonstrated NAC's lack of effectiveness.

In addition, Loveman et al (2014) (Loveman *et al.*, 2014[75]) used an assumed cost which does not match the list price for nintedanib, and as a result the ICER generated cannot be accurate.

A critical appraisal of Loveman et al (2014) (Loveman *et al.*, 2014[75]) is included in Appendix A. Details of the cost effectiveness model used in the study are not such that an informed critique can be made, however the model offers a simplistic two-state perspective on decline in lung function.

5.2 *De novo analysis*

5.2.1 Patient population

Marketing authorisation for nintedanib in IPF is for adults with IPF. The application was based on the findings of the phase III trials, thus the model populations was based on these trials and included patients within the marketing authorisation.

5.2.2 Model structure

The comparators considered in the analysis are:

BSC, assumed to be represented by the control arm of the phase III NDB clinical trial (Richeldi, 2014[4]):

- NDB 150mg twice daily continuous treatment (300mg per day)
- PFN 3 x 267mg three times daily (2,403mg per day)

Type of economic evaluation

The primary outcome measure of the economic evaluation is cost per QALYs. The incremental cost per QALY gained endpoint was selected as it enables an assessment of the value for money across a wide range of therapeutic areas. Cost per life-years (LYs) gained and exacerbation events avoided are also considered.

Perspective

The perspective taken when estimating costs in the economic evaluation is that of the NHS and PSS in England and Wales. All relevant healthcare costs are evaluated.

Time horizon

The economic evaluation estimates costs and health benefits over the cohort lifetime. A lifetime horizon is appropriate for this model as it captures all patients' relevant costs and HRQL impairment.

Economic model structure

Design and definition of health states

The economic model was designed with the objective to accommodate available evidence from the phase III clinical trial (Richeldi, 2014[4]), to depict an accurate representation of the patient condition, and to allow easy adaptation for sensitivity analysis of the main inputs.

In March 2015 a targeted review of the literature was conducted to identify other economic analyses in this disease area. The NICE clinical guideline for IPF (National Institute for Health and Care Excellence, January 2013[76]) were consulted, and the meetings of the NICE technology appraisal for pirfenidone (NICE TA 282 (NICE, 2013[17]), meeting held in Manchester on 29th January 2013) were attended. No other economic analyses were found at that time. Furthermore, clinicians were consulted during an advisory board meeting to present the assumptions and structure of the model, which were validated by the clinical experts (Dr. Toby Maher and Dr. Gisli Jenkins, meeting held in London on 23rd April, 2014 (Maher T, 2014[77])).

Further to this, to determine the model health states, different outcomes that could impact disease progression and deterioration of patient condition were considered. A targeted review of the literature identified studies that considered a single parameter (Alakhras et al. 2007 (Alakhras M, 2007[78]), Antoniou et al. 2008 (Antoniou KM, 2008[79]), Gribbin et al. 2006 (Gribbin *et al.*, 2006[67]), Hamada et al. 2007 (Hamada *et al.*, 2007[80]), Lederer et al. 2006a (Lederer *et al.*, 2006[81]), Lederer et al. 2006b (Lederer DJ, 2006[82]), Lopes et al. 2011 (Lopes AJ, 2011[83]), Manali et al. 2008 (Manali *et al.*, 2008[84]), Nathan et al. 2011 (Nathan *et al.*, 2011[85]), Zappala et al. 2010 (Zappala CJ, 2010[86])), or a risk scoring system (multiple parameters) (Ley et al. 2012 (Ley B, 2012[87]), Du Bois et al. 2011 (du Bois *et al.*, 2011[7])). Most studies analysed predictors of mortality rather than disease progression. Although this could be an indication of patient health status, it was possible that the definition of health states could come from a finer examination of those predictors. It followed that for those parameters that were considered relevant for the economic model, the literature for the minimal clinically important difference (MCID) was considered -albeit with reference to survival. A limitation identified in risk scoring systems was potential collinearity of the parameters considered; for example age and

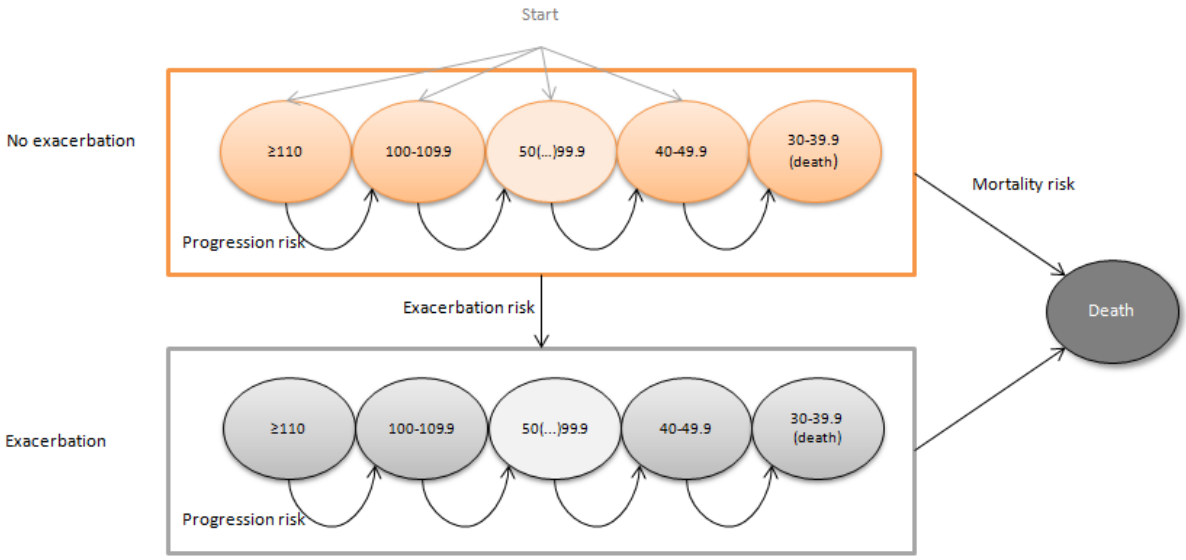
FVC per cent predicted (FVC%pred); a risk scoring system was not considered appropriate for the model.

Given that FVC "*is a widely used measure of* disease status and a common endpoint in clinical trials in IPF patients" (Du Bois et al. 2011b (Du Bois, 2011[88])), it was selected to be the main factor of disease progression. In particular, FVC%pred was preferred for a Markov model health state when compared with FVC (in mL), as it reflected the absolute state of patient condition adjusted for body capacity; age, gender, and height; this removed some of the heterogeneity of the health-state members, and it adhered to Markov model conventions. Furthermore, treatment effects for pirfenidone from the CAPACITY trials have been published in FVC%pred, allowing for an indirect treatment comparison with the data from the nintedanib INPULSIS trials.

According to many studies the MCID for FVC%pred ranged between a 2-6% change (Du Bois, 2011[88]), or a 10% change (Raghu *et al.*, 2011[13], Du Bois, 2011[88]). A 5-10% change has also been suggested to predict long-term outcomes including survival. After consultation with clinical experts (Dr. Toby Maher and Dr. Gisli Jenkins, meeting held in London on 23rd April, 2014 (Maher T, 2014[77])) and consideration of available evidence from the trial and in the literature, it was decided that a 10-point categorization of FVC%pred was appropriate for use in this analysis.

In addition to lung function, acute exacerbations of IPF are dramatic, singular events that are often fatal and a major cause of mortality and morbidity in IPF. Therefore, the final model structure was designed with health states that described the patient condition as a combination of both lung function (FVC%pred) and exacerbation (Figure 36).

Figure 36: Model structure



Note numbers in diagram relate to FVC%pred

Cohort transitions

A Markov model was analysed by cohort analysis using time-dependent and constant transition probabilities to describe decline in lung function and incidence of first acute exacerbations over the cohort's lifetime. A set of mutually exclusive health states was defined. It was assumed that each state represents the full health condition of its members. Possible transitions were defined between each of the health states and the probability of each transition occurring within a set period of time (a cycle) was assigned to each possible transition.

The cohort entered the model at different FVC%pred health states without exacerbation. The possible transitions were:

- death,
- loss of lung function (progression to a health state with lower FVC%pred),
- exacerbation,
- loss of lung function combined with exacerbation,
- remaining in the same health state.

Transitions were evaluated over a 3-month period. This cycle length was consistent with the clinical trial duration between observation timepoints.

Half-cycle correction was applied to all costs and health benefits.

It was assumed that once progressed to a lower FVC%pred the cohort could not regress back to health states with improved lung function (higher FVC%pred). Moreover, once an exacerbation occurred, the cohort could not move back to a health state without exacerbation and would continue in the health states with exacerbation history. Similar assumptions were made in the NICE model for PFN (NICE, 2013[17]).

Loss of lung function was defined as a 10-points decrease in FVC%pred within 3 months, which is in line with clinical expert opinion and the literature (Raghu *et al.*, 2011[13], Noble, 2011[40]).

Death could occur in two ways:

a) at any point in the model (and from any health state) based on the survival analysis of the clinical trial data

b) at the point that patients reached a level of FVC%pred of 30%, which was assumed to be an unsustainable level of lung function (National Institute for Health and Care Excellence, January 2013[76]).

The transitions used in the model and their respective clinical trial inputs are summarised in Table 96 below.

The relationship of health states and interdependence of events and transitions was studied with different analyses of the clinical trial data. In particular the dependencies of mortality and exacerbation and mortality and lung-function status (FVC%pred category) were explored. The final structure of the model transitions, with overall survival and exacerbation risks being dependent on treatment allocation and lung function being dependent on a number of predictors, was deemed to provide the best balance between the available evidence, and flexibility of the model for sensitivity analysis. The structure takes account of a patient's risk of decline in lung-function over time and their risk of exacerbation; the resulting costs and quality of life associated with this can then be captured.

Table 96: Model transitions and inputs

Transition	Input	Source
Mortality	Survival analysis of time to death and Conditional to progression to FVC%pred of 30%	Pooled phase II clinical trial (TOMORROW, 1199.30) and phase III (INPULSIS, 1199.32 and 1199.34) clinical trials data for PBO risk NMA OR values for NDB, PFN
Loss of lung function (disease progression)	Survival analysis of time to progression of FVC%pred (progression defined as a 10% point decline)	Phase III (INPULSIS, 1199.32 and 1199.34) clinical trials data for PBO risk NMA OR values for NDB, PFN
Exacerbation	Survival analysis of time to exacerbation (investigator reported and adjudication committee exacerbations, see section 0 for definitions)	Phase III (INPULSIS, 1199.32 and 1199.34) clinical trials data for PBO risk NMA OR values for NDB, PFN

FVC, Forced Vital Capacity; PBO, Placebo; NDB, nintedanib; PFN, pirfenidone; NMA, network meta-analysis; OR: odds ratio

Table 97: Features of the *de novo* analysis

Factor	Chosen values	Justification
Time horizon	Lifetime	Set to cover the lifetime of the patients, in order to fully incorporate the costs and health outcomes of IPF
Were health effects measured in QALYs; if not, what was used?	Yes	NICE reference case
Discount of 3.5% for utilities and costs	Yes	NICE reference case
Perspective (NHS/PSS)	Yes	NICE reference case

*PSS, personal social services; QALYs, quality-adjusted life years; IPF, idiopathic pulmonary fibrosis

5.2.4 Intervention technology and comparators

As discussed in section 5.2.1, the indication outlined in the marketing authorisation for nintedanib is modelled as to be consistent with the available data. The model

implements a comparison of nintedanib vs. pirfenidone (and also vs. BSC) for the same population (adults with IPF). This differs from the reimbursed population for pirfenidone recommended by NICE; the NICE guidance is for patients with 50-80% of their predicted FVC (NICE, 2013[89]). The model compares all adults with IPF in both arms because of a) the additional complexity required within the model of having the pirfenidone arm made up of a mixture of patients on BSC and pirfenidone b) there is evidence to suggest that pirfenidone is used in patients with an FVC >80% predicted (Chaudhuri N, 2014[90]) and c) TA282 has been scheduled into NICE's Technology Appraisal work programme for reappraisal to investigate the use of pirfenidone in patients with an FVC greater than 80% predicted (NICE, 2014[10]).

The intervention and comparators included within the model reflect the decision problem described in section 1.1.

5.2.5 Continuation rules

NICE guidance for pirfenidone currently states that patients incurring a greater than 10% decline in their predicted FVC in one year should discontinue (NICE, 2013[17]). In the base case, the cost effectiveness model does not employ this discontinuation rule for pirfenidone, however it is investigated in the sensitivity analyses. This discontinuation rule is not included in the base case because the clinicians at the advisory board carried out on 23rd April 2014 were highly critical of this stopping rule and considered it difficult to impose (Boehringer Ingelheim, 2014[91]).

5.3 *Clinical parameters and variables*

5.3.1 Clinical data incorporation into model

The economic model was designed to accommodate available evidence from the phase III clinical trial (Richeldi, 2014[4]), to depict an accurate representation of the patient condition, and to allow easy adaptation for sensitivity analysis of the main inputs.

Treatment efficacy transition probabilities

Source of evidence

The model captured three types of transitions related to treatment efficacy: mortality (overall survival), acute IPF exacerbations, and decline of lung function (progression based on FVC%pred).

The baseline risk of mortality, disease progression, and acute exacerbations was derived from patients in the PBO arm of both clinical trials (Richeldi, 2014[4], Richeldi, 2011[29]) (representing BSC; as is usual for a placebo-controlled clinical trial, the protocol of the INPULSIS clinical trials provided rules for the use of concomitant medication during the trial period. In general, patients were allowed to use a range of

background medication that closely resembled “best supportive care” in this disease). The risk of those events was extrapolated beyond the observed trial follow-up period using a statistical analysis presented in the following sections.

The relative effectiveness of NDB and PFN was obtained from an NMA performed on these matching outcomes (see Section 4.10). In the cost-effectiveness model, the efficacy of NDB and PFN was calculated by assuming proportional hazards and applying an odds ratio (OR) to the baseline risk (BSC arm). The following sections present first the calculations related to the baseline risk and secondly the estimates of relative efficacy for the active treatment comparators.

Survival Analysis

Using the individual patient data (IPD) from the phase III INPULSIS clinical trials (Boehringer Ingelheim, 2014[92]) a parametric model was fitted to determine the transition probabilities for the BSC arm from the start of the model. The use of parametric modelling is common when the analysis needs to extend beyond the clinical trial observational period. The base-case analysis assumes that the parametric model is applied for the full duration of the economic model; that also includes the first year of the analysis where clinical trial data were available. This allows a more robust representation of uncertainty from the trial results and a formal exploration via a probabilistic sensitivity analysis (PSA).

Five different parametric models have been explored for mortality, exacerbation, and progression probabilities: exponential, Gompertz, lognormal, loglogistic, and Weibull. All parametric models were fitted using the “streg” command in STATA (data analysis and statistics software).

Goodness of fit was assessed using the statistical criteria – AIC (Akaike Information Criterion). In general, the smallest AIC value represents the best-fit parametric model. The magnitude of the absolute AIC value is not meaningful in itself as it depends on the set of data. Models with an AIC value within 2-3 points of the lowest one (best option) could also be seen as suitable candidates.

As a second step the results of the parametric modelling were compared with evidence from the literature (visual inspection/face validity). A number of criteria were considered for choosing the parametric model for the economic analysis:

a) the AIC value (smallest value preferred)

b) face validity/visual inspection and comparison with published cohorts

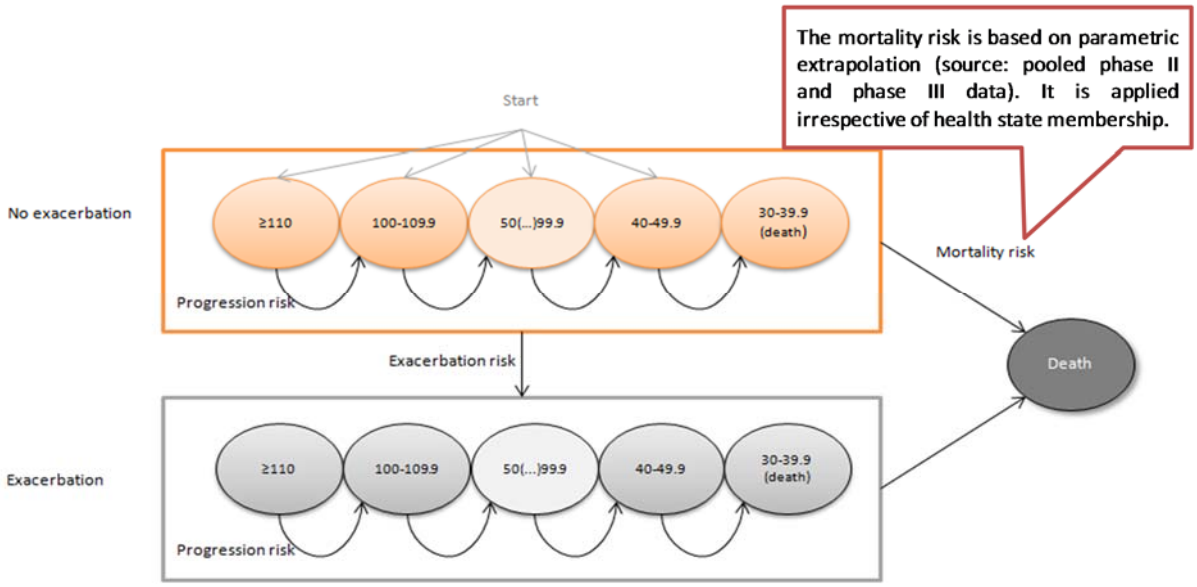
c) precedent in the literature (where evidence was available)

d) clarity of understanding and computation

Overall survival

This section presents the parameters used for overall survival in the model (Figure 37). Note model structure is discussed in more detail in section 12.

Figure 37: Model structure - overall survival



Parametric model extrapolation:

Parametric model extrapolation using pooled phase II (Richeldi, 2011[29]) and phase III (Richeldi, 2014[4]) clinical data was used to derive overall survival parameters for the BSC arm. The pooled data were used to increase the sample size, and strengthen the results, given the similarities between the phase II and phase III trials. Goodness of fit for five parametric models for overall survival is presented in Table 98.

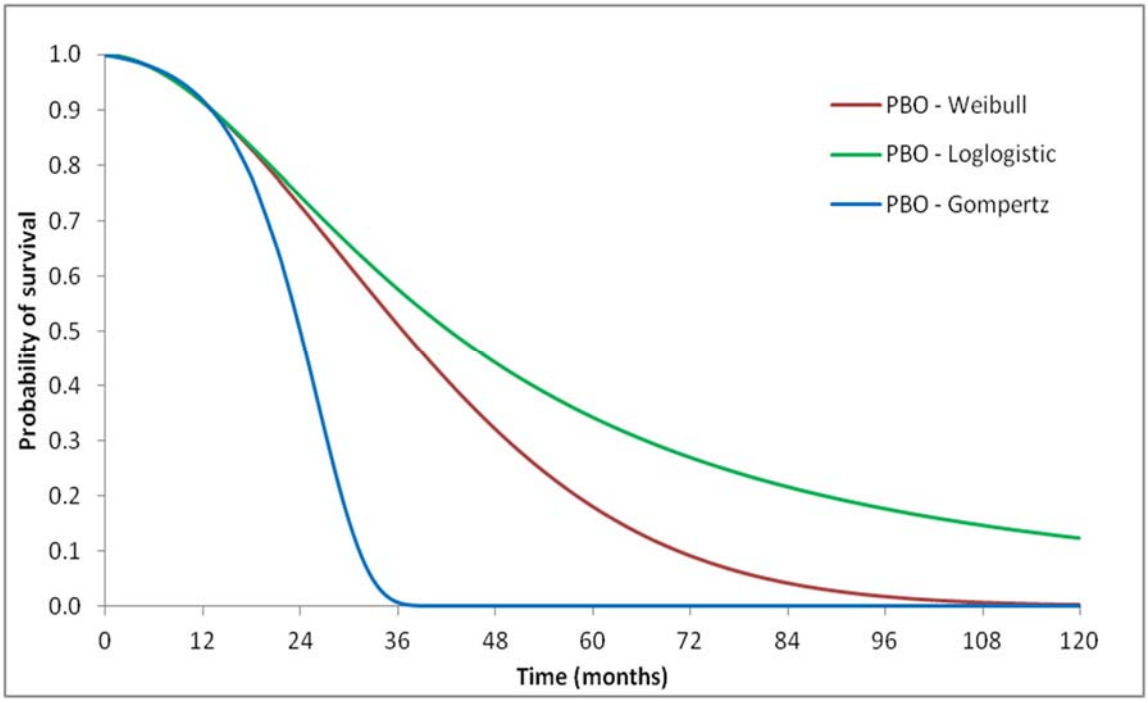
Table 98: Goodness of fit - overall survival

Model	Exponential	Gompertz	Loglogistic	Lognormal	Weibull
AIC	379.86	364.26	366.72	370.64	366.33

The lowest AIC value was reported for the Gompertz distribution, but the Weibull and loglogistic AIC values were very similar.

Overall survival, for the placebo (PBO) arm, was extrapolated to 10 years using the loglogistic, Weibull, and Gompertz parametric models. The 10-year extrapolation is depicted in Figure 38 below.

Figure 38: Overall survival - 10 year extrapolation



The coefficients for the three parametric models are presented in Table 99.

Table 99: Coefficients of parametric models - overall survival

Model	Variable	Coefficient	Std. Err.	Z	P>z	95% Conf. Interval	
Gompertz	***constant***	-9.503	0.375	-25.37	<0.001	-10.237	-8.769
	gamma	0.005	0.001	4.06	<0.001	0.003	0.008
Loglogistic	***constant***	7.164	0.187	38.4	<0.001	6.799	7.530
	ln_gamma	-0.633	0.139	-4.56	<0.001	-0.905	-0.361
Weibull	***constant***	7.217	0.192	37.6	<0.001	6.841	7.593
	ln_p	0.610	0.140	4.36	<0.001	0.336	0.885

The variance-covariance matrices corresponding to each of the three parametric models for overall survival are presented in Table 100.

Table 100: Variance-covariance matrix - overall survival

Model	Variable	*constant*	*Gamma*
Gompertz	***constant***	0.140	
	gamma	0.000	0.000
Loglogistic		***constant***	***ln_gamma***
	constant	0.035	
	ln_gamma	0.023	0.019
Weibull		***constant***	***ln_p***
	constant	0.037	
	ln_p	-0.025	0.020

Overall survival equations

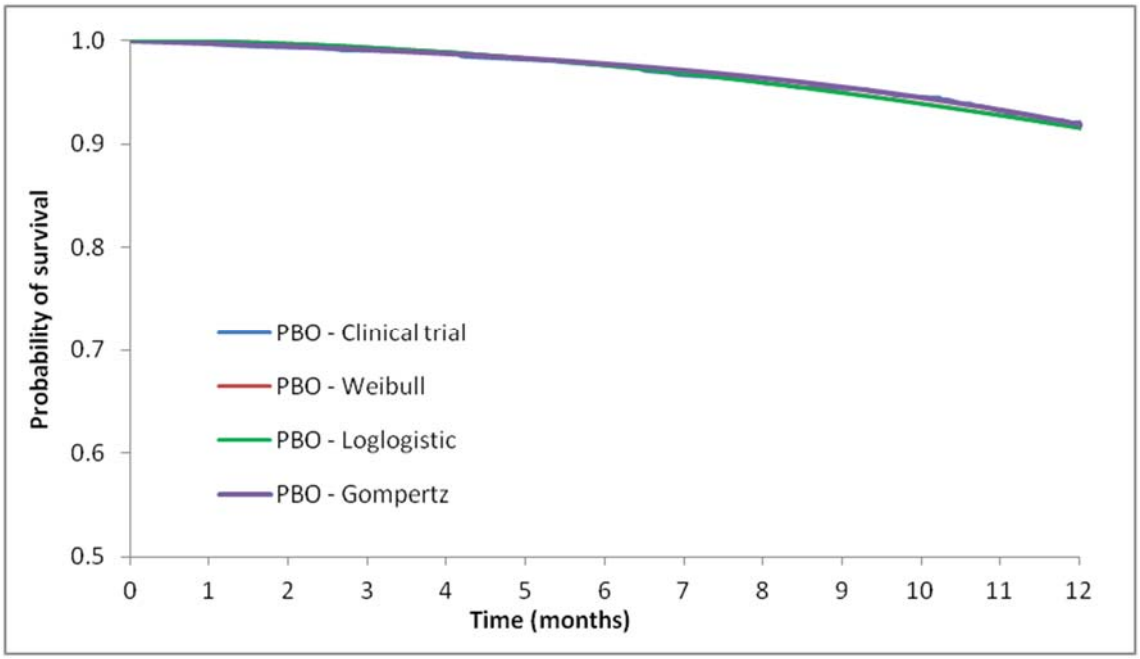
The model uses the coefficients reported in Table 99 above and the following functions, for each of the three parametric models for the PBO arm:

Gompertz: survival(time_in_days)= exp(-exp(cons)*1/gamma*
(exp(gamma*time_in_days)-1))
Loglogistic: survival(time_in_days)=1/{1+(exp(-constant)
*time_in_days)^(1/gamma)}
Weibull: survival(time_in_days)=exp(-exp(-p*constant)*(time_in_days^p))

Fit of parametric models:

Figure 39 compares the fit of the selected parametric models (Gompertz, loglogistic, and Weibull) to the Kaplan-Meier curves from the clinical trial (trial period: 12 months) for the PBO arm.

Figure 39: OS model fit vs. Kaplan-Meier for placebo



*Curves are overlapping and may not be entirely distinguishable.

Validation:

In order to validate the 10-year overall survival extrapolation, the three survival curves (Gompertz, loglogistic and Weibull) were compared to other published data on similar cohorts and in the same disease area. To this end, a targeted literature review in

PubMed was performed to search for recent observational studies following IPF patients over a period of two years or longer. Seven observational studies were identified:

- Kondoh and colleagues, 2010 (Kondoh *et al.*, 2010[93])
- Nathan and colleagues, 2011 (Nathan *et al.*, 2011[85])
- Raghu and colleagues, 2004 (Raghu *et al.*, 2004[94])
- Douglas and colleagues, 1998 (Douglas *et al.*, 1998[95])
- Lederer and colleagues, 2006 (Lederer *et al.*, 2006[81])
- Hamada and colleagues, 2007 (Hamada *et al.*, 2007[80])
- Manali and colleagues, 2008 (Manali *et al.*, 2008[84])

The identified studies had a sample size in the range of 25 to over 1,500. Results of the studies were presented in different subgroups: +/- exacerbations, ethnicity, lung function (positive air pressure), and treatment received. The populations' characteristics were comparable to the patients in the TOMORROW and INPULSIS trials. The baseline characteristics of the cohorts from the NDB clinical studies (Richeldi *et al.*, 2011[1], Richeldi, 2014[4]) and from the selected observational studies are presented in Table 101 below.

Table 101: Baseline population characteristics across trials

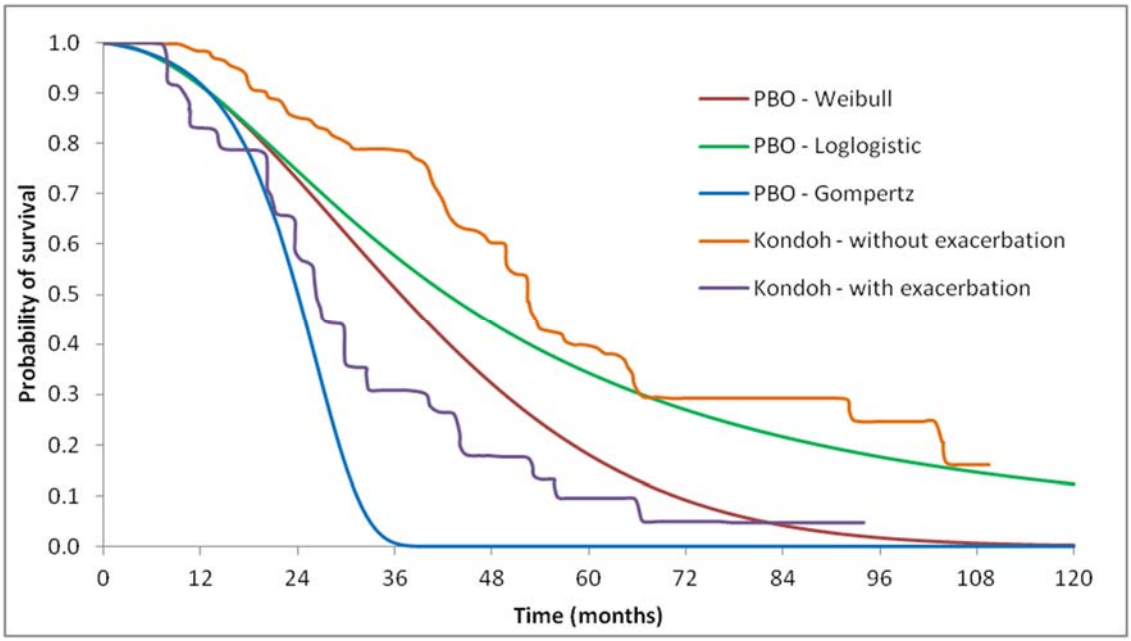
Study	Available subgroups	Patients number	Male, number (%)	Age (SD)	FVC%pred (SD)	DLco% pred (SD)
Clinical trial (phase II)	-	428	320 (74.8%)	65.1 (8.56)	81.3 (17.9)	47.5 (12.6)
Clinical trial (phase III, INPULSIS)		1061	841 (79.26%)	66.76	79.55	N/A
Kondoh et al. (2010)	With acute exacerbations	23	20 (89.96%)	62.9 (8.9)	73.5 (20.9)	57.3 (21.6)
	Without acute exacerbations	51	51 (100%)	64.6 (6.7)	79.1 (18.2)	60.2 (17.6)
Nathan et al.(2011)	-	357	256 (71.7%)	65	61.01	40.96
Raghu et al. (2004)	-	168	111 (66%)	63.4 (8.6)	64.1 (11.3)	36.8 (10.6)
Douglas et al. (1998)	Prednisone	12	10 (83%)	69.9 (4.0)	66.8 (14.9)	46.9 (10.8)
	Colchicine	14	10 (71%)	65.9(12.3)	62.4 (14.8)	51.9 (18.1)
Lederer et al. (2006b)	Black	299	130 (43%)	46 (10)	44 (15)	N/R
	Hispanic	173	88 (51%)	50 (10)	45 (16)	N/R
	White	2163	1408 (65%)	54 (9)	51 (17)	N/R
Hamada et al. (2007)	Normal PAP (<17 mmHg)	37	33 (89.2%)	62 (7)	79 (23)	51 (15)
	High PAP (<17 mmHg)	24	20 (83.33%)	63 (9)	71 (20)	36 (10)
Manali et al. (2008)	-	25	12 (48%)	64 (2)	73 (2)	49 (3)

*PAP – Positive Airway Pressure; mmHg – unit measure for pressure.

The study by Kondoh et al (Kondoh *et al.*, 2010[93]) was particularly useful because of its long follow-up; it was therefore further compared to the OS parametric extrapolations. The study evaluated the frequency, risk factors, and impact of IPF acute exacerbations on survival. It retrospectively studied 74 patients diagnosed with IPF and followed them for a minimum period of 3 years (maximum 9 years) (Kondoh *et al.*, 2010[93]).

A computer programme (Grafula 3 software, version 2.10) was used to extract and digitise the KM curves from the Kondoh study (Kondoh *et al.*, 2010[93]) and present them alongside the survival curves derived from the model for the PBO arm. Figure 40 shows that the Gompertz model underestimated survival in comparison with the observational data. Both the Weibull and the loglogistic models appeared within the two KM curves that reflected the mortality of patients with and without acute exacerbations. Since in the nintedanib trial cohort the frequency of an acute exacerbation event was low, the loglogistic model (closer to the no exacerbation cohort of the Kondoh study) would be more appropriate for the base-case analysis. Furthermore, it was noted that the Kondoh et al. cohort was followed up from 2000 to 2009 (maximum), whereas the model used in this report extrapolated outcomes for patients from 2015 to 2025. Therefore, with improvements in the management of IPF patients over the next 10 years, the KM curve of patients with exacerbation could be an overestimate of mortality. The Weibull and Gompertz models were explored in sensitivity analysis.

Figure 40: Overall survival extrapolations for the placebo arm compared against Kondoh et al 2010



Overall survival for NDB and PFN

In order to estimate probabilities of death for the active treatments (NDB and PFN), the baseline risk was synthesised with the results of an NMA (See Section 4.10 and Appendix B). For each comparator, the baseline mortality risk was multiplied by the corresponding OR value from the NMA, assuming a constant relationship over time.

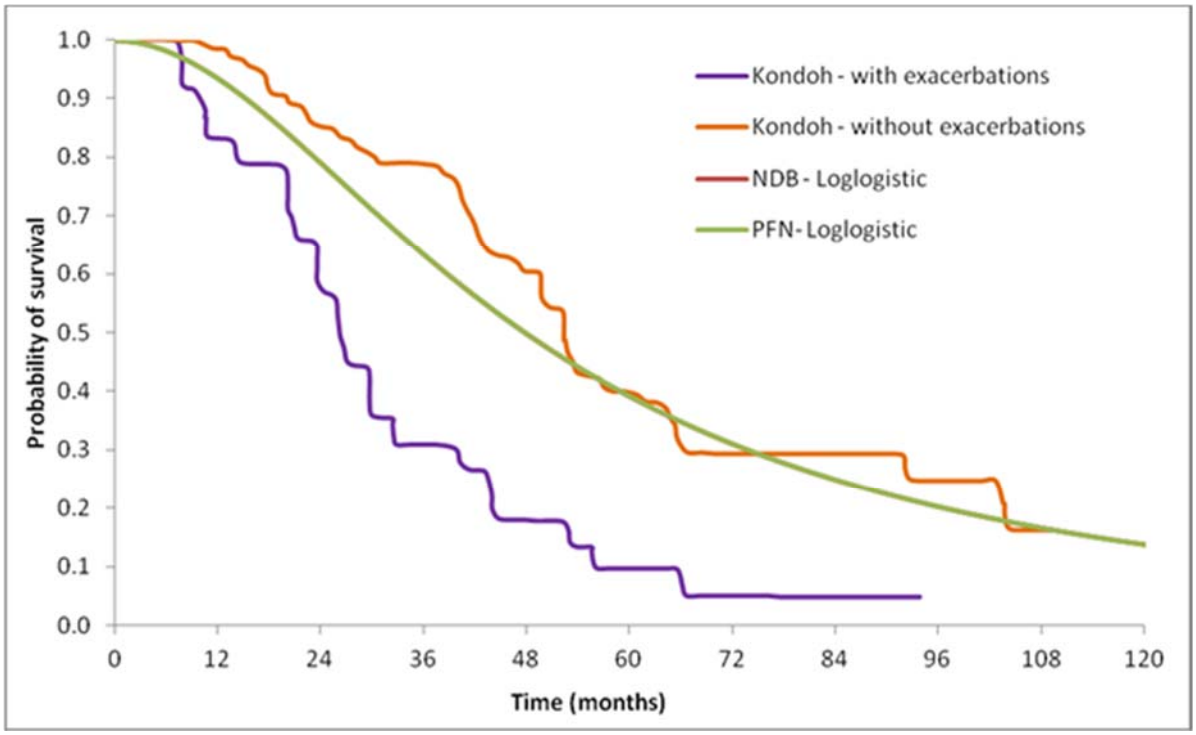
The OR values used in this model are presented in Table 102 below. Since there was no variation observed in the NMA results scenario analyses, the base-case analysis used the all-evidence scenario; other OS NMA scenarios were explored in sensitivity analyses. Details of this analysis and results are presented elsewhere (See Appendix B).

Table 102: OR values for OS - NMA results

Comparison	OR median value	95% Confidence interval
NDB vs. PBO	0.70	0.45 - 1.10
PFN vs. PBO	0.70	0.46 – 1.05

For comparison, the OS loglogistic extrapolations for the active treatment arms (NDB and PFN) compared against the Kondoh study with and without exacerbation cohorts are presented in Figure 41.

Figure 41: Overall survival loglogistic extrapolations for the active arms compared against Kondoh et al

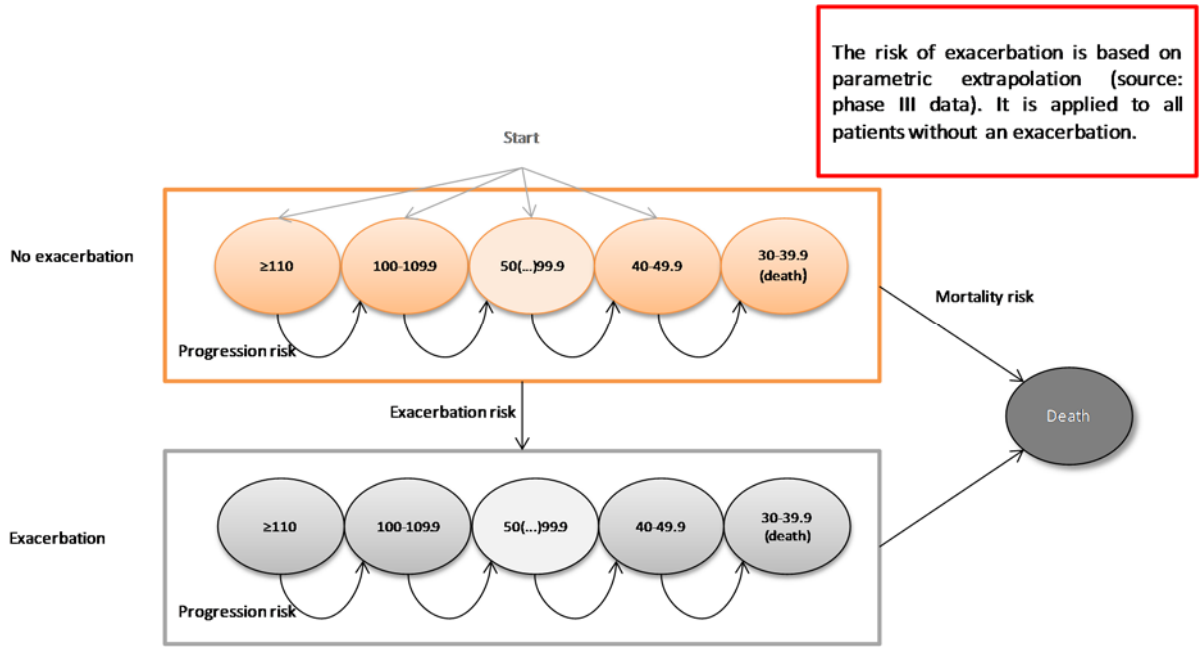


*Curves for NDB and PFN are overlapping

Time to first acute exacerbation

Time to first acute exacerbation is relevant to the following transitions presented in Figure 42 below.

Figure 42: Model structure - exacerbations



Available clinical evidence

Time to first acute IPF exacerbation over 52 weeks, was a key secondary endpoint in the phase III clinical trials (Richeldi, 2014[4]). The parameter was recorded in two ways:

1. based on investigator-reported adverse events, according to the selection criteria described by the trial protocol;
2. based on a blinded adjudication committee, which reviewed all cases that fulfilled the criteria of an acute IPF exacerbation according to the investigator. The adjudication committee classified each case as “confirmed,” “suspected,” or “not” acute IPF exacerbation. Note: cases were adjudicated as confirmed IPF exacerbations if all criteria in the definition of an acute IPF exacerbation were fulfilled. Suspected IPF exacerbations were clinical deteriorations of IPF that were of unknown cause but did not fulfil the above criteria due to missing data.

Parametric model extrapolation: investigator reported

Data for time-to-first-acute-exacerbation from the phase III clinical trial were used in the parametric model extrapolation for the PBO arm [post-hoc analysis of INPULSIS data; (Richeldi, 2014[4]). The goodness of fit for five parametric models for time-to-first-acute-exacerbation (investigator reported) is presented in Table 103.

Table 103: Goodness of fit: time-to-first-acute-exacerbation (investigator reported)

Model	Exponential	Gompertz	Loglogistic	Lognormal	Weibull
AIC	275.65	274.40	274.93	276.29	274.83

Since the AIC differences between the models were very small (1-2 AIC points), to ensure model parsimony the exponential model was selected, assuming a constant hazard. The alternatives would have required increased complexities of the decision analytic model in order to keep track of time-in-state before extrapolation.

The coefficient of the exponential model is presented in Table 104.

Table 104: Coefficient for exponential model - exacerbation (investigator reported)

Variable	Coefficient	Std. Err.	Z	P>z	95% Conf. Interval	
constant	8.430	0.177	47.69	<0.001	8.083	8.776

The estimated exacerbation risk applied in the model is 1.95% for the PBO arm per cycle (3 months).

Parametric model extrapolation: adjudication committee

A similar exercise was performed on the exacerbation data selected based on adjudication committee considerations. Goodness of fit for five parametric models for time-to-first-acute-exacerbation (adjudication committee) is presented in Table 105.

Table 105: Goodness of fit: time-to-first-acute-exacerbation (adjudication committee)

Model	Exponential	Gompertz	Loglogistic	Lognormal	Weibull
AIC	222.52	222.09	222.94	224.14	222.85

Since the AIC differences between the models were very small (1-2 AIC points), the exponential model was used for consistency and comparability with the previous type of exacerbation (investigator reported). The coefficient of the exponential model is presented in Table 106.

Table 106: Coefficients for exponential model - exacerbation (adjudication committee)

Variable	Coefficient	Std. Err.	Z	P>z	95% Conf. Interval	
constant	8.724	0.204	42.74	<0.001	8.324	9.124

The estimated baseline exacerbation risk applied in the model is 1.47% for the PBO arm per cycle (3 months).

The model uses the coefficients reported in Table 104 (investigator reported exacerbations) and Table 106 (adjudication committee) above and the following equations for the exponential model exacerbation function:

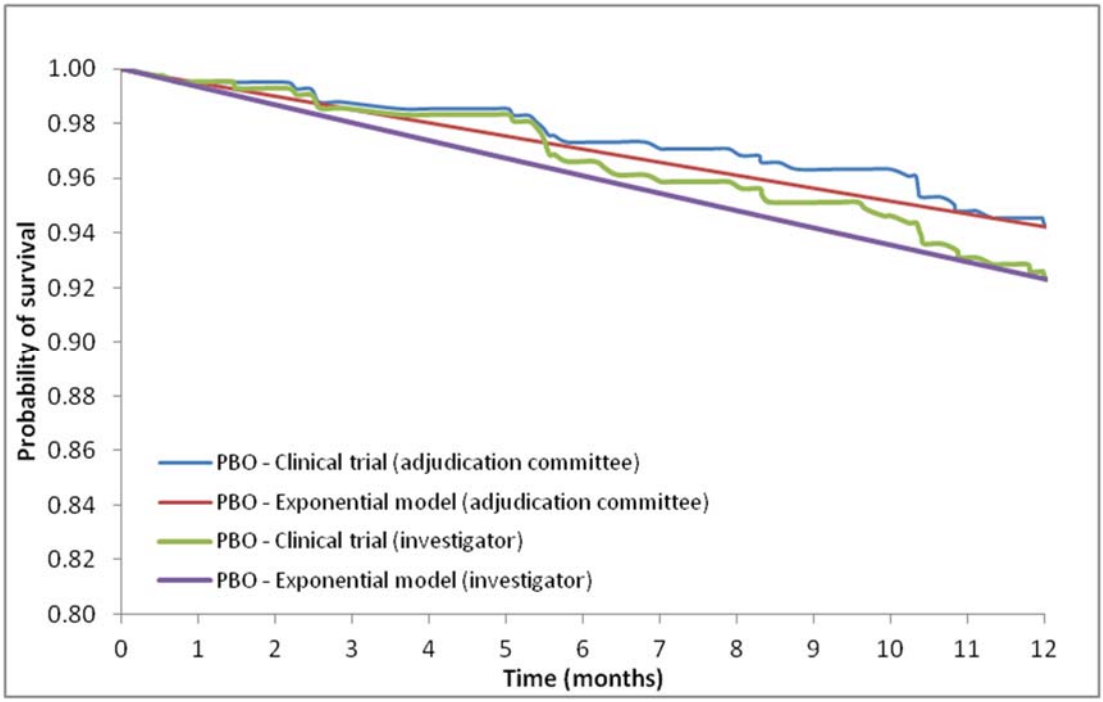
- Placebo arm: survival(time_in_days)=exp(-exp(-constant)*(time_in_days))

The model used the investigator reported exacerbation risk as base-case for consistency with the exacerbation definition used in the NMA exercise. The impact of the adjudication committee estimate on the model results was explored in sensitivity analysis.

Fit of parametric models

Similarly to the overall survival case, Figure 43 compares the fit of the exponential parametric models for both types of exacerbations to the Kaplan-Meier curves from the clinical trial (study period: one year) for time to exacerbation.

Figure 43: Exacerbation model fit vs. clinical trial Kaplan-Meier



*Curves are overlapping and may not be entirely distinguishable.

Exacerbations for NDB and PFN

Similarly to the OS exercise, the risks of exacerbation for NDB and PFN were informed by the NMA ORs applied to the baseline BSC risk, assuming a constant relationship over time. The NMA scenario used in the model excluded Azuma et al. (Azuma *et al.*,

2005[48]), Taniguchi et al. (Taniguchi, 2010[49]), and Homma et al. (Homma, 2012[54]) (See Appendix B, Scenario 3) because inclusion of these studies introduced a high level of heterogeneity to the other PFN and NAC studies. Furthermore, the fact that the above mentioned studies were Japanese-only trials with different treatment dosages than the European and US studies supported their exclusion from the base-case analysis. The fixed effects model was used as it performed better than the random effects model. Details of this analysis and results are presented elsewhere (See Appendix B). The OR values for each active treatment vs. PBO are presented in Table 107 below. Note other scenarios were explored in sensitivity analyses (see Appendix A).

Table 107: OR values for exacerbations - NMA Results

Comparison	OR median value	95% Confidence interval
NDB vs. PBO	0.56	0.35 – 0.89
PFN vs. PBO	1.01	0.22 – 4.50

* Please note that the NMA was only performed on investigator reported exacerbations.

Recurrent exacerbations

The economic model assumes that patients who experienced at least one exacerbation event are at risk of recurrent events. Due to lack of evidence on the incidence of recurrent events, the model assumes the same risk as for patients that have not had an exacerbation. Given that in general the outlook of patients with an acute IPF exacerbation is very poor, this is probably a conservative assumption. Furthermore, the low overall frequency of exacerbations combined with the limited remaining lifetime of the patients in the model results in a very low risk for recurrent exacerbation.

Loss of lung function

Patient disposition at start of model

For a more accurate reflection of the clinical trial setting, the cohort entered the model at different FVC%pred health states. The disposition of the cohort at the start of the model is presented in Table 108 and is based on phase III clinical trial data (Richeldi, 2014[4]).

Table 108: Patient distribution at the start of the model

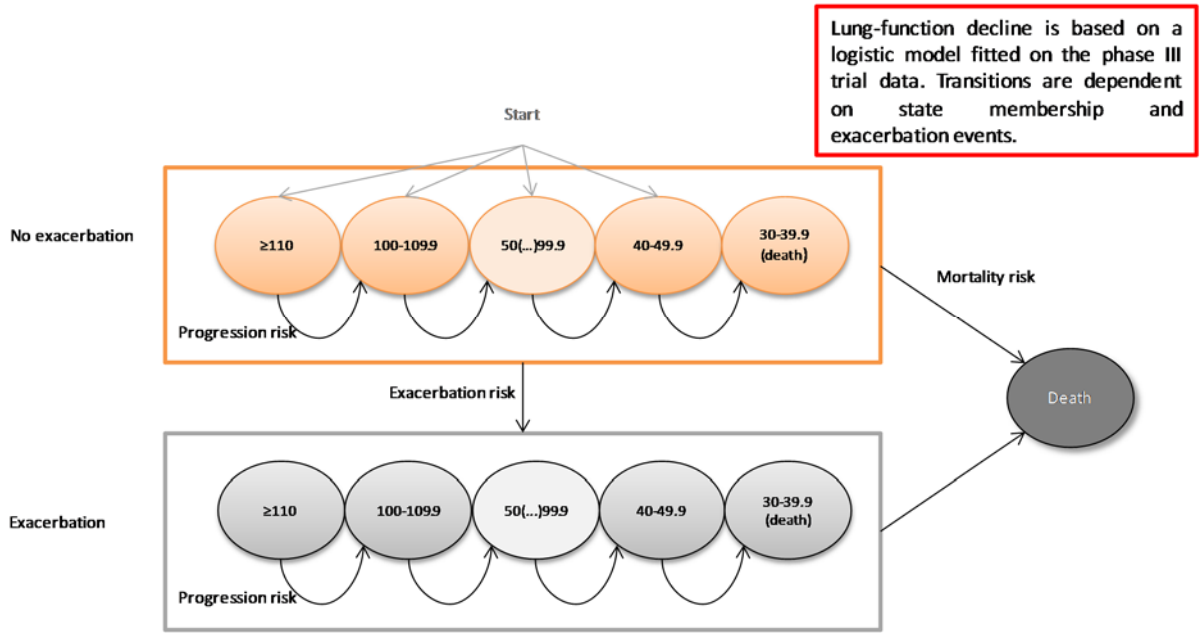
Health state	Distribution (%)
FVC%pred 110 and above	6.00%
FVC%pred 100-109.9	7.50%
FVC%pred 90-99.9	12.38%
FVC%pred 80-89.9	19.79%
FVC%pred 70-79.9	20.26%

Health state	Distribution (%)
FVC%pred 60-69.9	20.73%
FVC%pred 50-59.9	12.76%
FVC%pred 40-49.9	0.56%

*Note: all patients entered the model without recent history of exacerbation events

The model structure for loss of lung function is presented in Figure 44:

Figure 44: Model structure - loss of lung function



Phase III clinical trial patient-level data on lung function decline, defined as a 10-point drop in FVC%pred, were analysed for this model transition for the BSC arm (Richeldi, 2014[4]). A logistic regression model was used to capture several predictors of lung function decline and allows the analysis of recurrent events and the incorporation of additional covariates that may influence the probability.

Predictors of lung function:

a) Current FVC%pred state (FVC%pred value of patient before progression)

b) Exacerbation (i.e. progression before and after an exacerbation event)

c) Exacerbation during the analysed 3-month period (whether it occurred or not)

d) Interaction of the time variable with FVC%pred

In the analysis of the data, predictor a) was statistically significant. The other predictors were not statistically significant. Therefore, predictor a) was used in the base case model.

Lung function decline modelling (without exacerbation)

The logistic model used is *P=constant + b* fvcpp_prev~s,* where b is the value of the FVC%Pred at the start of the interval, *fvcpp_prev~s* is the current FVC%Pred covariate.

The coefficients of the model and variance-covariance matrix are presented in Table 109 and Table 110 below.

Table 109: Coefficients for lung function decline (without exacerbation)

Variable	Coefficient	St. Error	p-value	95% Confidence Interval	
fvcpp_prev~s	0.016	0.006	0.004	0.005	0.028
constant	-4.180	0.552	<0.001	-5.262	-3.097

*Where: *fvcpp_prev~s* is the current FVC%pred covariate.

Table 110: Variance-covariance matrix for lung function decline (without exacerbation)

Variable	***fvcpp_prev~s***	***constant***
fvcpp_prev~s	0.000	
constant	-0.003	0.305

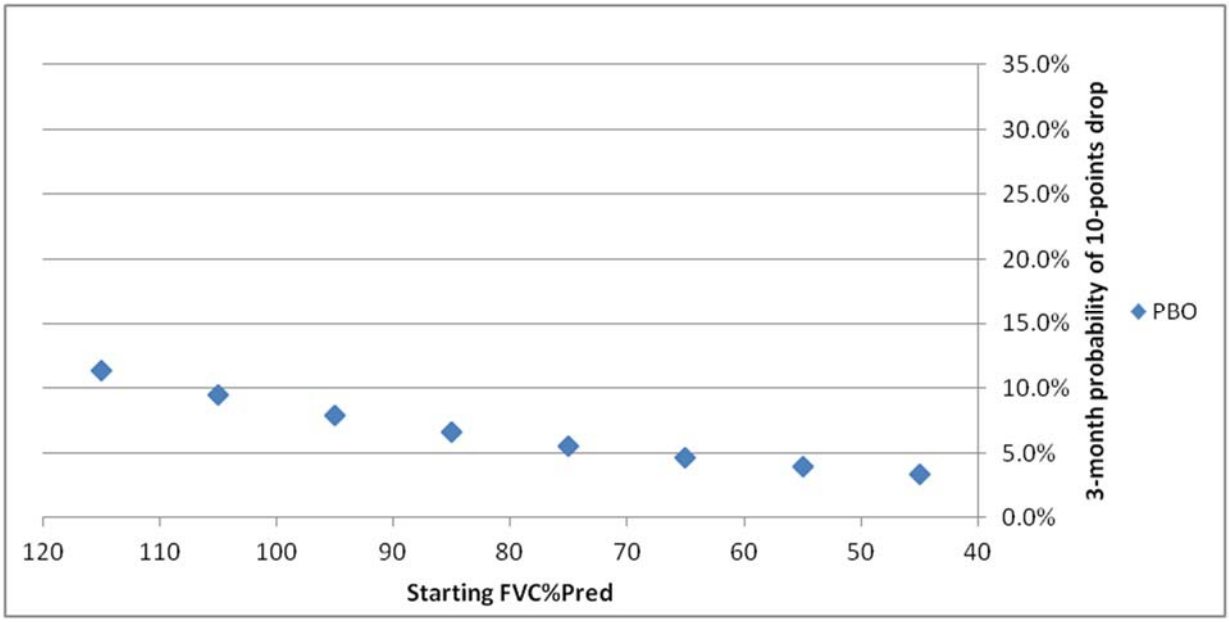
*Where: *fvcpp_prev~s* is the current FVC%pred covariate.

The coefficients in Table 109 are statistically significant. The probabilities of progressing during a 3-month interval for each FVC%pred value at the start of the interval are presented below in Table 111 and graphically in Figure 45. Note that the loss of lung function was modelled in categories of 10 percentage points. The model used the median FVC%pred value within each category (probabilities shown in Table 111) for each value; note the lowest and highest values of each FVC%Pred categories were explored in sensitivity analysis. A diminishing effect in progression as lung function was lost was observed.

Table 111: 3-month probability of progression (model without exacerbation, placebo arm)

FVC%pred at start of interval	PBO arm risk
115	11.3%
105	9.4%
95	7.9%
85	6.6%
75	5.6%
65	4.7%
55	3.9%
45	3.3%

Figure 45: Progression probabilities (model without exacerbation, placebo arm)



Lung function decline modelling (with exacerbation)

The coefficients and variance-covariance matrix of the loss of lung function model including the exacerbation interaction (i.e. following an exacerbation event) are presented in Table 112 and Table 113 below.

Table 112: Coefficients for lung function decline (with exacerbation)

Variable	Coefficient	St. Error	p-value	95% Confidence Interval	
fvcpp_prev~s	0.016	0.005	0.003	0.006	0.026
_lexa_1	0.814	1.066	0.445	-1.276	2.904
Constant	-4.068	0.456	<0.001	-4.962	-3.174

*Where: *fvcpp_prev~s* is the current FVC%pred covariate, and *_lexa_1* is the exacerbation related covariate.

Table 113: Variance-covariance matrix for lung function decline (with exacerbation)

Variable	***fvcpp_prev~s***	***_lexa_1***	***constant***
fvcpp_prev~s	0.000		
_lexa_1	0.000	1.137	
Constant	-0.002	-0.049	0.208

*Where: *fvcpp_prev~s* is the current FVC%pred covariate, and *_lexa_1* is the exacerbation related covariate.

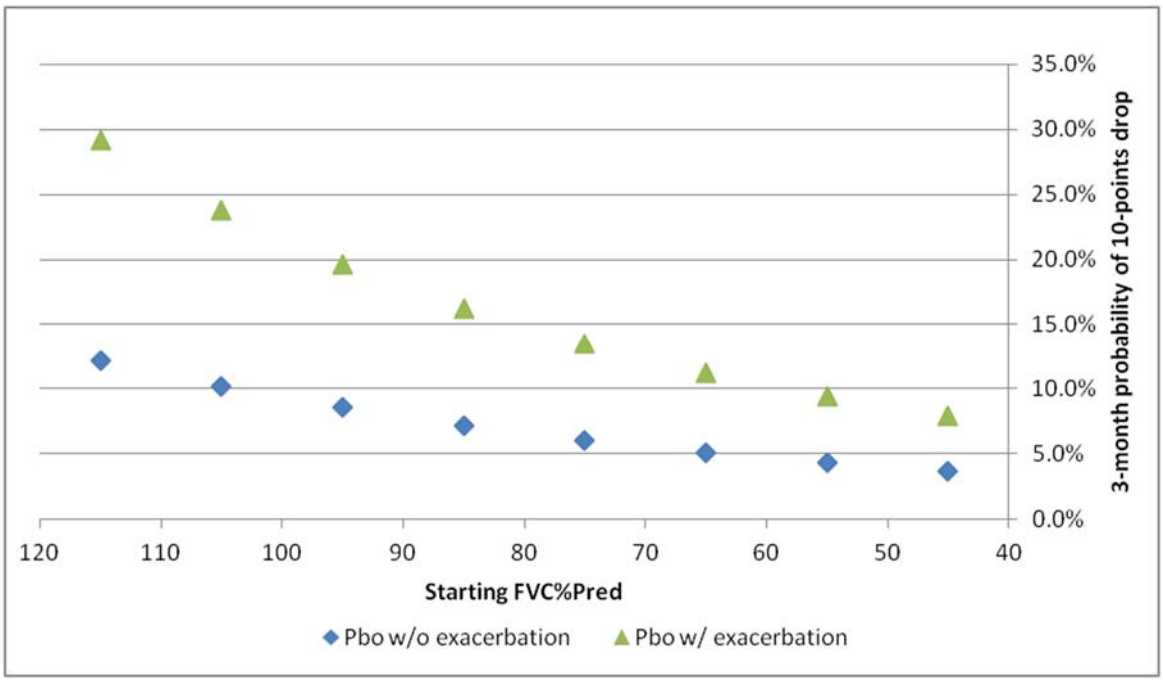
The probabilities of progressing during a 3-month interval for each FVC%pred value at the start of each interval are presented below in Table 114 and graphically in Figure 46. Similarly to the model without exacerbations, a diminishing effect in progression as lung function was lost was observed. Nevertheless, the absolute risk of progression was considerably higher following an exacerbation. Since the exacerbation coefficient was not statistically significant (see Table 112, p-value = 0.445), the model without the

exacerbation coefficient was considered as the base-case for the loss of lung function outcome. The second model considering the exacerbation covariate was explored in sensitivity analysis. Furthermore, note that the model of loss of lung function used the definition of exacerbation as reported by the investigator (10 observations). The alternative definition of adjudication committee exacerbations was not used due to an even lower number of observations.

Table 114: 3-month probabilities of progression (model with exacerbation, placebo arm)

FVC%pred at start of interval	No exacerbation at start of interval	Intervals starting after first exacerbation
115	12.2%	29.2%
105	10.2%	23.8%
95	8.5%	19.6%
85	7.2%	16.2%
75	6.0%	13.5%
65	5.1%	11.3%
55	4.3%	9.4%
45	3.7%	7.9%

Figure 46: Progression probabilities (model with exacerbation)



Loss of lung function for NDB and PFN

The risks of loss of lung function for NDB and PFN were informed by the NMA ORs applied to the baseline BSC risk, assuming a constant relationship over time. The NMA scenario implemented in the model excluded King et al. (King *et al.*, 2014[41]) (scenario 2 Appendix B) from the data because inclusion of this study introduced a high level of

heterogeneity when compared with Noble et al. (Noble, 2011[40]); other loss of lung function NMA scenarios were explored within the model (See Table 182). King et al. (King *et al.*, 2014[41]) was also excluded due to different patient characteristics, as its cohort had a lower average FVC%Pred than the other studies. The authors stated that the ASCEND selection criteria were chosen to "[…] allow enrolment of patients with an increased risk of disease progression" (King *et al.*, 2014[41]). Among other inclusion criteria they reported: a) IPF diagnosed at least 0.5 years before visit 2, b) FVC 50-90% predicted, and c) FEV1/FVC ≥ 0.8. The fixed effects model was used as it performed better than the random effects model. Details of this analysis and results are presented elsewhere (see Appendix B). The OR values for each active treatment vs. PBO are presented in Table 115 below.

Table 115: OR values for loss of lung function - NMA results

Comparison	OR median value	95% Confidence interval
NDB vs. PBO	0.54	0.42 – 0.69
PFN vs. PBO	0.69	0.47 – 1.00

Validity of model projections with clinical trial data

The distribution of the patients in FVC%pred categories after one year in the model was compared to the clinical trial results. In general, the modelled FVC%pred distribution closely matched the clinical trial observations in the BSC arm (Figure 47) and NDB arm (Figure 48).

Figure 47: Model synthesis and clinical trial results for the BSC arm (patients per health state at one year)

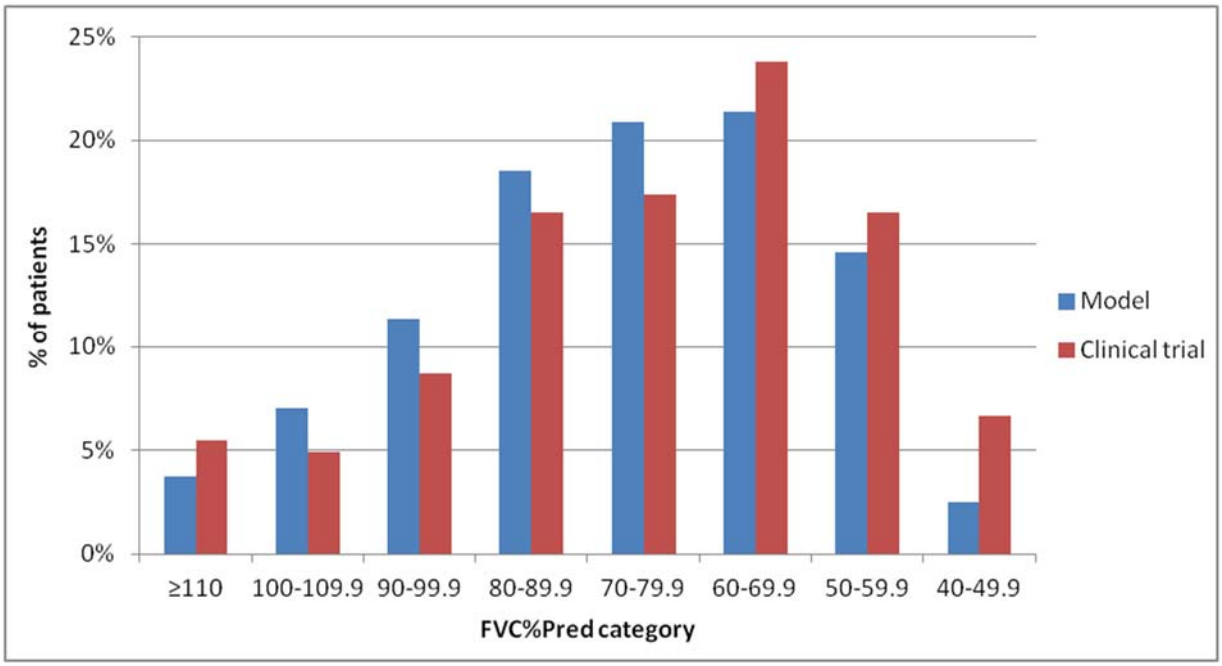
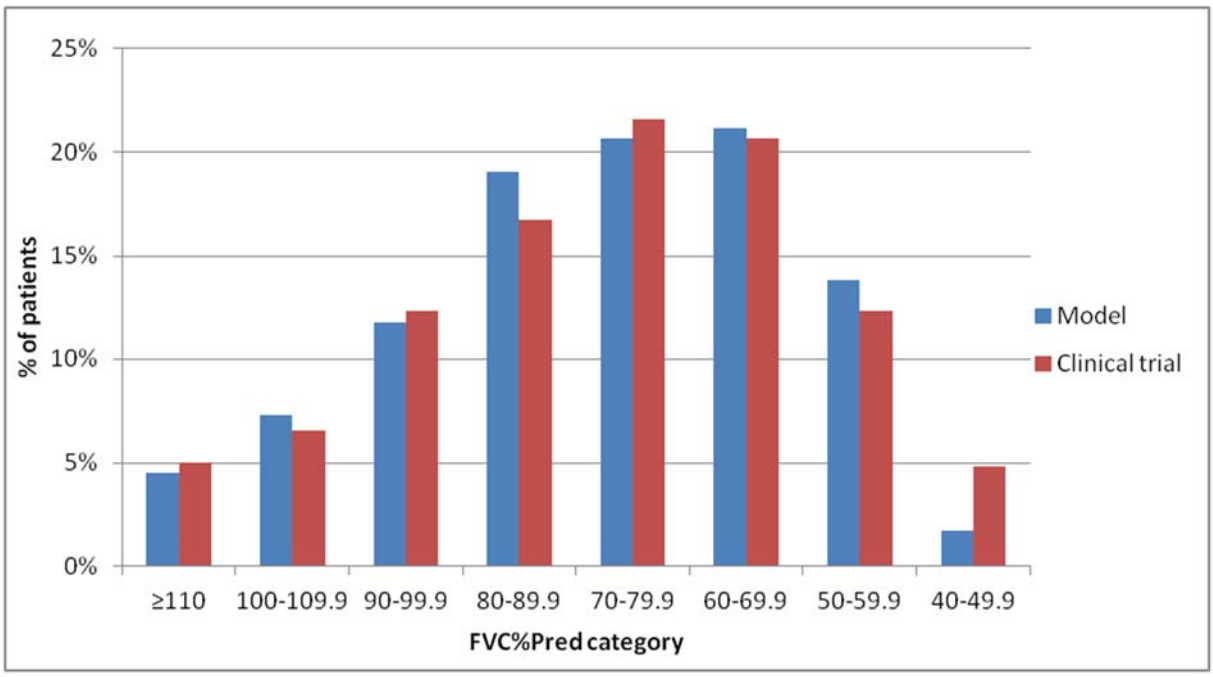


Figure 48: Model synthesis and clinical trial results for the nintedanib arm (patients per health state at one year)



Safety and treatment discontinuation

Selection of adverse events in the model

Safety in the model was analysed by selecting events (individual or grouped in classes) that satisfied all of the following criteria in at least one of the clinical studies considered:

- AEs with a significant impact on costs and QALYs: assumed to be those that were severe or serious
- AEs with an incidence greater or equal to 5%
- AEs with an incidence of 1.5 times greater between the two arms

In addition, the following AEs of particular focus to clinicians were implemented in the model regardless of whether they met above criteria or not:

- For NDB: gastrointestinal perforation
- For PFN: photosensitivity and rash

Serious adverse events

Based on the search of the individual studies and considering the above criteria, two AEs were common across both comparators (NDB and PFN); serious cardiac events

and serious gastro-intestinal (GI) events. There were no other events that qualified the above criteria and were common across any two comparators.

Similarly to the efficacy parameters previously presented, the incidence of each serious adverse event (SAE) was calculated for the PBO arm, using the number of patients experiencing each event throughout any of the clinical trials. The patient numbers and respective risks for each SAE are presented below for the PBO arm (Table 116).

Table 116: Incidence and risk of SAEs in the PBO arm

SAE	Serious cardiac events	Serious GI events
Total number of patients	423	
Number of patients experiencing the event	23	7
Risk of SAE per cycle (%)	1.39%	0.42%

*Source: INPULSIS Summary of Clinical Safety [2014]. Note that the observation period is 52 weeks. The risk values are cycle length adjusted.

The SAE risks for the active treatments (NDB and PFN) were estimated using the OR obtained from the NMA (Table 117).

The NMA scenario implemented in the model for serious cardiac events excluded Richeldi et al. (Richeldi *et al.*, 2011[1]) (See Appendix B, Scenario 2) because inclusion of this study introduced a high level of heterogeneity when compared with Richeldi et al. 2014 (Richeldi, 2014[4]). Regarding serious GI events, the all-evidence scenario was used. Other SAE NMA scenarios were explored in the sensitivity analyses (See Appendix B). Both sets of NMA results were generated by the fixed effects model as it performed better than the random effects model. Details of this analysis and results are presented elsewhere (See Appendix B).

Table 117: OR values for SAEs - NMA results

	Serious cardiac events		Serious GI events	
Comparison	**OR median value**	**95% CI**	**OR median value**	**95% CI**
NDB vs. PBO	0.92	0.53 – 1.63	2.35	1.05 – 5.88
PFN vs. PBO	1.27	0.66 – 2.49	0.60	0.23 – 1.45

Clinically important AEs

Additionally to the adverse events presented above, several other adverse events were implemented in the model as clinically important events. The patient numbers and respective risks for each SAE are presented below for each treatment arm (Table 118 for nintedanib, and Table 119 for pirfenidone).

Table 118: Incidence and risk of clinically important events for the nintedanib arm

Clinically important event	GI perforation events	
Treatment arm	**PBO**	**NDB**
Total number of patients	723	508
Number of patients experiencing the event	2.17	0
Risk of SAE per cycle (%)	0.08%	0%

*Source: FDA NDB prescribing information, value reported as percentage: 0.3%. Note that the observation period is 52 weeks. The risk values are cycle length adjusted.

Table 119: Incidence and risk of clinically important events for the pirfenidone arm

Clinically important event	Photosensitivity reaction		Rash	
Treatment arm	**PBO**	**PFN**	**PBO**	**PFN**
Total number of patients	347	345	347	345
Number of patients experiencing the event	6	42	40	111
Risk of SAE per cycle (%)	0.32%	2.32%	2.19%	6.79%

*Source: NICE STA (NICE, 2013[89]). Note that the observation period is 72 weeks. The risk values are cycle length adjusted

Liver enzyme elevations were assumed to be asymptomatic for patients. The model assumed that when these were detected, they contributed only to the overall discontinuation from treatment, and that there was no disutility or additional costs associated with them. Furthermore, regarding the safety of NDB, a warning appears in the treatment label advising against concomitant use of anticoagulation treatment and NDB (FDA highlights of prescribing information, 2014[96]). Because all NDB clinical trials excluded patients with high risk of bleeding, this risk could not be quantified for further analysis.

Treatment discontinuation

Although the majority of events are of mild or moderate severity, the nature of the adverse events is likely to have a greater effect on treatment tolerability and persistence. It was reported that over the entire clinical trial duration, 24.8% discontinued treatment in the NDB clinical trial (Richeldi, 2014[4]) and 32.9% discontinued in the PFN study (114 weeks, (NICE, 2013[17])).

Similarly to the efficacy parameters previously presented, the PBO overall discontinuation risk was calculated based on parametric modelling extrapolation of clinical trial data, while the discontinuation risks for the active treatments (NDB and PFN) were estimated via OR values obtained from the NMA, assuming proportional hazards.

Data on time-to-discontinuation for PBO were obtained from the phase II and phase III clinical trials to increase the sample size. The data were analysed assuming an exponential model (constant hazard) which was chosen based on having the lowest

AIC value (see Table 120). The analyses were performed excluding discontinuations due to death. It was assumed that patients died before discontinuation if the date of discontinuation coincided with the date of death or if it was the very next day. The coefficient of the exponential model for the PBO arm is reported in Table 121 below.

Table 120: AIC values for the current UK model (NMA)

Model	Exponential	Weibull	Gompertz	Lognormal	Loglogistic
AIC	730.5	732.4	730.5	743.9	734.1

Table 121: Coefficient for exponential model - overall discontinuation, placebo arm

Variable	**Coefficient**	**Std. Err.**	**P>z**	**95% Conf. Interval**	
cons	7.380	0.098	<0.001	7.190	7.570

The estimated overall discontinuation risk for the PBO arm was calculated to be 5.4% per month (exponential model, constant risk). The probability of discontinuation for NDB and PFN were informed by the NMA ORs applied to the baseline PBO risk. The all evidence scenario (See Section 4.10) was selected for the base-case, since there was no variation in the results of the NMA. Other overall discontinuation NMA scenarios were explored within the sensitivity analyses (see page 132). The fixed effects model was used as it performed better than the random effects model. The OR values used in this model are presented in Table 122 below. Details of this analysis and results are presented elsewhere (See Section 4.10 and Appendix B SMC sub)

Table 122: OR values for overall discontinuation - NMA results

Comparison	**OR median value**	**95% Confidence interval**
NDB vs. PBO	1.42	1.08 – 1.87
PFN vs. PBO	1.34	1.34 – 1.73

Note that the economic model assumed no discontinuation from BSC. The PBO baseline discontinuation risk was used only to calculate the risk of discontinuation for the active treatments.

5.3.2 Calculation of transition probabilities

Overall survival

Parametric model extrapolation using pooled phase II and phase III clinical data was used to derive the probability of death for the BSC arm.

Loglogistic model

The *survival function* of the loglogistic model was used:

$S(t)_{PBO}$=1/(1+(exp(-cons)*TimeInDays)^(1/gamma)),

where TimeInDays takes the value of the number of days, and cons and gamma are from the STATA output in Table 123:

Table 123: Loglogistic model for OS

Variable	Coef.	Std. Err.	z	P>z	95% Conf. Interval	
cons	7.164	0.187	38.4	<0.001	6.799	7.530
ln_gam	-0.633	0.139	-4.56	<0.001	-0.905	-0.361
gamma	0.531	0.074			0.405	0.697

The rate of death for the PBO arm over the period [t,t+1] was calculated from the cumulative hazard for periods [0,t] and [0,t+1]:

Rate=-(ln(S(t+1))-ln(S(t)))

The rate was converted to a probability: P=1-exp(-r*t).

Weibull model

The *shape* and *scale* of the Weibull parametric model were obtained for survival as such:

Shape=p from STATA output

*Scale=exp(-p*cons)* from STATA output:

Table 124: Weibull model for OS

Variable	Coef.	Std. Err.	z	P>z	95% Conf. Interval	
cons	7.217	0.192	37.6	<0.001	6.841	7.593
ln_p	0.610	0.140	4.36	<0.001	0.336	0.885
p	1.841	0.258			1.399	2.422
1/p	0.543	0.076			0.413	0.715

The probability of death was calculated:

Probability=1-exp(scale(TimeInDays-DaysInCycle)^shape-scale*TimeInDays^shape)*

Note that *DaysInCycle*=91.35 days, *TimeInDays* takes the value of the number of days given the cycle we are in.

Gompertz model

The survival function of the Gompertz model used was:

*S(t)PBO=exp(-exp(cons)*1/gamma*(exp(gamma*TimeInDays)-1)),*

where TimeInDays takes the value of the number of days, and cons and gamma are outputs of the survival analysis of clinical data in STATA.

The rate of death for the PBO arm over the period *[t,t+1]* was calculated from the cumulative hazard for periods *[0,t]* and *[0,t+1]*:

Rate=-ln(S(t+1))-ln(S(t)))

The rate was converted to a probability: *P=1-exp(-r*t)*

Probabilities of death for the active treatments

In order to estimate probabilities of death for the active treatments (NDB and PFN), the baseline risk (PBO probability of death) was synthesised with the results of the NMA (see section 4.10) and data on file (Boehringer Ingelheim, 2014[92]). For each comparator, the baseline mortality risk was multiplied by the corresponding OR value from the NMA, assuming a constant relationship over time. For instance, the mortality risk for NDB treatment is calculated as follows in each cycle:

$$NDB_P = BSC_P \times \frac{OR_{NDB/PBO}}{1 - BSC_P \times \left(1 - OR_{NDB/PBO}\right)}$$

Where: NDB_P is the probability for NDB treatment, BSC_P is the probability for BSC, and $OR_{NDB/PBO}$ is the OR of NDB vs. PBO.

Acute exacerbation

An exponential model on phase III PBO clinical data (constant risk) was used:

Recall the following formulas from the OS Weibull function:

Shape=p

*Scale=exp(-p*cons)*

Probability (exacerbation)=1-exp(scale(TimeInDays-DaysInCycle)^shape-scale*TimeInDays^shape)*

where *DaysInCycle*=91.35 days, *TimeInDays* takes the value of the number of days.

An exponential model is a special case of the Weibull model, in which the shape=p=1. Therefore:

*Probability=1-exp(-exp(-cons)*DaysInCycle)*

where cons is the output of the survival analysis of clinical data in STATA (see Table 125).

Table 125: STATA survival analysis output – time to acute exacerbation

Investigator reported exacerbation						
Variable	**Coef.**	**Std. Err.**	**z**	**P>z**	**95% Conf. Interval**	
cons	8.430	0.177	47.69	<0.001	8.083	8.776
Adjudication committee reported exacerbation						
Variable	**Coef.**	**Std. Err.**	**z**	**P>z**	**95% Conf. Interval**	
cons	8.724	0.204	42.74	<0.001	8.324	9.124

Similarly to the OS exercise, the risks of exacerbation for NDB and PFN were informed by the NMA ORs applied to the baseline BSC risk, assuming a constant relationship over time (see section 4.10 and data on file (Boehringer Ingelheim, 2014[92])).

<u>Loss of lung function</u>

Phase III clinical trial patient-level data on lung function decline, defined as a 10-point drop in FVC%pred, were analysed for this model transition for the BSC arm (Richeldi *et al.*, 2014[12], Ingelheim, 2014[97]). A logistic model was used to capture several predictors of lung function decline:

Current FVC%pred state (FVC%pred value of patient before progression)

Exacerbation (i.e. progression before and after an exacerbation event)

The **base-case logistic model** used is the model without exacerbations:

P=cons + b fvcpp_prev~s*, where *cons* is a constant value, *b* is the value of the FVC%Pred at the start of the interval, *fvcpp_prev~s* is the current FVC%Pred covariate:

Table 126: Base-case logistic model for loss of lung function (without exacerbations)

Variable	mean	SE	p-value	95% CI	
fvcpp_prev~s	0.016	0.006	0.004	0.005	0.028
_cons	-4.180	0.552	<0.001	-5.262	-3.097

A sensitivity analysis scenario used the logistic value that takes into account the exacerbation coefficient as well. Note that since the coefficient is not statistically significant, it was not used as base-case.

P=cons + b fvcpp_prev~s+_lexa_1*, where *cons* is a constant value, *b* is the value of the FVC%Pred at the start of the interval, *fvcpp_prev~s* is the current FVC%Pred covariate:

Table 127: Logistic model for loss of lung function (with exacerbations considered)

Variable	mean	SE	p-value	95% CI	
fvcpp_prev~s	0.016	0.005	0.003	0.006	0.026
_lexa_1	0.814	1.066	0.445	-1.276	2.904
_cons	-4.068	0.456	<0.001	-4.962	-3.174

The risks of loss of lung function for NDB and PFN were informed by the NMA ORs applied to the baseline BSC risk, assuming a constant relationship over time (see section 4.10 and data on file (Boehringer Ingelheim, 2014[92])).

<u>Overall discontinuation</u>

Similarly to the exacerbation probabilities, an exponential model on phase II and phase III PBO clinical data (constant risk) was used.

The probability of discontinuation given an exponential model is:

*Probability =1-exp(-exp(-cons)*DaysInCycle)*

Where *DaysInCycle*=91.35 days, and *cons* is the output of the survival analysis of clinical data in STATA:

Table 128: STATA output for risk of discontinuation

Variable	Coef.	Std. Err.	z	P>z	95% Conf. Interval	
cons	7.380	0.098		<0.001	7.190	7.570

The probability of discontinuation for NDB and PFN were informed by the NMA ORs applied to the baseline PBO risk (see section 4.10 and data on file (Boehringer Ingelheim, 2014[92])).

Note that transition matrices can be found in Appendix C.

5.3.3 Change in transition probabilities over time

The probability of death (overall survival) was assumed to be time-dependent and it was based on parametric modelling (Section 5.3.2).

The probability of exacerbation (investigator reported or adjudicated) was implemented using an exponential model, which provides a constant hazard over time:

- Investigator reported exacerbations: The goodness of fit for five parametric models for time-to-first-acute-exacerbation (investigator reported) was assessed. The AIC differences between the models were very small (1-2 AIC points); therefore to ensure model parsimony we selected the exponential model, assuming a constant hazard over time. The alternatives would have required increased complexities of the decision analytic model in order to keep track of time-in-state before extrapolation.
- Adjudication committee exacerbations: A similar exercise was performed on the exacerbation data selected based on adjudication committee considerations. Goodness of fit for five parametric models for time-to-first-acute-exacerbation was assessed. Since the AIC differences between the models were very small (1-2 AIC points), the exponential model was used for consistency and comparability with the previous type of exacerbation (investigator reported).

The probability of loss of lung function was derived from a logistic regression model, assuming a constant probability. The goal of this analysis was to estimate the probability of experiencing a drop of 10% in FVC%Pred during any 3-month cycle and control for any other parameter that could influence progression. This is a recurrent event and its probability could not be analysed using time-to-event techniques like in the other outcomes. A logistic regression was selected as this allows the analysis of recurrent events and the incorporation of additional covariates that may influence the probability.

The trial data were separated into four 3-month time intervals (cycle length), and assessed whether a 10-points drop was observed during each interval. The data was analysed by taking into account that each patient contributed up to 4 data points to the analysis (repeated measures technique). The position of the time-interval (i.e. 0-3 vs. 3-6 vs. 6-9 vs. 9-12 months) was checked; it did not significantly impact on the probability of the event. The logistic regression model was built by trying several covariates, before fine tuning it to the ones used (FVC%Pred value at start of interval, treatment) based on significance.

The probability of treatment discontinuation was implemented as an exponential model, which assumes a constant hazard over time. The exponential and Gompertz parametric models provided the best fit to the discontinuation data. To ensure model parsimony the exponential model was used.

5.3.4 Clinical expert assessment of the applicability

Boehringer Ingelheim selected clinical experts to review assumptions within the model on the basis they were widely published and were involved in clinical trials and guidelines and guidance development.

Two clinical experts were approached to validate the assumption within the model at an advisory board held on the 23rd April 2014. The advisory board was facilitated by both members of Boehringer Ingelheim and a health economic consultancy Symmetron and minutes were produced (Boehringer Ingelheim, 2014[91]).

The clinicians were:

- Gisli Jenkins – Clinical Associate Professor and Reader in Pulmonary Biology, Faculty of Medicine and Health Sciences at the University of Nottingham
- Toby Maher – Consultant Respiratory Physician on the Interstitial Lung Disease Unit at the Royal Brompton Hospital, and honorary Senior Lecturer at Imperial College, London and honorary Senior Research Associate at University College London

Clinicians were aware that the advisory board was to discuss aspects of the nintedanib for IPF HTA submission, and they were aware of the nintedanib clinical trials. During the advisory board, the clinical assumptions of the model were checked and discussed between the clinicians. The details and minutes of the advisory board were recorded (Boehringer Ingelheim, 2014[91]).The clinicians were in agreement of the model structure and assumptions made.

5.4 *Measurement and valuation of health effects*

5.4.1 Health-related quality-of-life data from clinical trials

Health state utility values

The economic analysis used patient reported EQ-5D data collected in the INPULSIS clinical trials to determine health state utility values (Richeldi, 2014[4]). Baseline utility was represented by values associated with lung function; an acute exacerbation was included as a decrement in the baseline utility. Further to that, the model also assumed a utility decrement (disutility) due to treatment-related adverse events.

Utility based on lung function (baseline utility)

The EQ-5D values were compiled from the phase III INPULSIS trial by FVC%pred group (Richeldi, 2014[4]). The analysis controlled for exacerbation events, i.e. data

before exacerbations were used. Several assumptions were made to incorporate the results of the data analysis to the economic model:

For health states above FVC%pred of 90%: assume the same utility value as for FVC90%pred. This assumption was made because the values for FVC%pred ≥90% were all around an EQ-5D value of 0.84.

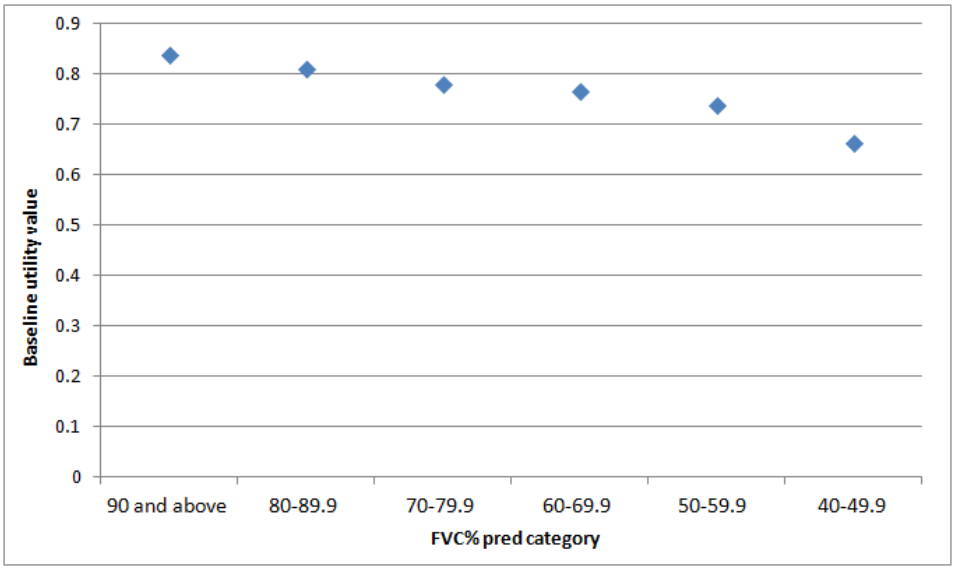
For FVC%pred values < 40%: assume the utility is 0 (dead).

The EQ-5D values for each FVC%pred category are presented in Table 129 and Figure 49 below:

Table 129: EQ-5D utility values by FVC%pred group

FVC%pred	Mean EQ-5D utility	SD	Number
90 and above	0.8380	0.1782	458
80-89.9	0.8105	0.2051	684
70-79.9	0.7800	0.2244	788
60-69.9	0.7657	0.2380	809
50-59.9	0.7387	0.2317	490
40-49.9	0.6634	0.2552	98

Figure 49: Baseline utility values



Acute exacerbation-related decrement in utility

The utility decrements related to acute exacerbations were compiled from the INPULSIS phase III trial for both investigator reported and adjudicated exacerbations (Ingelheim, 2014[97]). The utility change was assumed to be driven by the exacerbation event. The EQ-5D change was analysed over time assuming that patients experience an acute phase (in the 1st month following an exacerbation), and a post-acute phase (in the following 2+ months). Note that the data were too limited to permit the analysis of 2, 3, or 4 months decrements.

The analysis presented a logical trend in disutility experienced following an acute exacerbation:

- There was a severe change in the 1st month after the exacerbation, followed by a recovery and smaller and longer-lasting disability in the subsequent months
- Adjudicated exacerbations had a more severe impact than the investigator-reported ones.

The results of the exacerbation-related disutility analysis are presented in Table 130. These are assumed to be conservative estimates, as it is likely that the worst patients were missing not at random (as they were unable or unwilling to attend the next study visit).

Table 130: Exacerbation-related disutility

Outcome	Time followed by exacerbation	Mean	SE	LLCI	HLCI
Investigator exacerbations disutility	1st month	-0.14	0.047	-0.231	-0.049
	subsequent months	-0.078	0.032	-0.142	-0.015
Adjudication committee exacerbations disutility	1st month	-0.274	0.059	-0.39	-0.157
	subsequent months	-0.033	0.053	-0.137	0.07

SE, Standard error; LLCI, Lower level confidence interval; HLCI, Higher level confidence interval

5.4.2 Mapping

As EQ-5D was measured directly, no mapping was required.

5.4.3 Systematic search for relevant HRQL

A systematic literature review was conducted to identify relevant Health State Utility Values (HSUVs) for the *de novo* cost-effectiveness model for nintedanib. The systematic review for HSUVs was informed by a scoping stage as recommended by the Decision Support Unit (DSU) technical support document on the identification, review and synthesis of HSUV from the literature (Papaioannou *et al.*, 2011[98]).

Identification of studies

Database search

The following computerised bibliographic databases advised by the CRD (Dissemination, 2009[24]) were searched on 27/08/2014:

- MEDLINE;
- MEDLINE In-Process;
- Embase.

MEDLINE and Embase were searched via the Ovid® search platform.

Search strategy

The scoping stage of the review revealed limited evidence on publications reporting EQ-5D, thus the search strategy was not limited to studies reporting the measure. As no existing search filter for HRQL has been validated (Paisley *et al.*, 2005[99]), the search strategy was designed to incorporate disease and patient population terms consistent with the clinical searches in section 4.1 and terms for HRQL, generic and disease-specific measures of HRQL. The search was not limited by country, language, study design or date but studies published in languages other than English were not reviewed in full. The search strategies by databases are provided in Appendix A.

Conference proceedings search

In August 2014, conference proceedings were searched from 2012 to 2014 to identify publications reporting HSUVs for adults and elderly with IPF for the following conference sites:

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (www.ispor.org)

International Health Economics Association (iHEA) (www.healtheconomics.com).

Study selection

The study selection for HRQL studies was performed as a two-stage process.

The first stage consisted of review of abstracts and full texts of identified publications as detailed in section link to description in clinical section. However, different eligibility criteria were developed to capture studies reporting health state utility values for adults with IPF (see Table 131). As recommended by the DSU 9 guidance document, the review was not limited by the conventional PICOS criteria but instead considered the health states of interest for the *de novo* cost-effectiveness model for nintedanib, HSUVs and broad study design criteria (Papaioannou *et al.*, 2011[98]). As recommended by NICE and SMC, the review aimed to capture studies using EQ-5D. Other measures such as the SF-36 and disease-specific measures of HRQL, as well as publications reporting mapping of other generic or disease-specific measures to EQ-5D in IPF were also sought in case no EQ-5D data were available.

The second stage assessed the appropriateness of the selected studies for their use in the nintedanib *de novo* cost-effectiveness analysis for a NICE submission. Studies reporting EQ-5D scores were sought to be relevant. Measures other than EQ-5D were only considered relevant in case a publication reporting a mapping technique of that measure in question to EQ-5D was identified in the search.

Table 131: Eligibility criteria for the HRQL search

Criteria	Inclusion criteria	Exclusion criteria
Health states/ adverse events	Adult patients with IPF (including subgroups of patients e.g. subgroups based on 6MWT SpO2) with the following: *Health states*: • IPF, FVC >110% predicted; • IPF, FVC 100% predicted; • IPF, FVC 90% predicted; • IPF, FVC 80% predicted; • IPF, FVC 70% predicted; • IPF, FVC 60% predicted; • IPF, FVC 50% predicted; • IPF, FVC 40% predicted. *Adverse events:* • IPF exacerbations; • Gastro-intestinal event; • Skin disorder; • Dizziness; • Anorexia.	• Paediatric and younger than 18; • Healthy participants; • Non-human. • No health states defined by FVC% predicted • No adverse events of interest
Outcomes (HSUV)	*Health states:* • Preference measurements *Adverse events:* • Utility decrements (disutility) measurements (scores)	• No preference measurements
Study design	• RCTs that record HRQL data • Observational studies • HRQL studies • Cost-utility analyses	• Reviews and systematic literature reviews*; • Case reports; • Critical appraisals; • Update or commentaries on data published elsewhere; • Notes, letters or editorials.

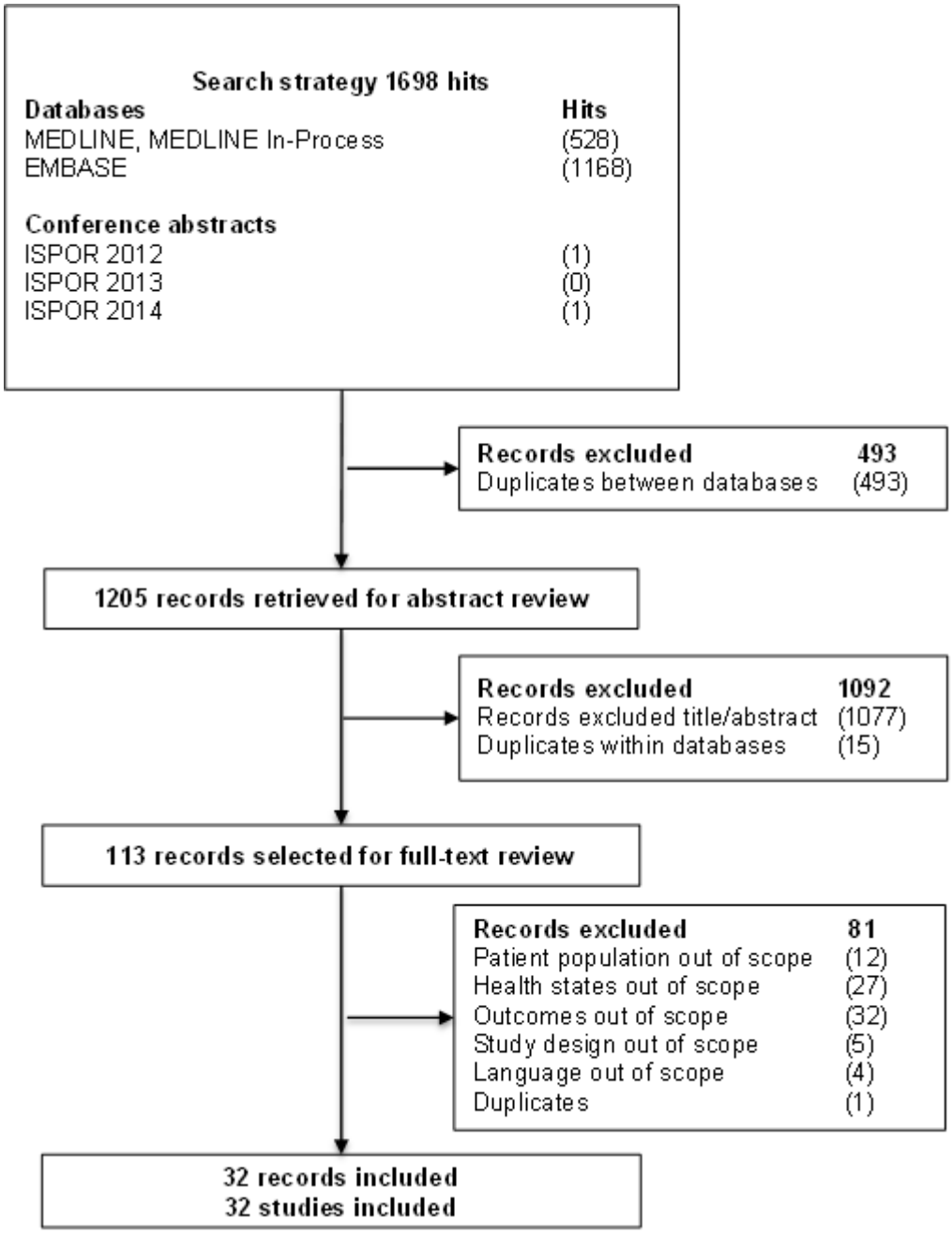
*Systematic literature reviews will be excluded from the main database search but their reference lists will be searched for relevant studies not captured by the main database search and to be included as hand search in the review.

Data Extraction

The process of data extraction was the same as is detailed in section link to relevant section in clinical section. Data on the population characteristics (*e.g.,* age, gender, ethnicity, smoking status, years since IPF diagnosis) and HRQL outcomes (*e.g.,* HSUV defined by FVC% predicted, adverse events disutility scores) were extracted. Details on data extracted are provided in Appendix A.

The search retrieved 1,698 records from databases and conferences. Following duplicate removal and abstract review, a total of 113 records were selected for full-text appraisal against the pre-defined eligibility criteria described above. 32 records were selected for data extraction (see Appendix A and C). Figure 50 illustrates the study selection process.

Figure 50: Study flow diagram for HRQL study selection (PRISMA diagram)



5.4.4 Details of studies in which HRQL was measured

A summary of the 32 selected studies is presented in Table 132.

Only two (King *et al.*, 2011[100], Zisman *et al.*, 2010[101]) studies reported EQ-5D scores for patients with IPF. Both studies were randomised, double-blind, placebo-controlled trials, with one investigating the safety and efficacy of bosentan (King *et al.*, 2011[100]) and the other of sildenafil (Zisman *et al.*, 2010[101]). The studies were consistent in their definition of IPF, both using the ATS/ERS statement. The sample sizes of the trials

were considered to be sufficient in terms of size, with 616 (King *et al.*, 2011[100]) and 180 (Zisman *et al.*, 2010[101]) patients, respectively. As the *de novo* model health states were defined based on FVC and FVC values were only reported at baseline, only the baseline HRQL estimates could be used for the analysis. Those studies were considered as appropriate for the analysis.

23 studies administered other generic (e.g. SF-36, WHOQOL-BREF) and 20 administered disease-specific (e.g. SGRQ, ATAQ-IPF) measures. As no mapping method for the identified generic or disease-specific measures to EQ-5D scores in IPF patients was identified, all studies using those measures of HRQL were considered not appropriate for the analysis.

No publications, reporting the HRQL impact of the adverse events of interest (IPF exacerbations, gastro-intestinal events, skin disorder, dizziness, anorexia) in IPF, were identified.

Table 132: Summary of retrieved records pertaining to the HRQL search

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Mishra et al. 2011 (Mishra *et al.*, 2011[102]) India	Patients with IPF and without ILD† receiving doxycycline orally continuously (100 mg and once or twice daily based on body weight of less then or more than 50 kg) • Population is in accordance with ATS, ERS standards • Mean age 70.67±11.25	Patients were selected and included from the referral OPD	6	SGRQ	100%	Mean % of predicted FVC : 61.38% Mean % of predicted FVC after treatment (24 weeks): 66.67%	Baseline: 50.9 24 weeks: 18.4	Baseline: 8.38 24 weeks: 6.39	Not appropriate: small sample, no mapping methodology of SGRQ to EQ-5D in IPF patients available
Swigris et al. 2011 (Swigris *et al.*, 2011[14]) USA	Pulmonary rehabilitation (standard 6-week programme) for IPF and COPD† patients. • Population is in accordance with ATS standards • 85.7% male • Mean age 71.5±7.4 • 81.3% ex-smokers	In 2008 and 2009 we recruited from the Denver metropolitan area a convenience cohort of 21 patients with IPF	21	SF-36	67%	NR; single time point measure	Baseline: PF: 31.9 RP: 36.4 BP: 45.0 GH: 38.3 V: 47.2 SF: 45.1 RE: 45.7 MH: 51.8 PCM: 33.1 MCS: 53.6	Baseline: PF: 2.4 RP: 2.3 BP: 2.2 GH: 1.7 V: 2.2 SF: 2.0 RE: 2.6 MH: 2.0 PCS: 1.9 MCS: 2.2	Not appropriate: small sample, no mapping methodology of SF-36 to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
King et al. 2008 (King *et al.*, 2008[103]) Multiple countries	IPF patients who received Bosentan (oral bosentan 62.5 mg twice daily for 4 weeks, uptitrated to bosentan 125 mg twice daily thereafter (target dose)) vs. placebo. • Population is in accordance with ATS standards/ ERS consensus guidelines • Bosentan patients: o 69% male o Mean age 65.3±8.4 o 2.8% current smokers • Placebo patients: o 76% Male o Mean age 65.1±9.1 o 1.2% current smokers		71	SF-36, SGRQ	69%	bosentan group: 'Mean percent of predicted FVC : 65.9% (10.50); Mean FVC percent predicted at 12 months 59.5%	Baseline: SF-36: 45.7 SGRQ: 44.3 After 12 Months: SF-36: NR SGRQ: 46.1	Baseline: SF-36: 18.1 SGRQ: 18 After 12 months: SF-36: NR SGRQ: 19.9	Not appropriate: no mapping methodology of SGRQ or SF-36 to EQ-5D in IPF patients available
			83			Placebo: 'Mean percent of predicted FVC : 69.5% (12.60) Mean FVC percent predicted at 12 months 56.9%	Baseline: SF-36: 45.2 SGRQ: 45.9 After 12 months: SF-36: NR SGRQ: 51.1	Baseline: SF-36: 19 SGRQ: 16.9 After 12 months: SF-36: NR SGRQ: 23.7	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
King et al. 2011 (King *et al.*, 2011[100]) Multiple countries	Patients with a proven diagnosis of IPF made within the last 3 years before enrolment treated with bosentan (62.5 mg twice daily, up-titrated after 4 weeks to a target dose of 125 mg twice daily (or remaining at 62.5 mg twice daily if body weight<40 kg) vs. placebo group : • Population is in accordance with ATS standards/ ERS statement • Bosentan patients: o 72.7% Male o Mean age 63.8±8.4 o 38.1% non-smokers • Placebo patients: o 63.6% Male o Mean age 63.2±9.1 o 32.1% non-	Adult patients with IPF, less than 3 years' duration, and with diagnosis confirmed by surgical lung biopsy	407	SF-36, EQ-5D, EQ-VAS	81.8%	Bosentan group: Mean percent predicted FVC : 74.9% (14.8); single time point measure	SF-36: PF: 61.1; RP: 63.1; PI: 69.9; GHP: 52.1; V: 55.5; SF: 77.6; RF: 79.3 MHI: 73.6 EQ-5D: 0.758; EQ-VAS: 70.4	SF-36: PF: 25.4; RP: 30.0; PI: 26.5; GHP: 21.5; V: 21.9; SF: 24.3; RF: 26.2; MHI: 20.1; EQ-5D: 0.185; EQ-VAS: 18.7	Appropriate
		Matching placebo patients	209		90%	Placebo group; single time point measure	SF-36: PF: 58.2; RP: 59.2; PI: 68.4; GHP: 48.7; V: 52.3; SF: 72.5; RF: 74.7; MIX: 71.3; EQ-5D: 0.718;	SF-36: PF: 28.9; RP: 32.4; PI: 31.1; GHP: 24.1; V: 24.4; SF: 30.5; RF: 31.6; MIX: 22.9; EQ-5D: 0.242;	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	smokers						EQ-VAS: 69.5	EQ-VAS: 19.4	
Lubin et al. 2014 (Lubin *et al.*, 2014[104]) USA	Patients with IPF and CHP† • All patients with IPF were diagnosed according to consensus criteria; • 75% Male • Mean age 70±8.0 • 30% non-smokers	Patients were identified from an ongoing longitudinal cohort of patients with ILD seen at the University of California, San Francisco from January 2010 to August 2012	102	SF-36	100%	FVC % predicted 70 (18); One time point measure	SF-36: PCS: 39.2; MCS: 49.1	SF-36: PCS: 10.2; MCS: 14.2	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available
Lee et al. 2010 (Lee *et al.*, 2010[105]) Ireland	IPF patients: • Diagnosed by ATS criteria and none receiving oxygen • 70% Male • Mean age 72±4.1 • 30% non-smokers	IPF patients by random selection from our IPF database	10	SF-36	100%	FVC percent predicted 85.3%; One time point measure	SF-36: 62.1	SF-36: 18.16	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available
Zimmermann et al. 2007 (Zimmermann *et al.*, 2007[106])	IPF patients: • IPF diagnosis was based on clinical, functional and	Patients with documented IPF, identified through clinical records, who attended the Interstitial Lung	20	SF-36, SGRQ	91%	FVC% predicted 70.4 (9.4), range (40-113)	SF-36: PF: 46; BP: 60.6;	SF-36: PF: 18.3 BP: 31.9	Not appropriate: no mapping methodology of SF-36 or SGRQ to EQ-5D in IPF

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Brazil	radiological findings; • 60% Male; • Mean age 61.4±10.5.	Diseases Group as outpatients at Hospital das Clínicas, University of São Paulo (Brazil) over a period of 6 months					PR: 57.5; GH: 53.7; V: 49.2; SF: 56.9; EA: 46.6; MH: 66.8; Total score: 55.7 SGRQ: Symptom: 46.4 Activity: 62.4 Impact: 43.6 Total score: 48.4	PR: 39.8 GH: 24.1 V: 24.3 SF: 32.2 EA: 39.5 MH: 17 Total score: 28.4 SGRQ: Symptom: 20.3 Activity: 19 Impact: 20.9 Total score: 17.9	patients available
Zisman et al. 2010 (Zisman *et al.*, 2010[101])	Advanced IPF patients treated with Sildenafil vs. placebo: • Population is in accordance with ATS standards/ ERS statement • Sildenafil patients: o 84.0% Male o Mean age	IPF patients in an advanced stage, which was defined as a diffusing capacity for carbon monoxide of less than 35% of the predicted value	89	SF-36, SGRQ, EQ-5D, EQ-VAS	100%	Mean percent of predicted FVC : 54.89% (14.00)	<u>Baseline:</u> SGRQ: Total Score 54.55; Symptoms Score 58.20; Activity Score 71.20; Impacts Score 43.20;	<u>Baseline:</u> SGRQ: Total Score 16.46; Symptoms Score 17.75; Activity Score 17.50; Impacts Score 19.26;	Appropriate

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	69.8±8.7 o 23.6% non-smokers • Placebo patients: o 80.0% Male o Mean age 68.2±9.3 o 24.2% non-smokers	Matching placebo patients	91		100%	Mean percent of predicted FVC : 58.73% (14.12)	SGRQ Total 51.72; Symptoms Score 53.99; Activity Score 68.02; Impacts Score 39.77; SF-36: Aggregate Physical Score: 33.17 Aggregate Mental Score: 49.53; ИЗЖ 50ю55ж ПРЖ 36ю99ж ЬРЖ 51ю22ж ЗАЖ 29ю20ж КУЖ 45ю13ж RP: 34.67; SF: 42.33; V: 44.08; EQ-5D: 0.71; EQ-5D VAS: 66.49	SGRQ Total 51.72; Symptoms Score 53.99; Activity Score 68.02; Impacts Score 39.77; SF-36: Aggregate Physical Score: 9.19; Aggregate Mental Score: 9.76; BP: 10.98; GH: 9.64; MH: 9.07; PF: 8.49; RE: 12.14; RP: 11.39; SF: 10.88; V: 9.28; EQ-5D: 0.24 EQ-5D VAS: 17.45	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Mermigkis et al. 2013 (Mermigkis *et al.*, 2013[107]) Greece	Newly diagnosed IPF and moderate to severe receiving continuous positive airway pressure (CPAP) treatment; • Population is in accordance with ATS/ERS/ JRS/ALAT statement • 83.3% Male • Mean age 67.1±7.2	Patients who were referred and evaluated at the outpatient clinic and Pulmonary Department of the University of Crete during the period March 2010 to September 2011.	12	SF-36	100%	Mean percent of predicted FVC : '77.1 (18.1)	Baseline: 63.2; 1st month: 68.9 3rd month: 72.1 6th month: 74.4	Baseline: 13.9; 1st month: 13.5 3rd month: 12.9 6th month: 11.3	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available
Peng et al. 2008 (Peng *et al.*, 2008[108]) China	IPF patients • The diagnosis was based on established clinical and histological criteria; • 79% Male • Mean age 64±8 • 35.3% non-smokers	IPF patients who were referred to the outpatient clinic at the Institute of Respiratory Diseases of China Medical University between January 2002 and May 2007	68	Chinese version of SGRQ	100%	FVC, % predicted: 66 (18)	SGRQ: Symptom: 65 Activity: 56 Impact: 49 Total score: 54	SGRQ: Symptom: 16 Activity: 15 Impact: 19 Total score: 15	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
Swigris et al. 2010b (Swigris *et al.*, 2010[109])	Moderately severe IPF patients: • The diagnosis	Details of recruitment reported somewhere else (Swigris,	95	ATAQ-IPF	100%	Mean percent of predicted FVC : 65%	ATAQ-IPF scores: Cough: 16	ATAQ-IPF scores: Cough: 7	Not appropriate: no mapping methodology of ATAQ-IPF to EQ-5D in IPF patients

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
USA	was based on ATS criteria; • 82% male • Mean age 69.3±7.6 • 36.0% non-smokers	2005[110]).				(17)	Dyspnoea: 19 Forethought: 14 Sleep: 16 Mortality: 17 Exhaustion: 15 Emotional: Well-Being: 20 Social Participation: 15 Finances: 17 Independence: 14 Sexual Health: 15 Relationships: 17	Dyspnoea: 6 Forethought: 6 Sleep: 5 Mortality: 5 Exhaustion: 5 Emotional Well-Being: 6 Social Participation: 5 Finances: 7 Independence: 5 Sexual Health: 6 Relationships: 4	available
Kozu et al. 2011b (Kozu *et al.*, 2011b[111]) Japan	IPF and COPD† patient' groups who received 8-week out-patient pulmonary rehabilitation program vs. placebo group • The diagnosis was based on ATS criteria; • 82% male • Mean age 67.5±7.8	NR	45	SF-36	67%	FVC % predicted: 68.6 (16)	<u>Baseline:</u> PF: 35.7 RP: 33.9 BP: 62.4 GH: 34.7 V: 38.3 SF: 48.3 RE: 36.5 MH: 50.0	<u>Baseline:</u> PD: 18.7 RP: 21.7 BP: 30.3 GH: 19.9 V: 21.3 S: 23.7 RE: 30.1 MH: 18.7	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
							8 weeks: PF: 40.6 RP: 35.9 BP: 63.4 GH: 36.9 V: 43.9 SF: 50.3 RE: 38.7 MH: 52.6 6 months: PF: 37.8 RP: 30.4 BP: 62.5 GH: 34.4 V: 42.18 SF: 45.88** RE: 35.88 MH: 47.58	8 weeks: PF: 22.6 RP: 20.7 BP: 28.1 GH: 21.1 V: 21.0 SF: 25.3 RE: 31.3 MH: 20.5 6 months: PF: 23.0 RP: 23.7 BP: 30.3** GH: 21.5** V: 23.6 SF: 26.9 RE: 29.8 MH: 21.8	
Baddini Martinez et al. 2002 (Baddini Martinez *et al.*,	IPF patients: • IPF confirmed by open lung	Group of patients with IPF under treatment at Sao Paulo University	30	SF-36	NR	Mean percent of predicted FVC :	PF: 40.83 PR: 44.17	PF: 4.41 PR: 8.86	Not appropriate: no mapping methodology of SF-36 to EQ-5D in

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
2002[112]) Brazil	biopsy or was based on typical clinical and high-resolution CT findings only ; • 60% male • Mean age 58.6±2.0	Hospital outpatient clinic were sequentially enrolled in the study				61.9% (3.24)	PI: 78.57 GHP: 53.07 V: 50.50 SF: 60.83 ER: 64.44 MHI: 59.47	PI: 4.4 GHP: 4.33 V: 5.10 SF: 7.26 ER: 8.74 MHI: 4.15	IPF patients available
Kozu .et al. 2011c (Kozu *et al.*, 2011c[113]), Japan	Stable IPF patient enrolled in 8 weeks pulmonary rehabilitation programme and grouped by dyspnoea scale • The diagnosis was based on ATS criteria; • Grade 2 o 81.25% male o Mean age 66.4±7.7 • Grade 3 o 76.5% male o Mean age 67.8±7.4 • Grade 4 o 64.7% male o Mean age 68.1±7.6 • Grade 5	Consecutive patients with a diagnosis of IPF who were referred to the pulmonary rehabilitation programme (PRP) at Nagasaki University Hospital	Grade 2: 16	SF-36	100%	Grade 2 FVC% pred: 83 (11)	GH: 50.9 PF: 55.3 PR: 55.9 BP: 66.5 V: 54.7 SF: 62.5 ER: 66.7 MH: 61.6	GH: 11.0 PF: 7.2 PR: 15.9 BP: 25.1 V: 11.7 SF: 18.8 ER: 15.2 MH: 14.3	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available
			Grade 3: 17			Grade 3 FVC% pred: 67 (13)	GH: 35.8 PF: 34.1 PR: 22.4 BP: 57.2 V: 37.9 SF: 42.6 ER: 47.1 MH: 42.9	GH: 18.9 PF: 18.4 PR: 17.3 BP: 29 V: 21.5 S: 27.6 E: 28.2 M: 20.8	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	o 60% male o Mean age 68.7±7.5		Grade 4: 17			Grade 4 FVC,% pred: 60 (16)	GH: 24.1 PF: 20.3 PR: 23.2 BP: 65.6 V: 26.5 SF: 36.0 ER: 30.9 MH: 41.8	GH: 16.8 PF: 7 PR: 13.4 BP: 29.1 V: 18 SF: 15.2 ER: 21.4 MH: 17.2	
			Grade 5: 15			Grade 5 FVC,% pred: 51 (11)	GH: 19.1 PF: 16 PR: 19.6 BP: 65.6 V: 19.6 SF: 30 ER: 19.4 MH: 35	PH: 10.7 PF: 9.1 PR: 10.3 BP: 28.4 V: 15.3 SF: 14.8 ER: 15.3 MH: 12	
Ozalevli et al. 2010 (Ozalevli *et al.*, 2010[114]) Turkey	IPF patient enrolled in Home-based pulmonary rehabilitation programme • The diagnosis was based on ATS, ERS criteria; • 67% male • Mean age 62.8±8.5	Patients who performed home-based pulmonary rehabilitation on regular basis.	17	SF-36	88%	Baseline: Mean percent of predicted FVC : 71.6 (8.2) 12 weeks: Mean percent of predicted	Baseline: GH: 57.0 PF: 56.0 PR: 25.0 BP: 67.3 V: 52.0 SF: 75.8 ER: 29.0 MH: 49.9	Baseline: GH: 4.6 PF: 5.7 PR: 1.7 BP: 2.6 V: 4.9 SF: 2.7 ER: 1.3 MH: 6.7	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
						FVC : 73.9 (5.3)	12 weeks: GH: 74.0 PF: 58.7 PR: 68.3 BP: 72.0 V: 55.0 SF: 89.1 ER: 65.0 MH: 56.8	12 weeks: GH: 4.7 PF: 7.3 PR: 1.6 BP: 2.2 V: 4.2 SF: 1.8 ER: 1.4 MH: 5.4	
Mishima et al. 2013 (Mishima *et al.*, 2013[115]), Japan	Advanced IPF patient enrolled in pulmonary rehabilitation programme • The diagnosis was based on ATS, ERS criteria; • 68% male • Mean age 69±8	Patients with advanced IPF who completed four weeks of in-patient PR at Tosei General Hospital from January 2001 to December 2012 were recruited.	28	SGRQ	100%	FVC, % predicted, 53% (17)	69.8	11.5	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Nishiyama et al. 2008 (Nishiyama *et al.*, 2008[116]), Japan	IPF patients: • The diagnosis of IPF was made in accordance with the ATS/ERS statement • Rehabilitation group o 92% male o Mean age 68.1±8.9 • Control group o 60% male o Mean age 64.5±8 9.1	Subjects were consecutive patients referred to the outpatient clinic between 2000 and 2004	Rehabilitation group: 13	SGRQ	87%	FVC,% pred 68.7 (19.5)	Symptoms: 56.4 Activity: 64.7 Impacts: 39.7 Total: 50.2	Symptoms: 22.3 Activity: 17.1 Impacts: 17.6 Total: 16.3	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
			Control group: 15		100%	FVC,% pred 66.1 (13.2)	Symptoms 38.0 Activity 50.4 Impacts 29.9 Total 37.8	Symptoms: 25.8 Activity: 26.2 Impacts: 23.7 Total: 22.7	
Richeldi et al. 2011 (Richeldi *et al.*, 2011[1]), Multiple countries	IPF patients: • The diagnosis of IPF was made in accordance with the ATS and ERS criteria and who had received the diagnosis less than 5 years	Eligible patients, who were recruited from 92 sites in 25 countries, had also undergone HRCT less than 1 year before randomisation and had a FVC that was 50% or more of their	BIBF1120 50g x1d: 86	SGRQ (total score)	72%	FVC,% pred: 80.4 (17.8)	43.7	17.5	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
			BIBF1120 50g x2d: 86		79%	FVC,% pred: 79.8 (15.8)	42.5	17.0	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	before screening; • BIBF1120 50g x1d group o 75.6% male o Mean age 65.3±9.4 o 79.1% non-smokers • BIBF1120 50g x2d group o 72.1% male o Mean age 64.9±8.5 o 83.7% non-smokers • BIBF1120 100g x1d group o 75.6% male o Mean age 65.1±8.6 o 83.7% non-smokers • BIBF1120 150g x2d group o 76.5% male o Mean age 65.4±7.8 o 71.8% non-smokers • Placebo group o 74.1% male o Mean age 64.8±8.6 o 76.5% non-	predicted value	BIBF1120 100g x2d: 86		84%	FVC,% pred: 85.5 (19.2)	43.7	16.6	
			BIBF1120 100g x2d: 85		62%	FVC,% pred: 79.1 (18.5)	40.1	18.3	
			Placebo: 85		72%	FVC,% pred: 81.7 (17.6)	41.2	17.9	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	smokers								
Tzanaki set al. 2005 (Tzanakis *et al.*, 2005[117]), Greece	IPF patients vs. control: • The diagnosis confirmed with biopsy and radiographic findings; • IPF group o 84% male o Mean age 66±11 o 76% smokes • Controls group o 80% male o Mean age 63.5±10 o 83.3% smokes	Consecutive outpatients with clinically and histologically consistent IPF were followed up at University Hospital of Alexandroupolis hospital from January 1994 until July 1999	IPF patients: 25	SGRQ, QWB, HAD	100%	FVC, % pred. 68.8 (16)	SGRQ: Total:37.7; Symptoms: 25.3; Activity: 36.2; Impact: 29.6; QWB: 0.59; HAD: Anxiety: 6.3; Depression: 6.2.	SGRQ: Total: 18.9; Symptoms: 25.3; Activity: 21.4; Impact: 21; QWB: 0.29; HAD: Anxiety: 5.6; Depression: 5.7.	Not appropriate: no mapping methodology of SGRQ, QWB or HAD to EQ-5D in IPF patients available
			Healthy control group: 30		100%	FVC, % pred. 102 (14)	SGRQ: Total: 7.8***; Symptoms: 10***; Activity: 6***; Impact: 4.5***;	SGRQ: Total: 2.8; Symptoms: 3.3; Activity: 1.4; Impact: 1.2; QWB: 0.05;	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
							QWB: 0.866***; HAD: Anxiety: 3.0***; Depression: 62.4***.	HAD: Anxiety: 0.6; Depression: 0.4.	
Martinez et al. 2000 (Martinez *et al.*, 2000[58]), Brazil	IPF and matched control group: • The diagnosis of IPF was confirmed by an open lung biopsy or was based only on typical clinical and high resolution CT findings • IPF group o 59% male o Mean age 58.29±1.87 • Controls group o Mean age 58±1.89	A group of patients with IPF who attended the outpatient clinic at Hospital das Cinicas de Ribeirao Preto university hospital were studied	IPF group: 34	SF-36	100%	FVC% predicted: 62.41 (2.96)	PF: 42.79 PR: 44.12 PI: 76.91 GHP: 53.50 V: 50.44 SF: 60.29 ER: 60.78 MHI: 57.53	PF: 4.40 PR: 8.11 PI: 4.16 GHP: 3.90 V: 4.88 SF: 6.69 ER: 8.26 MHI: 4.07	Not appropriate: no mapping methodology of SF-36 to EQ-5D in IPF patients available
			Matched control group: 34		100%		PF: 86.18 PR: 90.44 PI: 76.74 GHP: 76.35 V: 73.53 SF: 82.72 ER: 91.18 MHI: 78.24	PF: 2.47 PR: 4.35 PI: 3.17 GHP: 2.48 V: 2.39 SF: 3.84 ER: 3.54 MHI: 2.35	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Han et al. 2010 (Han *et al.*, 2010[118]), USA	IPF patients analysed by gender: • The diagnosis of IPF was made in accordance with the ATS • 66.5% Male • Average male age 63.3±8.2 • Average female age 62.3±9.9	The study sample was identified through the multi-centre Lung Tissue Research Consortium and all subjects with a final clinical diagnosis of IPF at the time of data query were included in this analysis.	Male: 147;	SF-12, SGRQ	100%	Mean percent of predicted FVC :62.1 (16.8)	SF12 PCS Total: 35.6; PF: 34.1; RP: 37.7; BP: 49.2; GH: 39.4; SF12 MCS Total: 54.4***; V: 45.9; SF: 47.6*; RE: 48.1**; MH: 53.0***; Total SGRQ Score: 44.6; Activity: 57.7*; Impacts: 35.2; Symptoms: 49.7	SF12 PCS: Total: 12.5; PF: 12.2; RP: 13.0; BP: 11.8; GH: 13.3; SF12 MCS: Total: 9.7; V: 11.6; SF: 11.3; RE: 11.4; MH: 8.9; Total SGRQ Score: 23.6; Activity: 30.3; Impacts: 22.7; Symptoms: 25.0	Not appropriate: no mapping methodology of SF-36 or SGRQ to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
			Female:74		100%	Mean percent of predicted FVC: 64.4 (18.7)	SF12 PCS: Total: 37.7; PF: 32.7; RP: 36.8; BP: 49.0; GH: 42.2; SF12 MCS: Total: 48.3; V: 44.6; SF: 43.9; RE: 41.9; MH: 47.8; Total SGRQ Score: 48.2; Activity: 75.3; Impacts: 37.6; Symptoms: 48.9	SF12 PCS: Total: 9.9; PF: 10.8; RP: 12.6; BP: 10.4; GH: 11.7; SF12 MCS: Total: 12.2; V: 11.1; SF: 12.8; RE: 14.1; MH: 10.8; Total SGRQ Score: 20.1; Activity: 24.3; Impacts: 21.9; Symptoms: 20.3	
Verma et al. 2011 (Verma *et al.*, 2011[119]), Canada	IPF patients: • The diagnosis was based on ATS; • 65.7% male • Mean age 59.4±7.1	Patients were recruited through the Lung Transplant and the Interstitial Lung Disease clinics at the	137	Web-based versions of the SF-36 and the SGRQ	100%	FVC, % predicted - 61.7±19.8	SF-36: PF: 25.6; RR: 31; BP: 68; GH: 35.5;	SF-36 95% CI: PF: 21.9-29.4; RR: 26.9–35.1; BP: 63.4–72.7; GH: 31.9–39.0;	Not appropriate: no mapping methodology of SF-36 or SGRQ to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
		University Health Network in Toronto (Ontario) between November 2003 and March 2008.					V: 39.3; SF: 59.2 RE: 74.5 MH: 71.2 PCS : 29.4 MCS: 49.7 SGRQ: Symptoms: 59.8; Activity: 81.6; Impacts: 54.1; Total: 63.4	V: 35.8–42.9; SF: 54.05–64.38 RE: 69.6–79.3 MH: 67.9–74. 6 PCS : 27.9–30.9 MCS: 47.6–51.7 SGRQ: Symptoms: 56.2–63.4; Activity: 78.7–84.4; Impacts: 50.6–57.6; Total: 60.4–66.3	
Nishiyama et al. 2012 (Nishiyama *et al.*, 2012[120]), Japan	IPF patients: • The diagnosis was based on ATS and ERS; • 77% male • Mean age 66.3±8.2 • Current smokers: 16%	Eighty-seven consecutive IPF patients, who had undergone evaluations and completed the SGRQ at diagnosis in Tosei General Hospital, were included in this	87	SGRQ	100%	FVC% predicted: 75.0 (19.2)	Total: 39; Symptoms: 45; Activity: 48; Impacts: 31.6	Total: 20.2; Symptoms: 45; Activity: 48; Impacts: 20.7	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
		study, covering the period from April 2000 to July 2005							
Furukawa et al. 2013 (Furukawa *et al.*, 2013[121]), NR	IPF patients: • 86% male • Mean age 65.7±7.7	Retrospective analysis of patients with IPF who underwent multidimensional evaluation including the HRQL scoring of the SGRQ	118	SGRQ	100%	FVC% predicted: 78.3% (20.1)	Symptoms: 45.9 Activity: 42.6 Impacts: 28.5 Total: 36	Symptoms: 22.5 Activity: 24.4 Impacts: 19.8 Total: 20	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
Antoniou et al. 2006 (Antoniou *et al.*, 2006[122]), Greece	IPF patients: • The diagnosis was based on ATS and ERS; • IFN gamma-1b o 91% male o Mean age 66 (54-85) o 37.5% non-smokers, 50% ex-smokers, 9.4% current smokers o 49.4±24.3 months since diagnosis • Colchicine o 72% male o Mean age 69	Between March 2000 and June 2003, 68 patients were recruited from eight centres and 50 underwent randomisation after informed consent. Patients with mild-to-moderate IPF were eligible for the study if they had histologically proven IPF	32	SGRQ	64%	FVC, % pred: 71.8 (15.0)	NR [results presented as change from baseline rather than utility value]	NR [results presented as change from baseline rather than utility value]	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
			18		56%	FVC, % pred: 70.7 (17.7)	NR [results presented as change from baseline rather than utility value]	NR [results presented as change from baseline rather than utility value]	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	(42-82) o 44.4% non-smokers, 50% ex-smokers, 5.6% current smokers o 42.7±16.8 months since diagnosis								
Jones et al. 2011 (Jones *et al.*, 2011[123]), UK	• IPF patients: o The diagnosis was based on ATS and ERS; o 63% males; o Mean age 71.7±7; o 59% smokers • Control group: o 70% males; o Mean age 65.6±5.3; o 37% smokers	Patients were recruited from the Respiratory Unit, Morriston Hospital, Swansea	27	VAS; LCQ	100%	FVC, % pred: 80.4 (20.9)	VAS: 0*** LCQ: 20.8***	NR	Not appropriate: no mapping methodology of VAS or LCQ to EQ-5D in IPF patients available
			30		100%	FVC, % pred: 120.6 (13.7)	VAS: 38 LCQ: 15.9	NR	
Key et al. 2010 (Key *et al.*, 2010[124]), UK	IPF patients • The diagnosis was based on ATS and ERS; • 73.7% males; • Mean age 70.8±8.6; • 36.8% non-smokers and 57.9% smokers	Patients were recruited from two specialist ILD clinics at University Hospital Aintree (UHA) and University Hospital South Manchester (UHSM)	19	VAS; LCQ	NR	FVC,% pred: 78.5% (24.4)	Cough VAS: Day median: 32mm Night median: 9mm Cough LCQ: Total median:	Cough VAS: Day median: 2-77 Night median: 2-55 Cough LCQ: 6.95-20.88	Not appropriate: no mapping methodology of VAS or LCQ to EQ-5D in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	• 3 years since diagnosis						15.4		
Tzouvelekis et al. 2013 (Tzouvelekis *et al.*, 2013[125]), Greece	IPF patients • The diagnosis was based on ATS and ERS; • 86% males; • Mean age 64.4±7; • 100% ex-smokers • DLco<35%	Participants were recruited as part of phase Ib, non-randomized, no placebo-controlled, unicentric clinical trial	14	SGRQ	NR	FVC,% pred: 71.2±15.2	35.1	6.8	Not appropriate: no mapping methodology of SGRQ to EQ-5D in IPF patients available
						FVC 6 months post-infusion: 73.4% (18.1)	27.8	5.6	
						FVC 12 months post-infusion: 74.4% (17.5)	28.4	5.7	
			14		100%	FVC, %pred: 69.2 (14.6)	PF: 54.3 RP: 44.6 BP: 66.8 GH: 37.4 V: 52.5 SF: 58.0 RE: 69.0 MH: 65.1 PC summary: 36.1 MC summary 46.5	PF: 17.4 RP: 24.4 BP: 22.2 GH: 11.1 V: 13.3 SF: 14.4 RE: 44.3 MH: 17.9 PC summary: 9.1 MC summary: 10.9	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Jastrzebski et al. 2005 (Jastrzebski *et al.*, 2005[126]), Poland	• IPF patients: o The diagnosis was based on ATS and ERS; o 69% males; o Mean age 48.3; • COPD patients: o 50% males; o Mean age 53.4	Patients, qualified for lung transplantation in 2000-2004 at the Hospital of Pulmonary Diseases in Zabrze	16	SF-36	100%	FVC, % predicted: 44.0 (15.3)	PF: 45 RP: 43 BP: 62 GH: 28 V: 38 SF: 58 RE: 65 MH: 50 PH summary: 35 MC summary: 42	NR	Not appropriate: no mapping methodology of SF-36 in IPF patients available
			16		88%	FVC, % predicted: 45.1 (10.3)	PH: 18 RP: 11 BP: 43 GH: 20 V: 31 SF: 30 RE: 12 MH: 42 PC summary: 27 MC summary: 33	NR	

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
Swigris et al. 2010c (Swigris *et al.*, 2010[127]),	• IPF patients: o 73% males; o Mean age 65.12±8.93 o 98% non-smokers and 2% smokers o 92% Caucasian, 2% Black, 1% Asian, 5% Other	Previously reported in King et al. 2008 (King *et al.*, 2008[103])	100	SF-36; SGRQ	NR	FVC% predicted: 66.97 (12.17)	SF-36: PF: 35.4 RP: 37.8 BP: 47.6 GH: 37.8 V: 43.1 SF: 44.6 RE: 42.7 MH: 48.2 PH summary: 37 MH summary: 44.2 SGRQ: Total 44.8 Symptoms 50.1 Activity 60.6 Impact 33.7	SF-36: PF:10.3 RP: 11.6 BP: 10.7 GH: 9.4 V: 9.2 SF: 12.3 RE: 14.2 MH: 10.1 PH summary 10 MH summary 10.84 SGRQ: Total 19.5 Symptoms 21.9 Activity 22.8 Impact 20.6	Not appropriate: no mapping methodology of SF-36 or SGRQ in IPF patients available
Raghu et al. 2008 (Raghu *et al.*, 2008[128]), USA	IPF patients: • The diagnosis was based on ATS and ERS; • Etanercept o 76.1% males	Participants were recruited as part of a 48-week, randomized, double-blind, placebo-controlled, multicentre,	46	SF-36; SGRQ	74%	FVC, % predicted: 64.7 (14.1)	SF-36: PF: 36.6 MC: 54.3 SGRQ:	SF-36: PF: 11.1 MC: 9.3 SGRQ:	Not appropriate: no mapping methodology of SF-36 or SGRQ in IPF patients available

Study, Country	Population	Recruitment Info	No. of Respondents	Assessment Tools	Response Rate %	Health States and Time Period	Utility Values (mean)	SD and 95% CIs	Appropriateness for CEA
	o Mean age 65.2±7.7 • Placebo o 58.5% males o Mean age 65.1±7.1	phase II exploratory study evaluating the efficacy and safety of subcutaneous etanercept 25 mg twice weekly					40.8	18.1	
			41		76%	FVC, % predicted: 63.0 (12.7)	SF-36: PF: 37.9 MC: 48.9 SGRQ: 42.9	SF-36: PF: 10.3 MC: 12.0 SGRQ: 19.4	
Elfferich et al. 2011 (Elfferich *et al.*, 2011[129]), Netherlands	• IPF patients o 62.5% males o Mean age 63.1±11.8 • Sarcoidosis patients o 54.2% males o Mean age 45.4±11.4 • Control group o NR	Patients were recruited by the ILD care team of the outpatient clinic of the department of Respiratory Medicine of the Maastricht University Medical Centre	49	WHOQOL-BREF	72.1%	FVC% of predicted: 82.9 (27.3)	IPF patients: 10.5	3.3	Not appropriate: no mapping methodology of WHOQOL-BREF in IPF patients available
			441		75%	FVC% of predicted: 98.1 (19.9)	Sarcoidosis patients: 11.9	3.2	
			3678		NR	NR	NR	NR	

†Results of the another patients group were recorded, but not reported here as those were not consistent with the decision problem; *p < 0.05; **p < 0.01, ***p<0.001 vs. baseline

5.4.5 Comparison between values derived from the literature and those reported from the clinical trials

The two studies reporting values relevant for a comparison of HRQL with values from the nintedanib clinical trials recorded EQ-5D scores for patients receiving placebo of 0.718 (King *et al.*, 2011[100]) and 0.71 (Zisman *et al.*, 2010[101]). The value from King et al 2011 (King *et al.*, 2011[100]) relates to patients with a mean percent predicted FVD of 74.9%, and the value from Zisman et al 2010 (Zisman *et al.*, 2010[101]) relates to patients with a mean percent predicted of 58.73%. The HRQL scores from these two studies are, despite being slightly lower, broadly in-line with the values from the nintedanib clinical trials.

5.4.6 Adverse reactions

Adverse events related decrement in utility

Serious cardiac events

The serious cardiac events disutility was based on Ara and Brazier 2010 (Ara *et al.*, 2010[130]) from the following quality of life (QoL) values:

- Baseline value: 0.826 (0.802, 0.850) respondents of similar age not affected by the health condition
- Heart attack/angina: 0.628 (0.602, 0.653) for respondents affected by the health condition (and any other health condition)
- Estimated disutility value: -0.198
- An alternative disutility value of -0.099 was estimated, assuming half of the disutility from Ara and Brazier 2010 (Ara *et al.*, 2010[130]).

The base-case model conservatively assumed the higher disutility value (-0.198) for the serious cardiac events (with greater impact for NDB). The alternative disutility value (-0.099) was tested in one-way sensitivity analysis (OWSA).

Serious GI events

Quality of life data on serious GI events were analysed from the phase III clinical trial (Ingelheim, 2014[97]). Out of 26 patients with serious GI events (19 in the NDB arm and 7 in the PBO arm), one patient did not have any QoL data recorded, and the analysis was performed on 25 patients.

The impact of the serious GI events onto QoL was investigated in three ways:

1. using the breakdown before/during/after the serious GI event (expecting to observe a dip in QoL, then a recovery),
2. before/after the event (expecting to observe whether there is a long-term alteration of QoL after a serious GI event),
3. during the event or not (transient effect).

All three analyses controlled for exacerbations (i.e. patients with an acute exacerbations were only included up to the date of exacerbation) and FVC%Pred level. QoL was

modelled using a regression model with an intercept term, to reflect the fact that other factors may have an impact on QoL (i.e. FVC%Pred, treatment), and using a repeated-measures analysis to account for the fact that several data points were available per patient.

The model exploring a transient effect was the one that gave logical results. This reflected the assumption that QoL decreased when a serious GI event occurred, and eventually the patient recovered at a later time. The other two models (before/during/after and before/after) indicated that a serious GI event could have a beneficial long-term impact on QoL, which was considered to be illogical.

For the transient effect analysis, 25 patients had relevant QoL data and there were only 9 observations recorded during the event (7 in the NDB arm and 2 in the PBO arm). During a serious GI event, QoL dropped by -0.068 points (-0.201 to 0.065) and then recovered; therefore, the disutility of a serious GI event was assumed to be -0.068.

Skin disorders disutility

The skin disorder disutility was estimated from Ara and Brazier 2010 (Ara *et al.*, 2010[130]) from the following QoL values:

- Baseline value: 0.855 (0.834, 0.875) for respondents of similar age not affected by the health condition.
- Skin complaint: 0.773 (0.733, 0.812) for respondents affected by the health condition (and any other health condition).
- Estimated disutility value: -0.082.

GI perforation

Due to lack of evidence for the GI perforation disutility, a value for similar or related GI conditions was used in the model as a proxy. The base-case model uses the disutility of stomach ulcer/abdominal hernia/rupture estimated from Ara and Brazier 2010 (Ara *et al.*, 2010[130]) from the following QoL values:

Baseline value: 0.806 (0.781, 0.830) respondents of similar age not affected by the health condition.

Stomach ulcer/abdominal hernia/rupture: 0.688 (0.654, 0.720) for respondents affected by the health condition (and any other health condition).

Estimated disutility value: -0.118.

An alternative disutility value of -0.025 was assumed to be the GI bleeding disutility from Leontiadis et al. (Leontiadis *et al.*, 2007[131]) . This study reported EQ-5D data of 57 patients in the UK right after acute GI bleeding (at discharge or within 7 days of bleeding) and at 4 weeks follow-up. The baseline utility was not collected. EQ-5D results were reported as QoL in hospital (0.45) and QoL at home (0.78). Assuming the baseline utility to be 0.78, the GI bleeding utility to be 0.45, and the event duration to be 4 weeks, the disutility of a GI bleeding event was estimated to be -0.025. As a validation for this

assumption, another publication that reported a similar utility value for GI bleeding/ulcer (0.46) was identified, with an event duration of 1 month (Cooper K, 2010[132]).

The base-case model conservatively assumed the higher disutility value (-0.118) for the GI perforation (with greater impact for NDB). The alternative disutility value (-0.025) was tested in OWSA.

A summary of the disutility values is presented in Table 133.

Table 133: Summary of disutilities due to AEs

Adverse event	Disutility value	Source/assumption
Serious cardiac events	-0.1980	Ara and Brazier [2010]
Serious GI events	-0.0680	INPULSIS Summary of Clinical Safety 2014 post-hoc analysis
Skin disorders*	-0.0820	Ara and Brazier [2010]
GI perforation	-0.1180	Leontiadis et al. [2007]

*Rash and photosensitivity reaction were grouped as “skin disorders”

5.4.7 The patient experience in the health states in terms of HRQL in the cost effectiveness analysis

The HRQL of patients varies by their FVC%pred (split into 10% categories); a gradual decline in HRQL is experienced as percent predicted FVC decreases.

Patients suffering an acute exacerbation experience a sudden decline in HRQL for one month following the event, followed by a recovery and smaller and longer-lasting disability in the subsequent months. The estimates used in the model are assumed to be conservative, as it is likely that the worst patients were missing not at random (as they were unable or unwilling to attend the next study visit).

The decrease in HRQL of carers has not been captured.

5.4.8 HRQL change over time

Mean utilities from the clinical trials were used for the HRQL in different lung function states to accommodate the memory-less feature of the Markov approach. The change in HRQL is related to events/change in state rather than time.

5.4.9 Baseline HRQL assumed in the cost effectiveness analysis

The baseline HRQL was the same as the HRQL of patients in the health states.

5.4.10 Adjustment of health state utility values

No adjustment.

5.4.11 Additional health effects found in the literature

None

5.4.12 Summary of the utility values chosen for the cost effectiveness analysis and justification

See section 0 for a summary of the utilities used in different lung function health states, the disutilities applied for exacerbation and an explanation for these. See section 5.4.6 for a summary of disutilities applied for adverse events and an explanation for these.

5.4.13 Clinical expert opinion on utility values used

Clinical opinion was sought on the utility values used within the model at an advisory board held on the 23rd April 2014 (Boehringer Ingelheim, 2014[91])(see section 5.2.2 for further details). The clinicians agreed that the values used are plausible.

5.5 *Cost and healthcare resource use identification, measurement and valuation*

5.5.1 Parameters to estimate cost effectiveness

See section 5.6.

5.5.2 Identification of relevant costs and healthcare resource use

Resource use and unit costs search

Search objective

A systematic literature review was conducted to identify existing studies reporting resource use and/or unit costs for nintedanib or its comparators in adults with IPF.

Identification of studies

Databases

The following computerised bibliographic databases advised by the CRD (Dissemination, 2009[24]) were searched on 22/10/2014:

- MEDLINE;
- MEDLINE In-Process;
- Embase;
- EconLit;
- NHS Economic Evaluation Database (NHSEED).

MEDLINE, Embase and EconLit were searched via the Ovid® search platform, and the NHSEED was searched via the CRD website.

Search strategy

Nintedanib, pirfenidone, N-acetylcysteine, best supportive care and their alternative names were included in a search strategy that incorporated disease and patient population terms consistent with the clinical searches in section 4.1. In addition to those terms, intervention terms and terms for resource use and costs search filters were

added. The search was not limited by country, language, study design or date but studies published in languages other than English were not reviewed in full. The search strategies by databases are provided in Appendix A.

Conference proceedings search

Conference proceedings were searched from 2012 to 2014 to identify publications on resource use and/or costs in adults and elderly with IPF, treated with nintedanib or its comparators for the following conference sites:

- International Society for Pharmacoeconomics and Outcomes Research (ISPOR) (www.ispor.org);
- International Health Economics Association (iHEA) (www.healtheconomics.com).

Study selection

The process of study selection in terms of abstract and full-text review was the same as is detailed in section 4.1. However, different PICOS criteria were developed to capture unit costs and resource use studies. Those are presented in Table 134.

Table 134: Eligibility criteria for unit costs and resource use search

Criteria	Inclusion criteria	Exclusion criteria
Population	• Adult patients with IPF (including subgroups of patients such as subgroups based on 6MWT SpO2 and %VC)	• Paediatric and younger than 18; • Healthy participants; • Non-human.
Intervention/Comparators	• Nintedanib • Pirfenidone • N-acetylcysteine (monotherapy) • Best supportive care (placebo)	• NAC triple therapy • Lung transplantation
Outcomes	• Unit costs; • Resource use (e.g. number of hospitalisations).	• No outcomes
Study design	• Economic evaluations: CUA, CBA, CEA, CCA, CMA; • Unit costs and resource use studies.	• Reviews and systematic literature reviews* of economic evaluations; • Case reports; • Critical appraisals; • Update or commentaries on data published elsewhere; • Notes, letters or editorials.

*Systematic literature reviews will be excluded from the main database search but their reference lists will be searched for relevant studies not captured by the main database search and to be included as hand search in the review.

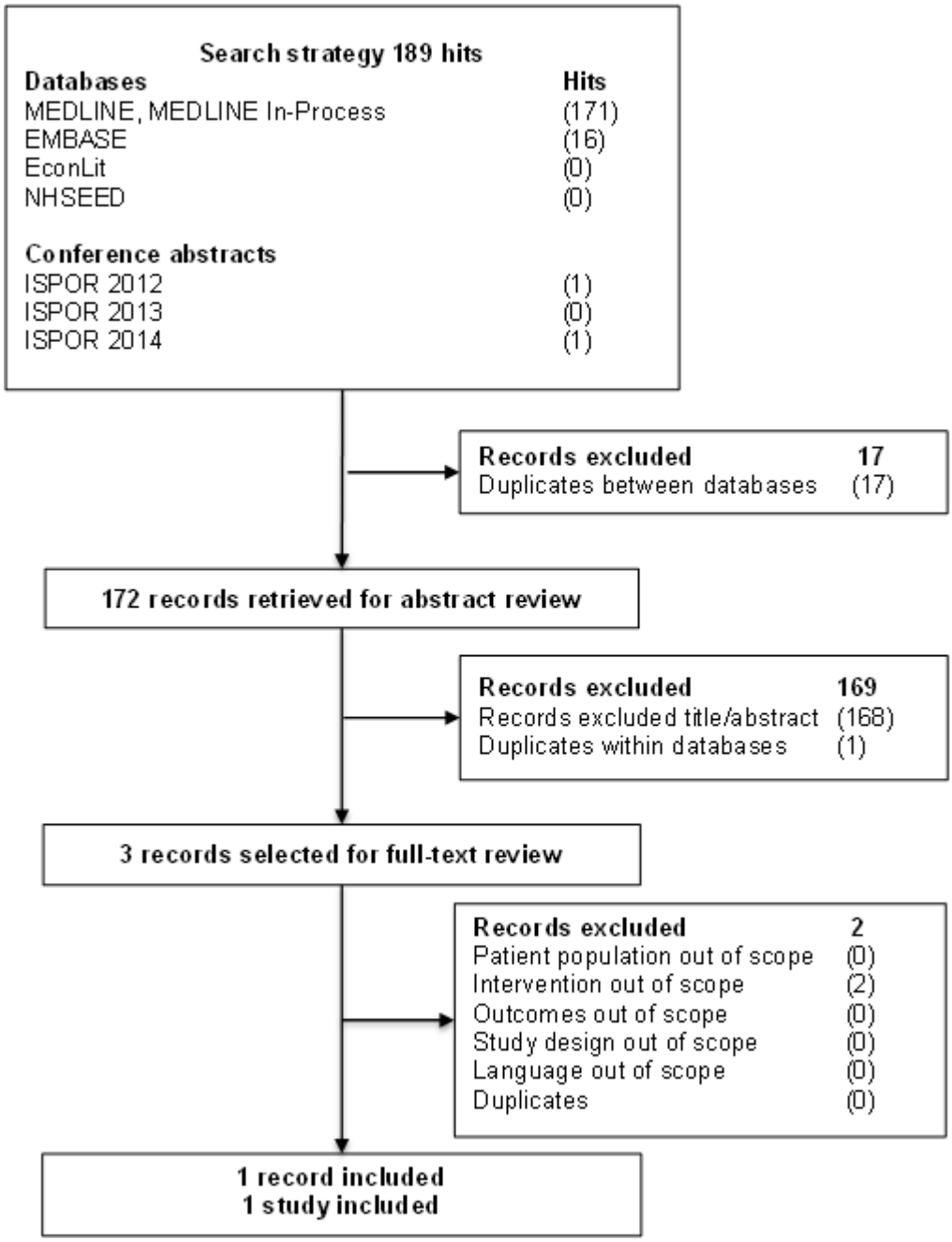
Data Extraction

The process of data extraction for economic evaluation studies was the same as is detailed in section 4.1.

A total of 189 records were identified in the systematic literature review for resource use and unit costs studies of nintedanib or its comparators in adults with idiopathic pulmonary fibrosis. Following duplicate removal, 172 abstracts were retained for assessment against the predefined eligibility criteria described above. Data extraction

was not performed as only one record in the form of an abstract was deemed to be relevant to the decision problem. Figure 51 illustrates the study selection process.

Figure 51: Study flow diagram for resource use study selection (PRISMA diagram)



The identified abstract (Parfrey, 2013[133]) reported hospital resource use data collected over nine-month observation period of a multi-centre, retrospective, cohort review undertaken across four NHS Trusts. A total of 100 patients, 76% males with mean age 69.3 years, starting pirfenidone for IPF through the named patient programme (NPP) before June 2012 were enrolled. During the first six months of therapy, 15 IPF-related

hospitalisations (6 due to acute exacerbations), with mean length of stay 11.0 + 7.5 days were reported. One patient was admitted to the Intensive Care Unit, for 5 days, and 18 patients had IPF-related Accident and Emergency department visits. Three patients had IPF-related day-case and 67 outpatient clinic visits, with a mean of 2.1±2.0 outpatient clinic visits per patient.

5.5.3 Costing by NHS reference costs or payment-by-results (PbR) tariffs.

NHS reference costs were used to cost the treatment of the intervention being appraised. A full description of the costing of the condition and its treatment is provided in sections 5.5.5/5.5.6/5.5.7.

5.5.4 Clinical expert assessment of costs and healthcare resource use values

Some discussion was performed around resource use within the model at an advisory board with two clinicians held on the 23rd April 2014 (Boehringer Ingelheim, 2014[91]) (see section 5.2.2 for further details).

- The clinicians found that exacerbation analysis was reasonable.
- The clinicians suggested end of life costs should be included.
- Regarding oxygen therapy use, the clinicians suggested that the clinical trial is excluding patients with emphysema and therefore, taking the trial data as a benchmark for oxygen use might not be appropriate. Both clinicians found the estimates very high and suggested considering a lower benchmark. They suggested focusing the analysis only on ambulatory costs. In the UK, patients who need overnight and ambulatory oxygen they would use a concentrator which is a relatively cheap solution.

Intervention and comparators' costs and resource use

5.5.5; 5.5.6; 5.5.7 and 5.5.8 Costing within the cost effectiveness model

Cost inputs

The cost inputs include the following components:

- Treatment related costs (since the active arms involved oral treatments, it was assumed that there was no administration cost associated with the treatments)
- Drug acquisition costs
- Treatment-related adverse events costs
- Liver function test costs
- Concomitant medications
- Background follow-up costs
- Oxygen use costs
- Exacerbation costs
- End of life costs

Drug acquisition costs

The cost of NDB is at parity of the published list price to PFN in the UK, £71.70 per day. No cost was assumed for BSC, as it reflected the PBO (control) arm of the clinical trial.

The daily cost of PFN used in the model was £71.70 (list price; pirfenidone has a confidential discount in place) per day (MIMS (Monthly Index of Medical Specialties), March-May 2014[134]). The PFN dosage was assumed to be 2,403mg/day (Electronic Medicines Compendium (eMC), Accessed May 2014[9]).

Treatment-related AE costs

The treatment-related AE cost inputs (values, sources and assumptions) are presented in Table 135 below.

Table 135: Adverse event-related costs

Adverse event	Value (£)	Source/assumption
Serious cardiac events	£2,054.18	NHS Reference Costs 2012/13, non-elective inpatients, long stay, weighted average of Actual or Suspected Myocardial Infarction with CC Score 4-13, service code 340, Respiratory medicine
Serious GI events	£1,749.45	NHS Reference Costs 2012/13, non-elective inpatients, long stay, weighted average of Inflammatory Bowel Disease, without Interventions, with CC Score 3-4 and 5+, service code 340, Respiratory medicine
Skin disorders	£373.07	NHS Reference Costs 2012/13, non-elective inpatients, short stay, Rash or Other Non-specific Skin Eruption, service code 340, Respiratory medicine
GI perforation	£2,352.99	NHS Reference Costs 2012/13, non-elective inpatients, long stay, weighted average of Gastrointestinal Bleed with CC Score 5-9+, service code 340, Respiratory medicine

Liver function tests costs

Patients on NDB experienced elevated hepatic enzyme values, with an incidence of 2.2% in the NDB arm and 0.5% in the PBO arm (Ingelheim, 2014[135]). Further to that, the PFN Summary of Product Characteristics (SPC) (Electronic Medicines Compendium (eMC), Accessed May 2014[9]) also reports that patients on PFN experienced elevations in hepatic enzyme levels.

Liver function tests were assumed to be routinely performed on patients receiving NDB and PFN. The cost of a liver panel blood test was estimated at £3.01 (NHS Reference Costs 2012/13, NHS Trusts and PCTs combined Direct Access: Pathology Services: DAPS05 Haematology (Department of Health, 2013[8])). The model assumed that all patients on active treatment would incur this cost, at a quarterly frequency (every 3 months, i.e. every cycle). This frequency schedule is the same as the maintenance test frequency recommended by the PFN SPC (Electronic Medicines Compendium (eMC), Accessed May 2014[9]).

Concomitant medications

According to the INPULSIS Summary of Clinical Efficacy (Boehringer Ingelheim, 2014[136]), concomitant medications were allowed throughout the trial. For completeness, concomitant on-treatment therapies were extracted and their costs were estimated. Concomitant medications with an incidence of over 10% in either treatment arm (NDB or PBO) were selected. The medications used and the respective incidence for the pooled INPULSIS data (trials 1199.32 and 1199.34) are presented in Table 136 below.

Table 136: Concomitant medication - incidence (>10%)

Concomitant therapy	PBO N (%) Total n = 423	NDB 150mg bid N (%) Total n = 638
Antidiarrhoeal agent	22 (5.2)	223 (35.0)
P-glycoprotein inhibitor	139 (32.9)	192 (30.1)
Systemic corticoid	123 (29.1)	143 (22.4)
Proton pump and inhibitor and H2-receptor antagonists	77 (18.2)	143 (22.4)
Antioxidant/expectorants	76 (18.0)	98 (15.4)
Bronchodilator	86 (20.3)	89 (13.9)
Antitussive	61 (14.4)	71 (11.1)

*Source: Table 3.1.2.1.5:2, INPULSIS Summary of Clinical Efficacy 2014

The probabilities adjusted by clinical trial duration (52 weeks) and cycle length (3 months) are presented below (Table 137).

Table 137: Concomitant medications - probabilities per cycle

Concomitant therapy	PBO arm	NDB arm
Antidiarrhoeal agent	0.0133	0.1024
P-glycoprotein inhibitor	0.0952	0.0859
Systemic corticoid	0.0826	0.0616
Proton pump and inhibitor and H2-receptor antagonists	0.0491	0.0616
Antioxidant/expectorants	0.0485	0.0411
Bronchodilator	0.0553	0.0368
Antitussive	0.0382	0.0291

In order to cost each concomitant therapy, the most prevalent drug within each class was considered. The most prevalent drugs for each therapy class are presented below (Table 138).

Table 138: Most prevalent drug for each concomitant therapy class

Concomitant therapy class	Most prevalent drug	Incidence N (%)
Antidiarrhoeal agent	Loperamide hydrochloride	1 (0.3)
P-glycoprotein inhibitor	Omeprazole	57 (14.5)
Systemic corticoid	Prednisone	46 (11.7)

Concomitant therapy class	Most prevalent drug	Incidence N (%)
Proton pump and inhibitor and H2-receptor antagonists	Omeprazole	57 (14.5)
Antioxidant/expectorants	Carbocisteine	9 (2.3)
Bronchodilator	Budesonide with formoterol	17 (4.3)
Antitussive	Codeine phosphate	16 (4.1)

*Note that the % values are calculated from the total number of patients who have taken concomitant medications (N=394).

The costs per cycle for each treatment arm were estimated for each concomitant therapy class by taking into account the daily drug cost, daily dosage, and probabilities per cycle (from Table 139).

Table 139: Concomitant therapy cost per cycle

Concomitant therapy class	Cost per cycle
Antidiarrhoeal agent	£9.81
P-glycoprotein inhibitor	£3.75
Systemic corticoid	£162.43
Proton pump and inhibitor and H2-receptor antagonists	£162.43
Antioxidant/expectorants	£72.16
Bronchodilator	£0.21
Antitussive	£5.99

*The unit drug costs were obtained from MIMS June-August 2014 (printed edition) (MIMS (Monthly Index of Medical Specialties), March-May 2014[134]); the daily dosages were approximated from each drug's therapeutic indications (Electronic Medicines Compendium (eMC), Accessed May 2014[9])

The final cost per cycle for the BSC arm was calculated to be £25.62, while the NDB arm was £24.48. Whilst the incremental difference of £1.14 per cycle favoured NDB, the overall impact was considered to be negligible, and was not explored further in the model.

Background follow-up costs

The background follow-up costs were compiled using the IPD data from INPULSIS (Ingelheim, 2014[97]). The following health care resources and their respective descriptions of components were available in the INPULSIS data and were further analysed for the economic model:

- **Hospital:** average number of hospitalisations, average duration of each hospitalisation (days), percentage (%) hospitalisation associated with intensive care unit (ICU) stay, percentage (%) hospitalisation associated with Emergency Room (ER) overnight stay, percentage (%) hospitalisation associated with ambulance use
- **ER**: average number of ER visits, percentage (%) ER visits associated with ambulance use
- **Visits**: visits to general practitioner (GP), specialist, nurse, physiotherapist, other visits

- **Procedures**: chest High Resolution Computed Tomography (HRCT), chest X-ray, oxygen (O_2) requirement assessment, bronchoalveolar lavage, Computed Tomography (CT) pulmonary angiogram, right heart catheterization, general diagnostic (aggregation of several services: bronchoscopy, sputum examination, surgical lung biopsy, transbronchial biopsy, other)

The Health care Research Unit (HCRU) data were analysed and adjusted to match the model health states, i.e. by FVC%pred categories. A per-cycle probability (3-month probability) of incurring the resource use was calculated. The number of observations were large enough in each category except for the <40% and 40-49.9% FVC%pred group; as a result, the two categories were combined. The number of observations for each FVC%pred group is presented in Table 140.

Table 140: Number of HCRU observations in each FVC%pred group

FVC%pred group	Number of observations
<40 – 49.9*	85
50-59.9	470
60-69.9	793
70-79.9	801
80-89.9	653
90-99.9	455
100-109.9	260
≥110	224

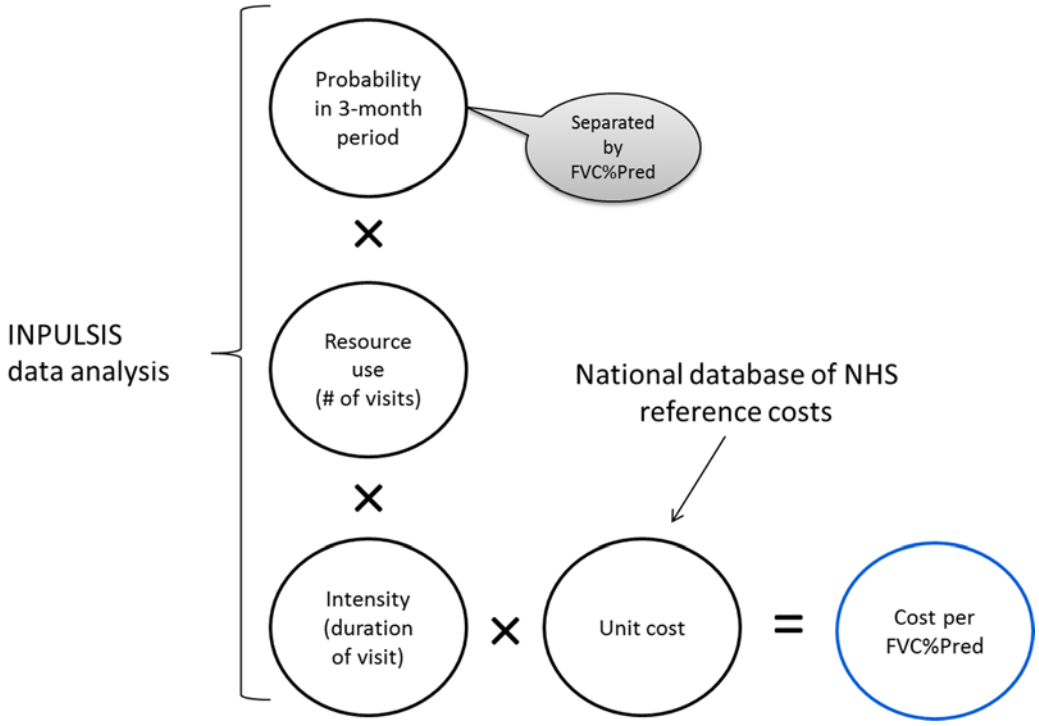
*The "<40" category contained 9 observations, while the "40-49.9" group had 76 observations.

Regarding the intensity of use within each HCRU group (i.e. hospital, ER, visits, and procedures) the number of observations was too low to permit a similar analysis by FVC%pred group. Therefore, the intensity of each resource use (e.g. the average number of hospitalisations, the average duration in hospital) was averaged-out across all FVC%pred groups. The number of patients was used as a weight for this calculation.

Synthesis of data

The synthesis of the INPULSIS data with unit costs from the NHS reference costs is presented in the diagram below (Figure 52):

Figure 52: Diagram of HCRU analysis



Hospitalisation cost synthesis

The cost inputs for all of the hospitalisation items (values, assumptions, sources) are presented in Table 141 below.

Table 141: Cost inputs for hospitalisation

Parameter description	Value used	Sources and assumptions
Hospitalisation	£303.73	Weighted average of DZ27D (Respiratory Failure without Intubation with Major CC), DZ27E (Respiratory Failure without Intubation with CC), DZ27F (Respiratory Failure without Intubation without CC); Cost inflated to 2012/13 from National Schedule of Reference Costs Year 2009/10'- NHS Trusts Elective Inpatient Excess Bed Day HRG Data (Department of Health, 2013[8])
ICU stay	£829.81	Weighted average of XC06Z (Adult Critical Care, 1 organ supported) and XC07Z (Adult Critical Care, 0 organs supported), Adult Critical Care Unit; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Critical Care (Department of Health, 2013[8])
Mechanical ventilation	£147.52	Weighted average of non-invasive ventilation support assessment 19 yrs and over (DZ37A), Respiratory medicine and Respiratory physiology; procedures in outpatients; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts (Department of Health, 2013[8])
ER overnight stay	£154.31	Weighted average across all types of A&E services, all treatment categories; filtered by admitted and non-dental; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Accident and Emergency Services (Department of Health, 2013[8])
Ambulance use	£204.53	Weighted average of ASH1 (hear and treat or refer), ASS01 (see and treat or refer), ASS02 (see and treat and convey); National Schedule of Reference Costs Year : 2012-13 - All NHS trust and NHS foundation trusts - ambulance services (Department of Health,

Parameter description	Value used	Sources and assumptions
		2013[8])

This cost synthesis of the hospitalisation resource use was the most complex (Figure 53), as it was composed of:

- Average number of hospitalizations per patient with at least one hospitalisation (1.124, SE 0.029)
- Average duration of hospitalisation (8.72 days, SE 0.692)
- Proportion of hospitalisation associated with an ICU stay (3.8%, SE 1.4%)
- Proportion of hospitalisation associated with mechanical ventilation use (1.1%, SE 0.8%)
- Proportion of hospitalisation associated with an ER overnight stay (4.8%, SE 1.6%)
- Proportion of hospitalisation associated with ambulance use (12.9%, SE 2.5%)

All of the above values were obtained from a post-hoc analysis of INPULSIS phase III trial and are per patient values, all-cause (not just IPF-related).

The sum and product of all of the items rendered a total hospitalisation cost of £3,044. This sum was then multiplied by the probabilities of each FVC%pred group (see Figure 54).

Figure 53: Cost synthesis: hospitalisation

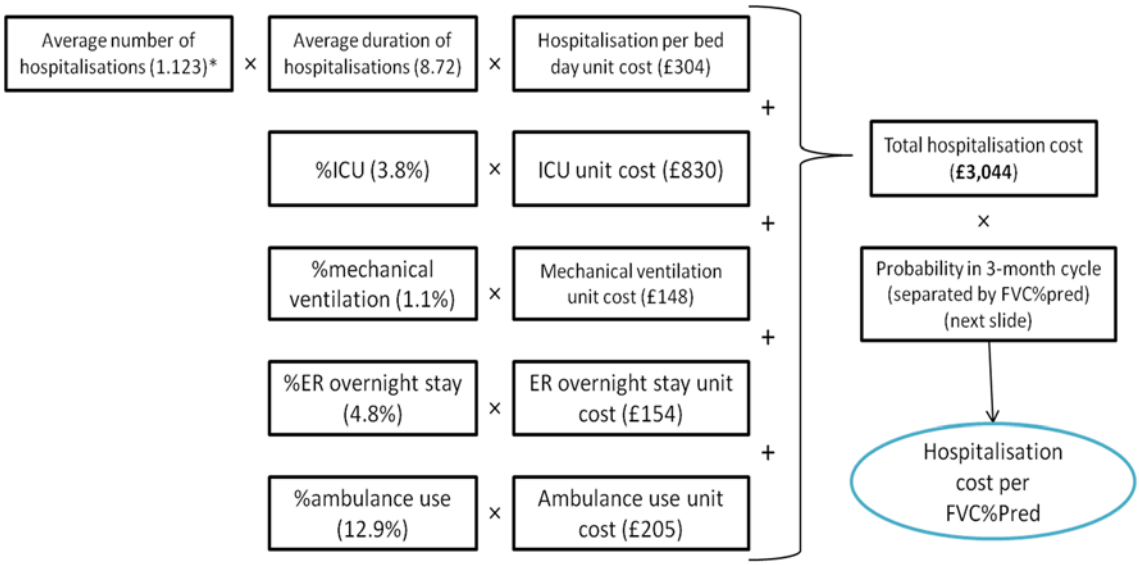


Figure 54: Hospitalisation cost per FVC%pred group

FVC%pred group	3-month probability of hospitalisation		3-month total hospitalisation costs
>=110	0.0491		£149.49
100-109.9	0.0423		£128.79
90-99.9	0.0505	× £3,044 →	£153.88
80-89.9	0.0368		£111.88
70-79.9	0.0474		£144.41
60-69.9	0.0416		£126.68
50-59.9	0.0681		£207.26
<40-49.9	0.1647		£501.38

A&E cost synthesis

The costs associated to each A&E item were assumed to be:

- Unit cost of A&E visit: £101.16
    - Source/assumption: Weighted average across all types of A&E services, all treatment categories; filtered by not admitted and non-dental; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Accident and Emergency Services (Department of Health, 2013[8]).
- Cost of ambulance use: £204.53 (same as in hospitalisation, Table 141 above Cost inputs for hospitalisation)
    - Source/assumption: Weighted average of ASH1 (hear and treat or refer), ASS01 (see and treat or refer), ASS02 (see and treat and convey); National Schedule of Reference Costs Year: 2012-13 - All NHS trust and NHS foundation trusts - ambulance services (Department of Health, 2013[8]).

The 3-month probabilities of incurring an ER cost for each FVC%pred group were:

Table 142: 3-month probabilities of ER visits

FVC%pred group	3-month probability
>=110	0.0357
100-109.9	0.0192
90-99.9	0.0308
80-89.9	0.0368
70-79.9	0.0300
60-69.9	0.0328

FVC%pred group	3-month probability
50-59.9	0.0298
<40-49.9	0.0235

The average number of ER visits (per patient) in INPULSIS was 1.094 (SE 0.038). The proportion of ER visits associated with ambulance use were 17.1% (SE 3.5%) (Ingelheim, 2014[135]).

Outpatient visits cost synthesis

GP visits

The unit cost of a GP visit was assumed to be £37.50 (PSSRU 2013 (Curtis L, 2013[137]); assumption: per patient contact visit lasting 11.7 minutes).

The 3-month probabilities of incurring a GP visit cost for each FVC%pred group were:

Table 143: 3-month probabilities of GP visit

FVC%pred group	3-month probability
>=110	0.1563
100-109.9	0.1654
90-99.9	0.1516
80-89.9	0.1884
70-79.9	0.1935
60-69.9	0.1828
50-59.9	0.1574
<40-49.9	0.1647

The average number of GP visits (per patient) was 1.585 (SE 0.054) (Ingelheim, 2014[97]).

Specialist visits

The unit cost of a specialist visit was assumed to be £119.0 (PSSRU 2013 (Curtis L, 2013[137]), assumption: average of medical consultant per contract hour, with (£139) and without (£99) qualifications].

The 3-month probabilities of incurring a specialist visit cost for each FVC%pred group were:

Table 144: 3-month probabilities of specialist visits

FVC%pred group	3-month probability
>=110	0.1116
100-109.9	0.1692
90-99.9	0.1758

FVC%pred group	3-month probability
80-89.9	0.1730
70-79.9	0.2260
60-69.9	0.1929
50-59.9	0.2085
<40-49.9	0.2824

The average number of specialist visits (per patient) was 1.823 (SE 0.052) (Ingelheim, 2014[97]).

Nurse visits

The unit cost of a nurse visit was assumed to be £74.97 (National Schedule of Reference Costs 2012-13, specialist nursing - asthma and respiratory nursing/liaison, adult, face to face; N08AF; Community Health Services – Nursing (Department of Health, 2013[8])).

The 3-month probabilities of incurring a nurse visit cost for each FVC%pred group were:

Table 145: 3-month probabilities of nurse visits

FVC%pred group	3-month probability
>=110	0.0089
100-109.9	0.0154
90-99.9	0.0110
80-89.9	0.0153
70-79.9	0.0237
60-69.9	0.0265
50-59.9	0.0191
<40-49.9	0.0118

The average number of nurse visits (per patient) was 1.634 (SE 0.186) (Ingelheim, 2014[97]).

Physiotherapist visits

The unit cost of a physiotherapist visit was assumed to be £50.44 (National Schedule of Reference Costs 2012-13, physiotherapist, adult, one to one; A08A1, Community Health Services - Allied Health Professionals (Department of Health, 2013[8])).

The 3-month probabilities of incurring a physiotherapist visit cost for each FVC%pred group were:

Table 146: 3-month probabilities of physiotherapist visits

FVC%pred group	3-month probability
>=110	0.0000
100-109.9	0.0038
90-99.9	0.0044
80-89.9	0.0092
70-79.9	0.0125
60-69.9	0.0164
50-59.9	0.0064
<40-49.9	0.0000

The average number of physiotherapist visits (per patient) was 3.514 (0.628) (Ingelheim, 2014[97]).

Other visits

The unit cost of an "other" visit was assumed to be £169.99 (PSSRU 2010: breakdown of inpatient attendance: per bed-day; cost inflated from 2009/10 to 2012/13 (Curtis L, 2010[138])).

The 3-month probabilities of incurring a visit cost classified as "other" for each FVC%pred group were:

Table 147: 3-month probabilities of other visits

FVC%pred group	3-month probability
>=110	0.0268
100-109.9	0.0308
90-99.9	0.0220
80-89.9	0.0429
70-79.9	0.0287
60-69.9	0.0404
50-59.9	0.0298
<40-49.9	0.0471

The average number of other visits (per patient) was 1.920 (SE 0.166) (Ingelheim, 2014[97]).

Procedures cost synthesis

Chest HRCT

The unit cost for chest HRCT was assumed to be £50.03 (National Schedule of Reference Costs 2012-13 (Department of Health, 2013[8]): weighted average of RA08A (CT scan, one area, no contrast, 19 yrs and over) and RA09A (CT scan, one area, with post contrast only, 19 yrs and over); Diagnostic imaging - direct access; respiratory medicine).

The 3-month probabilities of incurring a chest HRCT cost for each FVC%pred group were:

Table 148: 3-month probabilities of chest HRCT

FVC%pred group	3-month probability
>=110	0.0670
100-109.9	0.0692
90-99.9	0.0396
80-89.9	0.0567
70-79.9	0.0836
60-69.9	0.0731
50-59.9	0.1021
<40-49.9	0.1529

The average number of HRCT procedures (per patient) was 1.099 (SE 0.024) (Ingelheim, 2014[97]).

Chest X-ray

The unit cost of a chest X-ray was assumed to be £28.28 (National Schedule of Reference Costs 2012-13: DAPF; Directly Accessed Diagnostic Services (Department of Health, 2013[8])).

The 3-month probabilities of incurring a chest X-ray cost for each FVC%pred group were:

Table 149: 3-month probabilities of chest X-ray

FVC%pred group	3-month probability
>=110	0.1071
100-109.9	0.1269
90-99.9	0.1143
80-89.9	0.1317
70-79.9	0.1373
60-69.9	0.1072
50-59.9	0.0894
<40-49.9	0.2000

The average number of chest X-ray procedures (per patient) was 1.568 (SE 0.085) (Ingelheim, 2014[97]).

Oxygen requirement assessment

The unit cost for an oxygen requirement assessment was assumed to be £173.11 (National Schedule of Reference Costs 2012-13: weighted average of DZ38Z oxygen assessment and monitoring in Respiratory Medicine and Respiratory Physiology; procedures in outpatients (Department of Health, 2013[8])).

The 3-month probabilities of incurring an oxygen requirement assessment cost for each FVC%pred group were:

Table 150: 3-month probabilities of O_2 requirement assessment

FVC%pred group	3-month probability
>=110	0.0134
100-109.9	0.0000
90-99.9	0.0110
80-89.9	0.0123
70-79.9	0.0200
60-69.9	0.0277
50-59.9	0.0298
<40-49.9	0.0706

The average number of O2 assessment procedures (per patient) was 1.324 (SE 0.144) (Ingelheim, 2014[97]).

Bronchoalveolar lavage

The unit cost of bronchoalveolar lavage was assumed to be £105.43 (National Schedule of Reference Costs 2012-13: Minor Mouth or Throat Procedures, 19 years and over without CC (CZ01Y), Directly Accessed Diagnostic Services (Department of Health, 2013[8])).

The 3-month probabilities of incurring a bronchoalveolar lavage cost for each FVC%pred group were:

Table 151: 3-month probabilities of bronchoalveolar lavage

FVC%pred group	3-month probability
>=110	0.0000
100-109.9	0.0000
90-99.9	0.0000
80-89.9	0.0046
70-79.9	0.0025
60-69.9	0.0000
50-59.9	0.0021
<40-49.9	0.0000

The average number of bronchoalveolar lavage procedures (per patient) was 1.000 (SE 0.000) (Ingelheim, 2014[97]).

CT pulmonary angiogram

The unit cost of a CT pulmonary angiogram was assumed to be £50.03 (National Schedule of Reference Costs 2012-13 (Department of Health, 2013[8]): weighted average of RA08A (CT scan, one area, no contrast, 19 yrs and over) and RA09A (CT scan, one

area, with post contrast only, 19 yrs and over); Diagnostic imaging - direct access; respiratory medicine).

The 3-month probabilities of incurring a CT pulmonary angiogram cost for each FVC%pred group were:

Table 152: 3-month probabilities of CT pulmonary angiograms

3-month probability	3-month probability
>=110	0.0000
100-109.9	0.0000
90-99.9	0.0088
80-89.9	0.0031
70-79.9	0.0012
60-69.9	0.0050
50-59.9	0.0106
<40-49.9	0.0118

The average number of CT pulmonary angiogram procedures (per patient) was 1.176 (SE 0.176) (Ingelheim, 2014[97]).

Right heart catheterisation

The unit cost of a right heart catheterization procedure was assumed to be £54.35 (National Schedule of Reference Costs 2012-13: Minor Cardiac Procedures, with CC Score 0-1; EA44B, Directly Accessed Diagnostic Services (Department of Health, 2013[8])).

The 3-month probabilities of incurring a right heart catheterisation cost for each FVC%pred group were:

Table 153: 3-month probabilities of right heart catheterisation

FVC%pred group	3-month probability
>=110	0.0000
100-109.9	0.0038
90-99.9	0.0022
80-89.9	0.0031
70-79.9	0.0062
60-69.9	0.0050
50-59.9	0.0170
<40-49.9	0.0471

The average number of right heart catheterization procedures (per patient) was 1.080 (SE 0.055) (Ingelheim, 2014[97]).

General diagnostic

The "general diagnostic" item is an aggregation of bronchoscopy, sputum examination, surgical lung biopsy, transbronchial biopsy, and "other."

The unit cost for the general diagnostic item was assumed to be £105.43 (National Schedule of Reference Costs 2012-13: Minor Mouth or Throat Procedures, 19 years and over without CC (CZ01Y), Directly Accessed Diagnostic Services (Department of Health, 2013[8])).

The 3-month probabilities of the composite general diagnostic procedures for each FVC%pred group were:

Table 154: 3-month probabilities of general diagnostic

FVC%pred group	3-month probability
>=110	0.0580
100-109.9	0.0538
90-99.9	0.0681
80-89.9	0.0858
70-79.9	0.0824
60-69.9	0.0858
50-59.9	0.0894
<40-49.9	0.1176

The average number of general diagnostic procedures (per patient) was 1.580 (SE 0.065) (Ingelheim, 2014[97]).

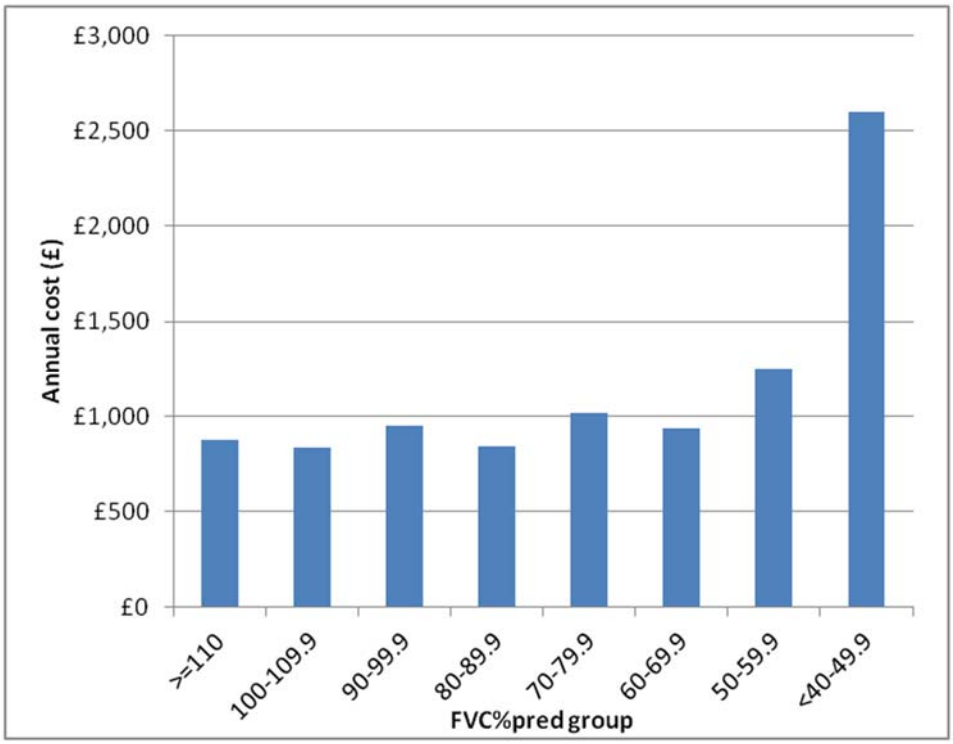
Total follow-up cost synthesis

The summation of all follow-up costs (hospitalisation, ER, visits, procedures) produced the following total costs per FVC%pred group (Table 155). Figure 55 presents the trend in the costs by FVC%pred category. The cost fluctuated around £900 for FVC%pred of 60-110+, and then increased for the lower FVC%pred groups (50-59.9, <40-49.9). As expected, the highest total cost was observed for the lowest FVC%pred group. Note that the total costs were mainly driven by the calculations on patient hospitalisation.

Table 155: Total cost per FVC%pred group

FVC%pred group	Per cycle	Annual
≥110	£219.19	£876.78
100-109.9	£209.37	£837.49
90-99.9	£236.57	£946.30
80-89.9	£210.37	£841.46
70-79.9	£253.49	£1,013.96
60-69.9	£233.51	£934.04
50-59.9	£312.53	£1,250.12
<40-49.9	£649.17	£2,596.66

Figure 55: Total annual costs for each FVC%pred group



The INPULSIS IPD also provided information about several services that were not included in the health care resource use (HCRU) analysis for the cost-effectiveness model due to reasons explained below:

Table 156: HCRU services excluded from the analysis

Group	Service	Observations	Reason for exclusion
Visits	Occupational therapy visit	4	Double-counting with physiotherapy
Sessions	Daily living assistance	28	Societal perspective, not well defined
	Rehabilitation therapy	51	Double-counting with physiotherapy
	Palliative/end of life care	3	Not well defined, low numbers

Oxygen use costs

Patients with IPF should receive supportive long-term oxygen supplementation in case of resting hypoxemia (National Institute for Health and Care Excellence, January 2013[76]).

Due to data inconsistencies and a small sample size, a thorough and meaningful analysis of the oxygen use costs from the INPULSIS data was not possible. The NICE Clinical Guidelines (National Institute for Health and Care Excellence, January 2013[76]) (2013a) sources and assumptions were used instead. Although there is poor correlation between breathlessness and FVC levels, guidelines suggest that a 10% drop in FVC%pred is significant and patients should be offered oxygen support. Considering

that patients with an FVC%Pred of 80% are in a relatively good health state, it was assumed that patients who drop below that level would require oxygen supplementation.

The cost of oxygen supplementation was estimated at £418 per 3-month cycle (NHS Reference Costs 2010/11, cycle-length adjusted from £1,600 per year, value inflated to 2012/13 costs (Department of Health, 2013[8])).

Acute exacerbation costs

Data from INPULSIS were analysed to calculate 3-month probabilities of visiting the hospital, using an ER, visiting a GP, and a specialist following an acute exacerbation event. With regards to hospitalisation, the HCRU data set was not complete, possibly due to the retrospective nature of collecting HCRU data within the trial combined with a high mortality and morbidity risk following an exacerbation. An analysis was therefore conducted on the safety data set where hospitalisations due to respiratory causes and exacerbations were recorded.

Table 157: 3-month probabilities of exacerbation costs

Health care resource	Total number of new exacerbations	Number who used service	3-month probability	Source
Hospitalisation	63	40	63.49%	Post-hoc analysis of INPULSIS data
ER		5	7.94%	
GP visit		5	7.94%	
Specialist visit		10	15.87%	

The HCRU data set was used for hospitalisations to obtain evidence on the intensity of the visits:

- The average number of hospitalisations was 1.3 (SE 0.18) (Ingelheim, 2014[97])
- The average duration of each hospitalization was 16.3 days (SE 4.72) (Ingelheim, 2014[97])

For the GP and specialist visits, the model assumed the same resource use as in the background follow-up cost synthesis (1.59 for GP visits, 1.82 for specialist visits; (PSSRU 2013 (Curtis L, 2013[137]), assumption: per patient contact visit lasting 11.7 minutes)).

The exacerbation cost synthesis was calculated as the total of the four cost items (3 – month probability × intensity/duration × unit cost) and is presented in Table 158 below:

Table 158: Exacerbation costs per item and total

Health care resource cost	Unit cost*	Total item cost
Hospitalisation	£303.73	£4,086.40
ER visit	£101.16	£8.03
GP visit	£37.5	£4.72
Specialist visit	£119.0	£34.44
Total exacerbation cost		£4,133.59

*The unit costs sources and assumptions are the same as described above

The total exacerbation cost (£4,134) was added to all patients who experience a new exacerbation case, for each treatment arm (BSC, NDB and PFN).

End of life costs

Clinical experts [Dr. Toby Maher and Dr. Gisli Jenkins] (Maher T, 2014[77]) suggested that palliative care is an important component of care for IPF patients. The model assumed that patients receive palliative care (in addition to background health care resources) as they reach the end of their life. The model applied an end of life (EoL) cost for the last year of patients' lives. The cost of EoL was formed of hospice and home care (excluding hospital) and it was estimated at £3,920.64 per cycle (£8 for hospice care, £9,098 for home care – cycle length adjusted, inflated to 2012/13 values (Hatziandreu E, 2008[139])).

5.6 *Summary of base-case de novo analysis inputs and assumptions*

Table 159: Summary of variables applied in the economic model

Category	Variable	Value	Variance	Source/assumption
General settings of the model				
General settings of the model	Starting age	66.75	Fixed	NDB Phase III trial CSR: Table 3.1.2.1.1:1, Summary of Clinical Efficacy 2014 (Ingelheim, 2014[140])
	Cycle length (days)	91.25	Fixed	Assumed based on three month follow-up interval of the INPULSIS clinical trial
	Cycle length (year)	0.25	Fixed	
	Discount rate – effect	3.50%	Fixed	NICE Reference Case
	Discount rate - cost	3.50%	Fixed	NICE Reference Case
	WTP	£30,000	Fixed	NICE Reference Case
Baseline characteristics				
Proportion of patients at each starting FVC%pred	≥110 FVC%pred group	0.06	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{110\ and\ above}$=64, N_{total}=1066(Ingelheim, 2014[97])
	100-109.9 FVC%pred group	0.08	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{100-109.9}$=80, N_{total}=1066 (Ingelheim, 2014[97])
	90-99.9 FVC%pred group	0.12	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{90-99.9}$ =132, N_{total}=1066 (Ingelheim, 2014[97])
	80-89.9 FVC%pred group	0.20	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{80-89.9}$=211, N_{total}=1066 (Ingelheim, 2014[97])
	70-79.9 FVC%pred group	0.20	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{70-79.9}$=216, N_{total}=1066 (Ingelheim, 2014[97])
	60-69.9 FVC%pred group	0.21	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{60-69.9}$ =221, N_{total}=1066 (Ingelheim, 2014[97])
	50-59.9 FVC%pred group	0.13	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{50-59.9}$ =136, N_{total}=1066 (Ingelheim, 2014[97])
	49-49.9 FVC%pred group	0.01	Fixed	NDB phase III INPULSIS trial post-hoc analysis; $N_{40-49.9}$ =6, N_{total}=1066 (Ingelheim, 2014[97])
Efficacy probabilities (Overall survival, acute exacerbation, loss of lung function)				
Baseline risks (PBO arm)				
Efficacy – baseline risks (PBO)	Overall survival	Time dependant (see section 5.3)	N/A	NDB phase II and phase III trials post-hoc analysis(Ingelheim, 2014[97])
	Acute exacerbations	(see section 5.3)	N/A	NDB phase III trial post-hoc analysis (Ingelheim, 2014[97])

Category	Variable	Value	Variance	Source/assumption
	Loss of lung function	(see section 5.3)	N/A	NDB phase III trial post-hoc analysis (Ingelheim, 2014[97])
Relative treatment effect for NDB and PFN				
Overall survival (OR)	NDB vs. PBO	0.70	0.45 to 1.10 (CI)	Scenario 1, NMA OS results, fixed effects
	PFN vs. PBO	0.70	0.46 to 1.05 (CI)	Scenario 1, NMA OS results, fixed effects
Acute exacerbations (OR)	NDB vs. PBO	0.56	0.35 to 0.89 (CI)	Scenario 3, NMA exacerbation results, fixed effects
	PFN vs. PBO	1.01	0.22 to 4.50 (CI)	Scenario 3, NMA exacerbation results, fixed effects
Loss of lung function (OR)	NDB vs. PBO	0.54	0.42 to 0.69 (CI)	Scenario 2, NMA progression results, fixed effects
	PFN vs. PBO	0.69	0.47 to 1.00 (CI)	Scenario 2, NMA progression results, fixed effects
Safety probabilities				
Baseline risks (PBO arm)				
Safety- baseline risks (PBO)	Serious cardiac event	1.39%	0.97% to 1.81% (CI)	Richeldi et al. 2014; $N_{PBO,\ serious\ cardiac\ events}$=23, $N_{PBO,total}$=423. Assume 30% variance around mean value. (Richeldi, 2014[4])
	Serious GI event	0.42%	0.29% to 0.54% (CI)	Richeldi et al. 2014; $N_{PBO,\ serious\ GI\ events}$=7, $N_{PBO,total}$=423. Assume 30% variance around mean value. (Richeldi, 2014[4])
Relative treatment effect for NDB and PFN				
Serious cardiac events (OR)	NDB vs. PBO	0.92	0.53 to 1.63 (CI)	Scenario 2, serious cardiac events, NMA results, excluding Richeldi 2011, fixed effects (Richeldi *et al.*, 2011[1]),
	PFN vs. PBO	1.27	0.66 to 2.49 (CI)	Scenario 2, serious cardiac events, NMA results, excluding Richeldi 2011, fixed effects (Richeldi *et al.*, 2011[1]),
Serious GI events (OR) 2.32% 0.32% 6.79% 2.19%	NDB vs. PBO	2.35	1.05 to 5.88 (CI)	Scenario 1, serious GI events, base-case analysis, NMA results - all evidence, fixed effects
	PFN vs. PBO	0.60	0.23 to 1.45 (CI)	Scenario 1, serious GI events, base-case analysis, NMA results - all evidence, fixed effects
Clinically important AEs				
Photosensitivity reaction	PBO arm	0.32%	0.22% to 0.41% (CI)	NICE STA 2011, p.156; $N_{PBO,\ photosensitivity\ reaction}$=6, $N_{PBO,total}$=347. Assume 30% variance around mean value.(NICE, 2013[17])
	PFN arm	2.32%	1.63% to 3.02% (CI)	NICE STA 2011, p.156; $N_{PFN,\ photosensitivity\ reaction}$=42, $N_{PFN,total}$=345. Assume 30% variance around mean value. (NICE, 2013[17])
Rash	PBO arm	2.19%	1.54% to 2.85% (CI)	NICE STA 2011, p.156; $N_{PBO,\ rash}$=40, $N_{PBO,total}$=347. Assume 30% variance around mean value. (NICE, 2013[17])

Category	Variable	Value	Variance	Source/assumption
	PFN arm	6.79%	4.75% to 8.82% (CI)	NICE STA 2011, p.156; $N_{PFN, rash}$=111, $N_{PFN,total}$=345. Assume 30% variance around mean value. (NICE, 2013[17])
GI perforation	PBO arm	0%	N/A	FDA NDB prescribing information 2014. (FDA highlights of prescribing information, 2014[96])
	NDB arm	0.08%	0.05% to 0.10% (CI)	FDA NDB prescribing information 2014, 0.03% from NDB pooled population (phase II and phase III trial data). Assume 30% variance around mean value. (FDA highlights of prescribing information, 2014[96])
Discontinuation probabilities				
Baseline risk (PBO arm)				
Overall discontinuation – baseline risk (PBO)	Probability PBO arm	5.38%	4.49% to 6.43%	NDB phase II and phase III trials post-hoc analysis. Based on an exponential parametric model.
Relative treatment effect for NDB and PFN				
Overall discontinuation (OR)	NDB vs. PBO	1.42	1.08 to 1.87 (CI)	Scenario 1, discontinuation (overall), base-case analysis, NMA results, all evidence, fixed effects
	PFN vs. PBO	1.34	1.04 to 1.74 (CI)	Scenario 1, discontinuation (overall), base-case analysis, NMA results, all evidence, fixed effects
Resource use				
NDB Phase III trial post-hoc HCRU analysis Resource use intensity	Frequency of liver function test (per cycle)	0.33	0.17 to 0.50 (CI)	Assumed patients receive quarterly tests as per the PFN SPC, (LLCI assume every 6 months, HLCI assume every 2 months).
	Number of hospitalisations	1.12	SE=0.03	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Average duration of each hospitalisation	8.72	SE=0.69	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Proportion in ICU	0.04	SE=0.01	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Proportion with mechanical ventilation	0.01	SE=0.01	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Proportion associated with ER overnight	0.05	SE=0.02	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Proportion associated with ambulance use (hospital)	0.13	SE=0.02	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Average number of ER visits	1.09	SE=0.04	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Proportion associated with ambulance use (ER)	0.17	SE=0.03	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of GP visits	1.59	SE=0.05	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of specialist visits	1.82	SE=0.05	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])

Category	Variable	Value	Variance	Source/assumption
	Number of nurse visits	1.63	SE=0.19	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of physiotherapist visits	3.51	SE=0.63	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of other visits	1.92	SE=0.17	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of chest HRCT	1.10	SE=0.02	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of chest x-ray	1.57	SE=0.09	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of O_2 requirement assessment	1.32	SE=0.14	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of bronchoalveolar lavage	1.00	SE=0.50	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of CT - pulmonary angiogram	1.18	SE=0.18	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of right heart catheterization	1.08	SE=0.06	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Number of general diagnostic	1.58	SE=0.06	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Exacerbations - Number of hospitalisations	1.30	SE=0.18	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
	Exacerbations - Average duration of each hospitalisation	16.30	SE=4.72	NDB Phase III trial post-hoc HCRU analysis (Ingelheim, 2014[97])
Unit costs				
Treatment cost	BSC daily cost	£0.00	Fixed	Assumed patients on BSC do not receive drug therapy.
	NDB daily cost	£71.70	£68.12 to £75.29 (CI)	Assumed equivalent to the cost of PFN.
	NDB daily cost (discounted)	[redacted]	[redacted]	Drug price. Assume 5% variance around mean value.
	PFN daily cost	£71.70	£68.12 to £75.29 (CI)	Drug price obtained from MIMS March-May 2014, printed version; dosage obtained from EMC 2013. Assume 5% variance around mean value.
Liver function test	Liver function test cost	£3.01	£1.68 to £3.97 (CI)	NHS reference cost 2012/13 NHS Trusts and PCTs combined Direct Access: Pathology Services: DAPS05 Haematology. Respiratory Medicine (Department of Health, 2013[8])
Adverse events cost	Serious cardiac event cost	£2,054.18	£1,318.39 to £2,521.95 (CI)	NHS Reference cost 2012/13, non-elective inpatients, long stay, weighted average of Actual or Suspected Myocardial Infarction with CC Score 4-13, Code 340, Respiratory medicine(Department of Health, 2013[8])
	Serious GI event cost	£1,749.45	£975.95 to £2,298.86 (CI)	NHS Reference cost 2012/13, non-elective inpatients, long stay, weighted average of Inflammatory Bowel Disease, without Interventions, with CC Score 3-4 and 5+,

Category	Variable	Value	Variance	Source/assumption
				service code 340, Respiratory medicine (Department of Health, 2013[8])
	Skin disorder cost	£373.07	£313.24 to £412.78 (CI)	NHS Reference cost 2012/13, non-Elective Inpatients (Short Stay) of Rash or Other Non-specific Skin Eruption, Code 340, Respiratory medicine (Department of Health, 2013[8])
	GI perforation cost	£2,352.99	£1,336.79 to £3,142.00 (CI)	NHS Reference cost 2012/13, non-elective inpatients, long stay, weighted average of Gastrointestinal Bleed with CC Score 5-9+, Code 340, Respiratory medicine (Department of Health, 2013[8])
Resource use cost	Hospitalisations (per day)	£303.73	Fixed	Weighted average of DZ27D (Respiratory Failure without Intubation with Major CC), DZ27E (Respiratory Failure without Intubation with CC), DZ27F (Respiratory Failure without Intubation without CC); Cost inflated to 2012/13 from National Schedule of Reference Costs Year : '2009-10' - NHS Trusts Elective Inpatient Excess Bed Day HRG Data (Department of Health, 2013[8])
	ICU (per day)	£829.81	Fixed	Weighted average of XC06Z (Adult Critical Care, 1 organ supported) and XC07Z (Adult Critical Care, 0 organs supported), Adult Critical Care Unit; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Critical Care (Department of Health, 2013[8])
	Mechanical ventilation (per day)	£147.52	Fixed	Weighted average of non-invasive ventilation support assessment 19 yrs and over - DZ37A, Respiratory medicine and Respiratory physiology; procedures in outpatients; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts (Department of Health, 2013[8])
	ER overnight stay	£154.31	Fixed	National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Accident and Emergency Services; Weighted average across all types, all treatment categories; admitted; non-dental (Department of Health, 2013[8])
	Ambulance use	£204.53	Fixed	Weighted average of ASH1 (hear and treat or refer), ASS01 (see and treat or refer), ASS02 (see and treat and convey); National Schedule of Reference Costs Year : 2012-13 - All NHS trust and NHS foundation trusts - ambulance services (Department of Health, 2013[8])
	ER visit	£101.16	Fixed	National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Accident and Emergency Services; Weighted average across all types, all treatment categories; not admitted; non-dental (Department of Health, 2013[8])
	GP visit	£37.50	Fixed	Per patient contact lasting 11.7 minutes; General practitioner — unit costs; PSSRU 2013, p. 198
	Specialist visit	£119.00	Fixed	Consultant: medical; £99 (139) per contract hour; PSSRU 2013, p. 252
	Nurse visit	£74.97	Fixed	Specialist Nursing - Asthma and Respiratory Nursing/Liaison, Adult, Face to face; N08AF; Community Health Services - Nursing; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts (Department of Health, 2013[8])
	Physiotherapist visit	£50.44	Fixed	Physiotherapist, Adult, One to One; A08A1, National Schedule of Reference Costs

Category	Variable	Value	Variance	Source/assumption
				- Year 2012-13 - NHS trusts and NHS foundation trusts; Community Health Services - Allied Health Professionals (Department of Health, 2013[8])
	Other visit	£169.99	Fixed	PSSRU 2010 p.31, costs from 2009/10, inflated to 2012/13; assume weighted average (£157.94) is the mean, general surgery is the low limit (£96) and mental health- elderly is the high limit (£224)
	Chest HRCT	£50.03	Fixed	Weighted average of RA08A (CT scan, one area, no contrast, 19 yrs and over) and RA09A (CT scan, one area, with post contrast only, 19 yrs and over); Diagnostic imaging - direct access; respiratory medicine ; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation (Department of Health, 2013[8])
	Chest x-ray	£28.28	Fixed	DAPF; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Directly Accessed Diagnostic Services (Department of Health, 2013[8])
	O_2 requirement assessment	£173.11	Fixed	Weighted average of DZ38Z oxygen assessment and monitoring, Respiratory medicine and Respiratory physiology; Procedures in Outpatients; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts (Department of Health, 2013[8])
	Bronchoalveolar lavage	£105.43	Fixed	Minor Mouth or Throat Procedures, 19 years and over without CC; CZ01Y; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Directly Accessed Diagnostic Services (Department of Health, 2013[8])
	CT - pulmonary angiogram	£50.03	Fixed	Assume same cost as Chest HRCT
	Right heart catheterization	£54.35	Fixed	Minor Cardiac Procedures, with CC Score 0-1; EA44B; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Directly Accessed Diagnostic Services (Department of Health, 2013[8])
	General diagnostic	£105.43	Fixed	Assume same as Bronchoalveolar lavage. Minor Mouth or Throat Procedures, 19 years and over without CC; CZ01Y; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Directly Accessed Diagnostic Services (Department of Health, 2013[8])
Baseline health state cost	≥110 FVC%pred group	£219.19	Fixed	Calculated using the INPULSIS HCRU and the unit costs for each resource.
	100-109.9 FVC%pred group	£209.37	Fixed	
	90-99.9 FVC%pred group	£236.57	Fixed	
	80-89.9 FVC%pred group	£210.37	Fixed	
	70-79.9 FVC%pred group	£253.49	Fixed	
	60-69.9 FVC%pred group	£233.51	Fixed	
	50-59.9 FVC%pred group	£312.53	Fixed	

Category	Variable	Value	Variance	Source/assumption
	49-49.9 FVC%pred group	£649.17	Fixed	
Oxygen use cost	Oxygen use – cost per cycle (applied to FVC%pred 40-79.9 groups)	£417.93	£208.96 to £626.89 (CI)	Weighted average of DZ27D (Respiratory Failure without Intubation with Major CC), DZ27E (Respiratory Failure without Intubation with CC), DZ27F (Respiratory Failure without Intubation without CC); Cost inflated to 2012/13 from National Schedule of Reference Costs Year : '2009-10' - NHS Trusts Elective Inpatient Excess Bed Day HRG Data (Department of Health, 2013[8])
Exacerbation cost – acute event	Exacerbation cost – acute event	£4,133.59	Fixed	Comprises of hospital, ER, GP and specialist cost at specific frequencies (INPULSIS trial analysis)
EoL cost	End of life (EoL) cost	£3,920.64	£2,744 to £5,097 (CI)	The potential cost savings of greater use of home- and hospice-based end of life care in England. Evi Hatziandreu, Fragiskos Archontakis, Andrew Daly in conjunction with the National Audit Office (2008), p 56. Cost inflated to 2012/13 cost. Assume 30% variance around mean value. (Department of Health, 2013[8])
Utilities				
Baseline utility values	FVC%pred 110 and above	0.8380	Fixed	Assumption, same as FVC%pred 90 value
	FVC%pred 100	0.8380	Fixed	Assumption, same as FVC%pred 90 value
	FVC%pred 90	0.8380	SE=0.0083	NDB Phase III trial post-hoc analysis; SD=0.1782, N=458 (Ingelheim, 2014[97])
	FVC%pred 80	0.8105	SE=0.0078	NDB Phase III trial post-hoc analysis; SD=0.2051, N=684 (Ingelheim, 2014[97])
	FVC%pred 70	0.7800	SE=0.0080	NDB Phase III trial post-hoc analysis; SD=0.2244, N=788 (Ingelheim, 2014[97])
	FVC%pred 60	0.7657	SE=0.0084	NDB Phase III trial post-hoc analysis; SD=0.2380, N=809 (Ingelheim, 2014[97])
	FVC%pred 50	0.7387	SE=0.0105	NDB Phase III trial post-hoc analysis; SD=0.2317, N=490 (Ingelheim, 2014[97])
	FVC%pred 40	0.6634	SE=0.0258	NDB Phase III trial post-hoc analysis; SD=0.2552, N=98 (Ingelheim, 2014[97])
	FVC%pred 30	0	Fixed	Assumption
Investigator exacerbations disutility	1st month	-0.1400	-0.2310 to -0.0490 (CI)	NDB Phase III trial post-hoc analysis (Ingelheim, 2014[97])
	subsequent month	-0.0780	-0.1420 to -0.0150 (CI)	NDB Phase III trial post-hoc analysis (Ingelheim, 2014[97])
Adjudication committee exacerbations disutility	1st month	-0.2740	-0.3900 to -0.1570 (CI)	NDB Phase III trial post-hoc analysis (Ingelheim, 2014[97])
	subsequent month	-0.0330	-0.1370 to 0.0700 (CI)	NDB Phase III trial post-hoc analysis (Ingelheim, 2014[97])

Category	Variable	Value	Variance	Source/assumption
Treatment-related adverse events disutility	Serious cardiac event	-0.1980	-0.2574 to -0.1386 (CI)	Ara and Brazier (baseline utilities): HEDS Discussion Paper 10/11. Heart attack/angina: 0.628 (0.602, 0.653) for respondents affected by the health condition (and any other health condition); 0.826 (0.802, 0.850) respondents of similar age not affected by the health condition. Assume 30% variance around mean value. (Ara *et al.*, 2010[130])
	Serious GI event	-0.0680	-0.2010 to 0.0650 (CI)	NDB phase III trial post-hoc analysis of serious GI events (Ingelheim, 2014[97])
	Skin disorder	-0.0820	-0.1066 to -0.0574 (CI)	Ara and Brazier (baseline utilities): HEDS Discussion Paper 10/11. Skin complaints: 0.773 (0.733, 0.812) for respondents affected by the health condition (and any other health condition); 0.855 (0.834, 0.875) respondents of similar age not affected by the health condition. Assume 30% variance around mean value. (Ara *et al.*, 2010[130])
	GI perforation	-0.1180	-0.1534 to -0.0826 (CI)	Ara and Brazier (baseline utilities): HEDS Discussion Paper 10/11.Stomach ulcer/abdominal hernia/rupture: 0.688 (0.654, 0.720) for respondents affected by the health condition (and any other health condition); 0.806 (0.781, 0.830) respondents of similar age not affected by the health condition. Assume 30% variance around mean value. (Ara *et al.*, 2010[130])

Table 160: Key model assumptions

Model input	Assumption	Source / rationale
Cycle length	The model cycle length was assumed to be 3 months.	The cycle length was selected to be consistent with the clinical trial intervals between observations and was considered a balanced interval for the model outcomes in discussion with clinical experts (Boehringer Ingelheim, 2014[91])
BSC model inputs	Efficacy and safety were assumed to be represented by the events observed in control (PBO) arm of the (phase III and phase II) NDB clinical trials.	It was assumed that both strategies shared similar background treatments. An analysis on concomitant medications showed that there was a small difference in the management of patients between the two NDB clinical trial arms (PBO and NDB). Since the perspective of the economic evaluation is the incremental cost-effectiveness of NDB vs. BSC, it was assumed that the efficacy and safety of BSC is reflected by the observed outcomes of the PBO arm of the trial.
Survival analysis implementation	Survival analysis extrapolation was assumed to be applied for the full duration of the economic model; that also included the first year of the analysis where clinical trial data were available.	This allowed a more robust representation of uncertainty from the trial results and a formal exploration via a probabilistic sensitivity analysis.
Baseline mortality risk (PBO arm)	It was assumed that death could occur at the point that patients reached a level of FVC%pred of 30-39.9%.	A discussion with clinical experts held regarding the point at which life becomes unsustainable with low FVC%Pred values [ref ad board]. A similar assumption was made in the NICE IPF Clinical Guideline (CG163) for a threshold of 35% FVC%pred [(NICE, 2013[18]), Apdx L]. To fit with the 10-percentage-points categories 40% was used. The impact of this to the incremental cost-effectiveness results was tested in sensitivity analysis and it was minimal.
	Risk of death was modelled independently of other outcomes (exacerbations, progression).	The following interactions and correlations between parameters were tested: a) Treatment-related mortality, treatment-related progression, and treatment-related exacerbation analysed independently. b) Treatment-related exacerbation and treatment-related progression were analysed independently; overall mortality was dependent on exacerbation only. c) Treatment-related exacerbation and treatment-related progression were analysed independently; overall mortality was dependent on progression only. d) Treatment-related exacerbation and treatment-related progression were analysed independently; overall mortality was dependent on exacerbation and progression. The final economic model included case (a). Case (b) was not selected because after synthesis of the probabilities of exacerbation and mortality, the final mortality risk was

Model input	Assumption	Source / rationale
		illogical (i.e. exacerbations were linked to better survival) therefore the analysis was not pursued further. For case (c) results were not statistically significant and no link was found between overall survival and progression. It was reasoned that either there were not enough data points for progression, or a 10-point decrease in lung function was not enough to influence survival. As a result, this dependency was not pursued any further. The last scenario (d) was problematic due to two reasons: 1. Statistically, the same problems as with the progression were encountered-dependent overall survival. 2. Including this link in the economic model would have rendered it too complicated, and the objective was to keep it simple, transparent, and user-friendly. It would have been too complex to follow patients' survival through health states as they experienced exacerbations. This analysis would have required tunnels for re-setting time and switching to a different survival equation to account for a different survival function when experiencing exacerbations.
Definition of baseline disease progression / loss of lung function (PBO arm)	Baseline disease progression was defined as a 10-point drop in FVC%Pred every three months (constant risk).	According to many studies the MCID for FVC%pred ranged between a 2-6% change (du Bois *et al.*, 2011[7]), or a 10% change (du Bois *et al.*, 2011[7]) (Raghu *et al.*, 2011[13]). After consultation with the clinical experts [ad board] and consideration of comparable evidence for the other model comparators (PFN and NAC), it was decided that a 10-point categorization of FVC%pred was a balanced range to capture granularity of outcomes without overcomplicating the model. The choice of a 3-months cycle length was consistent with the clinical trial intervals duration and the cycle length. The impact of the length of this interval was not assessed. If the interval was shorter or longer, the loss of lung function for the cohort could be underestimated or overestimated.
Progression / loss of lung function	It was assumed that once progressed to a lower FVC%pred the cohort could not regress back to health states with improved lung function (higher FVC%pred).	Similar assumptions were made in the NICE model on PFN (NICE, 2013[17]). Clinical expert opinion validated the assumptions (Boehringer Ingelheim, 2014[91]).
Definition of baseline exacerbation risk (PBO arm)	The model used the investigator reported exacerbations as base-case, and explored the effect of the adjudication committee defined exacerbations in sensitivity analysis.	There were more exacerbation events for the investigator than for the adjudication-committee definition (63 vs. 36). Moreover, it was considered more appropriate to use the investigator definition since it reflected more closely "real-world" clinical practice. Furthermore, most of the clinical trials (used in the NMA) did not capture adjudication-committee definitions. Finally, the use of

Model input	Assumption	Source / rationale
		investigator definition was more conservative for NDB treatment; the adjudication-committee definition showed larger difference between NDB and control.
Exacerbation risk	Exacerbation was assumed to be a constant hazard every three months (exponential model).	Several parametric models were considered based on INPULSIS trial data. Considering the AIC values and model parsimony, the exponential model was selected.
Recurrent exacerbation risk	Patients who experienced at least one exacerbation event were at risk of recurrent events, which was assumed to be the same risk as for patients that have not had an exacerbation.	It was considered that this is a reasonable assumption given the lack of other relevant evidence on the incidence of recurrent exacerbation events. The assumption (inclusion/exclusion of recurrent events) was tested in OWSA and had a minimal effect on the cost-effectiveness results.
Effect of exacerbations on mortality and progression	A link between exacerbation and mortality was not included in the model. It was assumed that exacerbation had no effect on loss of lung function in the base-case analysis.	The effect of exacerbations on mortality was previously explored, but illogical results were obtained [see the explanation behind the risk of death being modelled independently of other outcomes (exacerbations, progression) above]. The effect of exacerbations on loss of lung function was explored in sensitivity analysis, with minimal impact on the model results.
Baseline discontinuation risk (PBO arm)	Baseline discontinuation (PBO risk) was assumed to be a constant hazard every three months (exponential model).	Several parametric models were considered based on INPULSIS trial data. Considering the AIC values and model parsimony, the exponential model was selected.
IPD analysis on discontinuation to separate from death events	In INPULSIS, in some cases it was impossible to separate discontinuation due to death from actual treatment discontinuation; it was assumed that an event counted as death rather than study discontinuation if the date of discontinuation coincided with the date of death or if it was the very next day.	Simplifying assumption.
Use of clinical data for discontinuation	The model used clinical trial data for time-to-discontinuation for the NDB and PFN arms (rather than real-world data).	Since NDB and PFN are new treatments for idiopathic pulmonary fibrosis (IPF) patients, there is a lack of evidence on treatment tolerability and discontinuation in real life. In the model this was based on rates observed within the investigation trials. This is likely to underestimate “real-world” discontinuation. This was addressed with sensitivity analysis in the comparison NDB vs. PFN; there was a minimal impact on the model results.
Discontinuation for the BSC arm	Assumed no discontinuation from BSC.	Simplifying assumption- no other treatment to cycle to.
Applying NMA OR values to the baseline (PBO) risk – relative treatment effects (NDB and PFN arms)	The relative effects of NDB and PFN were assumed to be informed using NMA OR values.	In the absence of direct clinical trial evidence a network meta-analysis (NMA) was designed, using a Bayesian framework, to assess the efficacy and safety of NDB relative to its comparators. The NMA originally included N-Acetylcysteine (NAC) treatment. The strength of the NMA was that it included more trials than only INPULSIS and CAPACITY

Model input	Assumption	Source / rationale
		for NDB and PFN respectively. For consistency with the evidence generated in the NMA, and since the inclusion of NAC does not have an impact on the network for NDB, PFN and PBO, we used part of those NMA results (excluding NAC) in our submission; note that data were reported as odds-ratios (ORs) vs. PBO. Although direct evidence for NDB vs. PBO was available in the NDB clinical trials, the relative effects of the NMA were used instead to be consistent with PFN and NAC. Direct data between NDB and PBO were tested as a sensitivity analysis scenario. The direct evidence for mortality was isolated as having the highest impact on the model results.
	The same relative effects (OR) was applied to the baseline risks, independent of time.	There was a lack of information to explore the analysis of different ORs over time or other time-dependencies.
	Three-month estimates of baseline risk were synthesised in the model with approximately 1-year estimates of relative efficacy from the clinical trials. In effect the analysis assumed that the relative difference we observed across the comparators at the end of the trial, was constant and would hold for the intermediate intervals (3 months).	This was consistent with the assumptions made regarding a constant relationship of relative effects over time.
Adverse events	AEs with a significant impact on costs and QALYs were assumed to be those that were severe or serious.	Comparability with the evidence on safety of the other comparators (PFN and NAC) was considered; PANTHER did not report "any" AE. Limiting the events to severe and serious provided with a better basis for the EQ-5D analysis (more consistent in terms of the impact to the patient).
Liver enzyme elevations	These events were assumed to be asymptomatic for patients. The model assumed that when these events were detected (with appropriate liver function tests), they contributed only to the overall discontinuation from treatment, and that there was no disutility or additional costs associated with them.	Simplifying assumption; in clinical practice patients would discontinue treatment – an outcome analysed separately in the model.
Risk of bleeding	The risk of bleeding was not included in the cost-effectiveness model.	Because all NDB clinical trials excluded patients with high risk of bleeding, we could not quantify and assess this risk for further analysis.
BSC daily treatment cost	No treatment cost was assumed for the BSC arm.	An analysis on concomitant medications on INPULSIS showed a small difference between trial arms (PBO and NDB).
PFN daily dosage	The PFN dosage was assumed to be 2,403mg/day	eMC 2014
Liver function test frequency	The model assumed that all patients on active treatment would incur the cost of liver function test, at a quarterly frequency (every 3 months, i.e. every cycle). For sampling, it was assumed every 6 months for LLCI, and every 2	Frequency schedule was the same with the maintenance test frequency recommended by the PFN SPC [eMC 2014]

Model input	Assumption	Source / rationale
	months for HLCI.	
Cost of ER visit	Unit cost was assumed to be £101.16 per visit.	Weighted average across all types of A&E services, all treatment categories; filtered by not admitted and non-dental; National Schedule of Reference Costs - Year 2012-13 - NHS trusts and NHS foundation trusts; Accident and Emergency Services (Department of Health, 2013[8])
Cost of ambulance use	Unit cost was assumed to be £204.53 per use.	Weighted average of ASH1 (hear and treat or refer), ASS01 (see and treat or refer), ASS02 (see and treat and convey); National Schedule of Reference Costs Year: 2012-13 - All NHS trust and NHS foundation trusts - ambulance services (Department of Health, 2013[8])
Cost of GP visit	Unit cost was assumed to be £37.50 per visit.	PSSRU 2013, per patient contact visit lasting 11.7 minutes
Cost of a specialist visit	Unit cost was assumed to be £119.0 per visit.	PSSRU 2013, average of medical consultant per contract hour, with (£139) and without (£99) qualifications
Cost of a nurse visit	Unit cost was assumed to be £74.97 per visit.	National Schedule of Reference Costs 2012-13, specialist nursing - asthma and respiratory nursing/liaison, adult, face to face; N08AF; Community Health Services – Nursing (Department of Health, 2013[8])
Cost of a physiotherapist visit	Unit cost was assumed to be £50.44 per visit.	National Schedule of Reference Costs 2012-13, physiotherapist, adult, one to one; A08A1, Community Health Services - Allied Health Professionals (Department of Health, 2013[8])
Cost of “other” visit	Unit cost was assumed to be £169.99 per visit.	PSSRU 2010: breakdown of inpatient attendance: per bed-day; cost inflated from 2009/10 to 2012/13
Cost of chest HRCT	Unit cost was assumed to be £50.03 per procedure.	National Schedule of Reference Costs 2012-13: weighted average of RA08A (CT scan, one area, no contrast, 19 yrs and over) and RA09A (CT scan, one area, with post contrast only, 19 yrs and over); Diagnostic imaging - direct access; respiratory medicine (Department of Health, 2013[8])
Cost of chest X-ray	Unit cost was assumed to be £28.28 per procedure.	National Schedule of Reference Costs 2012-13: DAPF; Directly Accessed Diagnostic Services (Department of Health, 2013[8])
Cost of an oxygen requirement assessment	Unit cost was assumed to be £173.11 per assessment.	National Schedule of Reference Costs 2012-13: weighted average of DZ38Z oxygen assessment and monitoring in Respiratory Medicine and Respiratory Physiology; procedures in outpatients (Department of Health, 2013[8])
Cost of bronchoalveolar lavage	Unit cost was assumed £105.43 procedure.	National Schedule of Reference Costs 2012-13: Minor Mouth or Throat Procedures, 19 years and over without CC (CZ01Y), Directly Accessed Diagnostic Services (Department of Health, 2013[8])
	It was assumed that SE =0.5 (rather than 0 as obtained in the INPULSIS HCRU analysis)	Sampling error in Excel with SE=0.
Cost of CT – pulmonary angiogram	Unit cost was assumed to be same cost as chest HRCT (£50.03 per procedure).	No alternative unit cost was identified.
Cost of right heart	Unit cost was assumed to be	National Schedule of Reference Costs 2012-13:

Model input	Assumption	Source / rationale
catheterization procedure	£54.35 per procedure.	Minor Cardiac Procedures, with CC Score 0-1; EA44B, Directly Accessed Diagnostic Services (Department of Health, 2013[8])
Cost of general diagnostic (bronchoscopy, sputum examination, surgical lung biopsy, transbronchial biopsy, other)	Unit cost was assumed to be same cost as bonchoalveolar lavage (£105.43 per procedure)	Simplification assumption in order to include this category of costs.
Oxygen use	It was assumed that patients with an FVC%Pred of 80% and over were in a relatively good health state. It was further assumed that patients who dropped below a level of 80% FVC%Pred would require oxygen supplementation.	Simplification assumption due to lack of alternative guidance on this parameter.
	Unit cost was assumed to be £418 per 3-month cycle.	NHS Reference Costs 2010/11, cycle-length adjusted from £1,600 per year, value inflated to 2012/13 cost (Department of Health, 2013[8])
	It was assumed that HLCI=mean+LLCI for sampling distribution.	No upper limit precision parameter was reported
End of life	It was assumed that patients received palliative care (in addition to background health care resources) as they reached the end of their life. The model applied an end of life (EoL) cost for the last year of patients' life.	Clinical experts advised that palliative care is an important aspect of people's end of life care [Ad board]. Since this is a lifetime model, it does drive the incremental CE results.
	It was assumed that the cost of EoL was formed of hospice and home care (excluding hospital) and it was estimated at £3,920.64 per cycle (£8 for hospice care, £9,098 for home care).	Hatziandreu et al. 2008, cost was cycle length adjusted, inflated to 2012/13 values.
	Assumed 30% variation around mean value for sampling distribution.	No precision parameters were reported.
Use of clinical trial EQ-5D and HCRU data	The correlation of lung status and patient condition (health state) with HRQL (in the form of EQ-5D) and resource use was based on INPULSIS post-hoc analyses. The analysis assumed that the results of the clinical trial in terms of EQ-5D and resource use are generalisable for the UK population.	This was the only available evidence to perform such an analysis for IPF patients. All other known HRQL analyses were based on mapping from other instruments.
Baseline EQ-5D value for FVC%Pred ≥110 and 100-109.9	Assumed the same utility value as for FVC90%pred (0.8380).	This assumption was made because the values for FVC%pred ≥90% were all around an EQ-5D value of 0.84.
Baseline EQ-5D value for FVC%Pred 30-39.9	Assumed the utility is 0 (dead).	Common assumption.

Model input	Assumption	Source / rationale
Exacerbation-related disutility values	Exacerbations were assumed to be acute events that affect the health state of the patients. It was assumed that patients experienced an acute phase in the 1st month and a post-acute phase (in the following 2+ months), following an exacerbation.	This assumption was supported by the analysis of INPULSIS EQ-5D data [INPULSIS post-hoc analysis]. The analysis of INPULSIS data showed the following trend in disutility following an acute exacerbation: • There was a severe change in the 1st month after the exacerbation, followed by a recovery and smaller and longer-lasting disability in the subsequent months. Note that the data were too limited to permit the analysis of 2, 3, or 4 months decrements [INPULSIS post-hoc analysis]. • Adjudicated exacerbations had a more severe impact than the investigator-reported ones. Values were assumed to be conservative estimates, as it was likely that the worst patients were not missing at random (as they were unable or unwilling to attend the next study visit).
Serious cardiac events related disutility value	The base-case model assumed the higher disutility value (-0.198) for the serious cardiac events.	Conservative scenario for NDB. The base-case value has a greater impact on NDB.
	Assumed 30% variation around mean value for sampling distribution.	No precision parameters were reported.
Serious GI events related disutility value	The model assumed the disutility of a serious GI event to be -0.068.	During a serious GI event, HRQL dropped by -0.068 points (-0.201 to 0.065) and then recovered [INPULSIS post-hoc analysis].
GI perforation related disutility value	The base-case model assumed the higher disutility value (-0.118) for the GI perforation.	Conservative scenario for NDB. The base-case value has a greater impact on NDB.
	Assumed 30% variation around mean value for sampling distribution.	No precision parameters were reported.
Skin disorder related disutility value	Assumed 30% variation around mean value for sampling distribution.	No precision parameters were reported.

5.7 *Base-case results*

5.7.1; 5.7.2 Base-case results and base-case incremental results

Results for NDB vs. PFN

Base-case deterministic results for NDB vs. PFN at NDB list price

The base-case deterministic results of NDB vs. PFN are presented in Table 161. The total costs for the NDB arm were £85,087, and £87,479 for the PFN arm. The breakdown of the total costs is reported in Table 161 below. With an incremental total cost of -£2,392 and incremental total QALYs of 0.0471, the NDB strategy dominates the PFN one. The incremental LYs gained were 0.0036 over the modelled time horizon. The

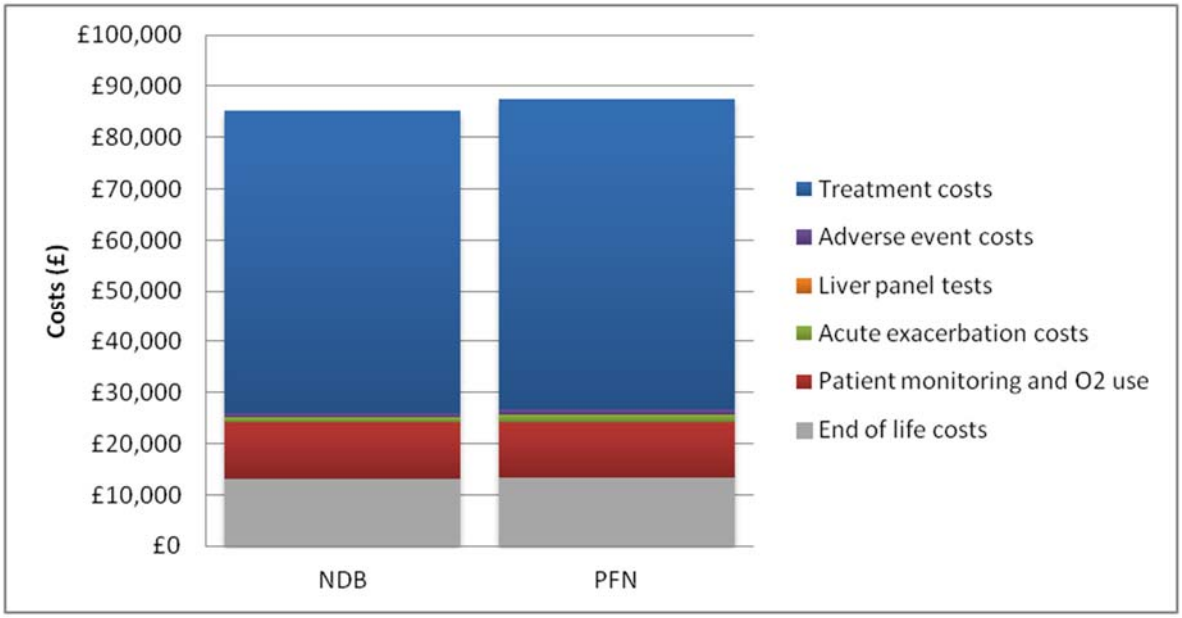
exacerbations avoided due to the NDB treatment were 0.075 over the duration of the model.

Table 161: Deterministic results for nintedanib vs. pirfenidone

	PFN	NDB	Incremental
Treatment costs	£60,707.68	£59,037.76	-£1,669.92
Adverse event costs	£1,054.92	£781.97	-£272.95
Liver panel tests	£9.31	£9.05	-£0.26
Patient monitoring and O2 use	£10,953.78	£10,825.15	-£128.64
Acute exacerbation costs	£1,501.72	£1,191.50	-£310.22
End of life costs	£13,251.74	£13,242.07	-£9.67
Total costs	£87,479.15	£85,087.49	-£2,391.65
Total QALYs	3.6249	3.6720	0.0471
LYs	4.8558	4.8593	0.0036
Exacerbation events	0.3633	0.2882	-0.0750
Net monetary benefit	£21,267.04	£25,071.71	
Cost-effectiveness	£24,133.02	£23,172.14	
ICER (per QALY)			NDB dominates
ICER (per LY)			NDB dominates
ICER (per exacerbation avoided)			NDB dominates

Figure 56 presents the analysis of total costs over the entire model duration.

Figure 56: Base-case total costs (nintedanib and pirfenidone)



Base-case deterministic results for NDB vs. PFN with NDB PAS

Table 162: Deterministic results for NDB vs. PFN

	PFN	NDB	Incremental
Treatment costs	£60,707.68	[redacted]	[redacted]
Adverse event costs	£1,054.92	[redacted]	[redacted]
Liver panel tests	£9.31	[redacted]	[redacted]
Patient monitoring and O2 use	£10,953.78	[redacted]	[redacted]
Acute exacerbation costs	£1,501.72	[redacted]	[redacted]
End of life costs	£13,251.74	[redacted]	[redacted]
Total costs	£87,479.15	£53,492.45	-£33,986.69
Total QALYs	3.6249	3.6720	0.0471
Lys	4.8558	4.8593	0.0036
Exacerbation events	0.3633	0.2882	-0.0750
Net monetary benefit	[redacted]	[redacted]	
Cost-effectiveness	[redacted]	[redacted]	
ICER (per QALY)			[redacted]
ICER (per LY)			[redacted]
ICER (per exacerbation avoided)			[redacted]

[redacted]

Results for NDB vs. BSC

Base-case deterministic analysis for NDB vs. BSC at NDB list price

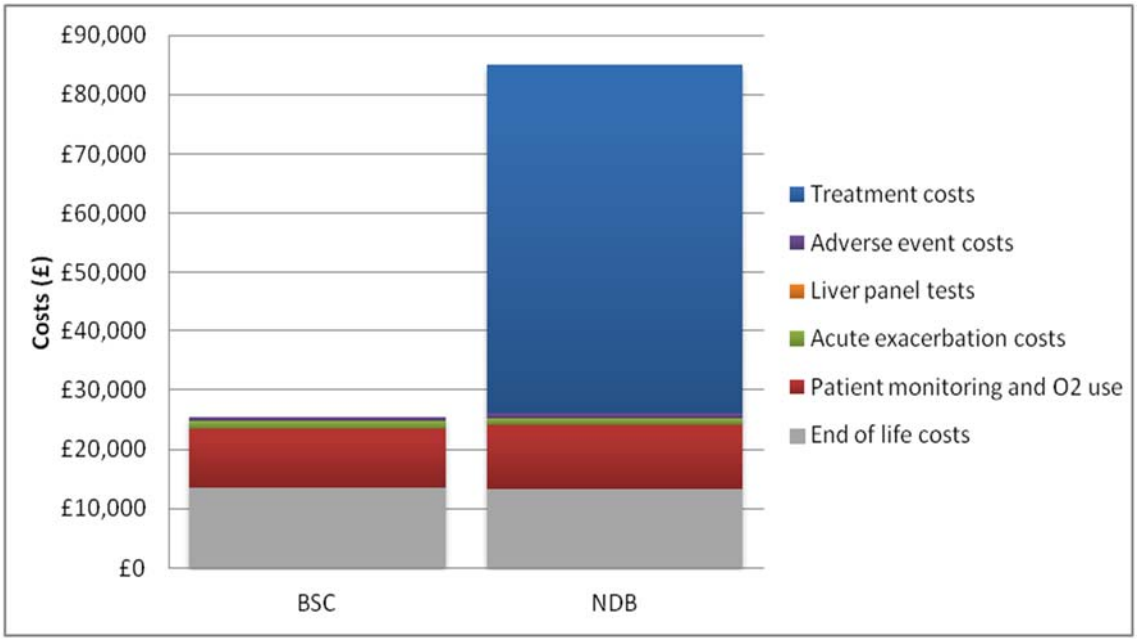
The base-case deterministic results of NDB vs. BSC are presented in Table 163. The total cost for the NDB treatment strategy was estimated to be £85,087, and £25,359 for the reference strategy (BSC). The breakdown of the total cost is reported in Table 163 below. With an incremental total cost of £59,728 and an estimated incremental QALY gained of 0.40, the estimated ICER for NDB vs. BSC was £149,361 per QALY gained. The incremental LYs gained with NDB were estimated to be 0.5 (6 months). The estimated number of acute exacerbation events avoided with NDB was 0.035.

Table 163: Deterministic results for nintedanib vs. BSC

	BSC	NDB	Incremental
Treatment costs	£0.00	£59,037.76	£59,037.76
Adverse event costs	£625.98	£781.97	£155.98
Liver panel tests	£0.00	£9.05	£9.05
Patient monitoring and O2 use	£9,955.76	£10,825.15	£869.38
Acute exacerbation costs	£1,336.83	£1,191.50	-£145.33
End of life costs	£13,440.51	£13,242.07	-£198.44
Total costs	**£25,359.09**	**£85,087.49**	**£59,728.40**
Total QALYs	**3.2721**	**3.6720**	**0.3999**
LYs	**4.3602**	**4.8593**	**0.4992**
Exacerbation events	**0.3234**	**0.2882**	**-0.0352**
Net monetary benefit	£72,803.32	£25,071.71	
Cost-effectiveness	£7,750.14	£23,172.14	
ICER (per QALY)			**£149,360.88**
ICER (per LY)			£119,656.46
ICER (per exacerbation avoided)			£1,698,840.58

Figure 57 presents the analysis of total costs over the entire model duration.

Figure 57: Base-case total costs (nintedanib and BSC)



Base-case deterministic analysis for NDB vs. BSC with NDB PAS

Table 164: Deterministic results for NDB vs. BSC

	BSC	NDB	Incremental
Treatment costs	£0.00	[redacted]	[redacted]
Adverse event costs	£625.98	[redacted]	[redacted]
Liver panel tests	£0.00	[redacted]	[redacted]
Patient monitoring and O2 use	£9,955.76	[redacted]	[redacted]
Acute exacerbation costs	£1,336.83	[redacted]	[redacted]
End of life costs	£13,440.51	[redacted]	[redacted]
Total costs	**£25,359.09**	**£53,492.45**	**£28,133.36**
Total QALYs	**3.2721**	**3.6720**	**0.3999**
LYs	4.3602	4.8593	0.4992
Exacerbation events	0.3234	0.2882	-0.0352
Net monetary benefit	£72,803.32	[redacted]	
Cost-effectiveness	£7,750.14	[redacted]	
ICER (per QALY)			[redacted]
ICER (per LY)			[redacted]
ICER (per exacerbation avoided)			[redacted]

[redacted]

Full incremental cost-effectiveness analysis at NDB list price

The full incremental cost-effectiveness analysis results are presented below (Table 165). The results are also presented graphically in Figure 58.

.

Treatments are ranked based on QoL benefit from the lowest to the highest. An ICER is estimated for each treatment compared to the previous most cost-effective and excluding strategies that are dominated (in this instance PFN).

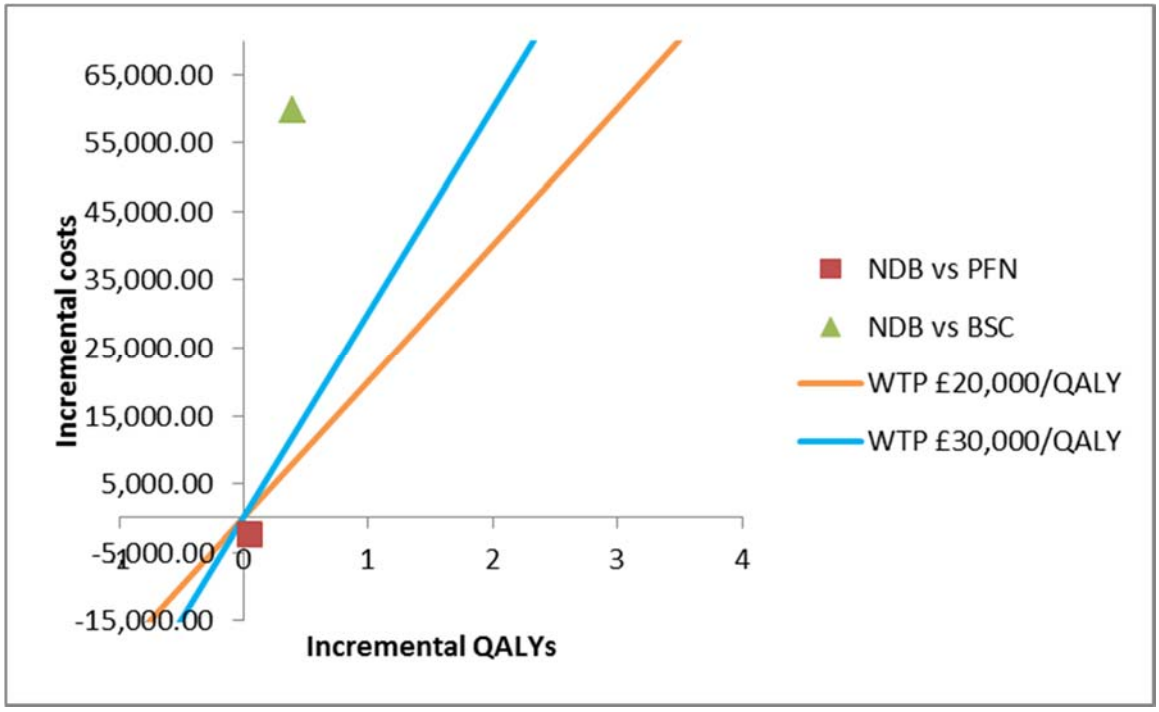
Table 165: Base-case results at nintedanib list price

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs. baseline (vs. BSC) (QALYs)	ICER (£) incremental (QALYs)
BSC	£25,359.09	4.3602	3.2721					

PFN	£87,479.15	4.8558	3.6249	62,120.05	0.4956	0.3528	£176,081	Dominated by NDB
NDB	£85,087.49	4.8593	3.6720	-2,391.66	0.0035	0.0471	£149,361	£149,361

* ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Figure 58: Scatter plot: incremental cost-effectiveness analysis NDB vs BSC and NDB vs PFN (without nintedanib PAS)



Full incremental cost-effectiveness analysis with NDB PAS

Table 166: Base-case results at nintedanib PAS price

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs. baseline (vs. BSC) (QALYs)	ICER (£) incremental (QALYs)
BSC	£25,359.09	4.3602	3.2721					
PFN	£87,479.15	4.8558	3.6249	62,120.05	0.4956	0.3528	£176,081	Dominated by NDB
NDB	[redacted]	4.8593	3.6720	[redacted]	0.0035	0.0471	[redacted]	[redacted]

*ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

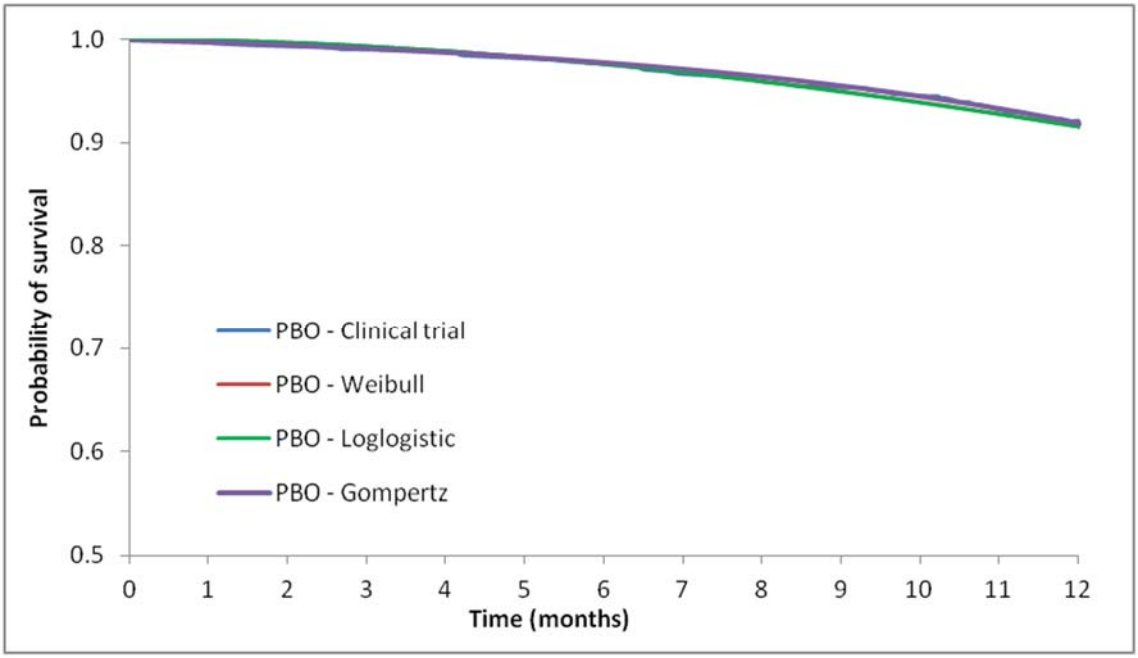
[redacted]

5.7.3 Clinical outcomes from the model

Validation of overall survival at the end of one year: model projections vs. clinical trial(s)

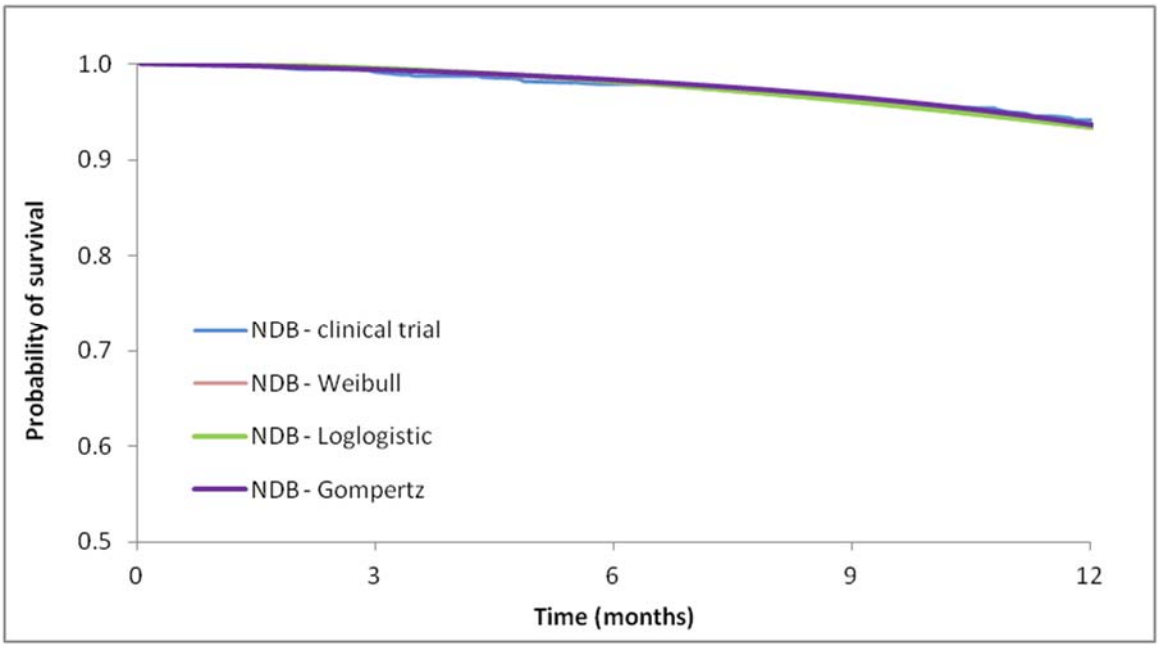
Figure 59 and Figure 60 compare the fit of the selected parametric models (Gompertz, loglogistic, and Weibull) to the Kaplan-Meier curves from the clinical trial (trial period: 52 weeks) for the PBO arm and NDB arm, respectively. Note that the clinical trial survival data for the PBO and NDB arms are based on the TOMORROW and INPULSIS clinical trials pooled data. The model projections closely match the clinical trial data for the first year of the model.

Figure 59. OS models fit vs. clinical trial KM – PBO arm



*Curves are overlapping and may not be entirely distinguishable.

Figure 60. OS models fit vs. clinical trial KM – NDB arm



*Curves are overlapping and may not be entirely distinguishable.

Figure 61 compares the fit of the selected parametric models (Gompertz, loglogistic, and Weibull) to the Kaplan-Meier curves from the clinical trial (trial period: 72 weeks) for the PFN arm. Note that the clinical trial survival data for the PFN arm was extracted using a digitation programme (Grafula) from a Kaplan-Meier curve (NICE, 2013[17]). The model projections match the clinical data for the duration of the clinical trial.

Figure 61: OS models fit vs. clinical trial KM - PFN arm

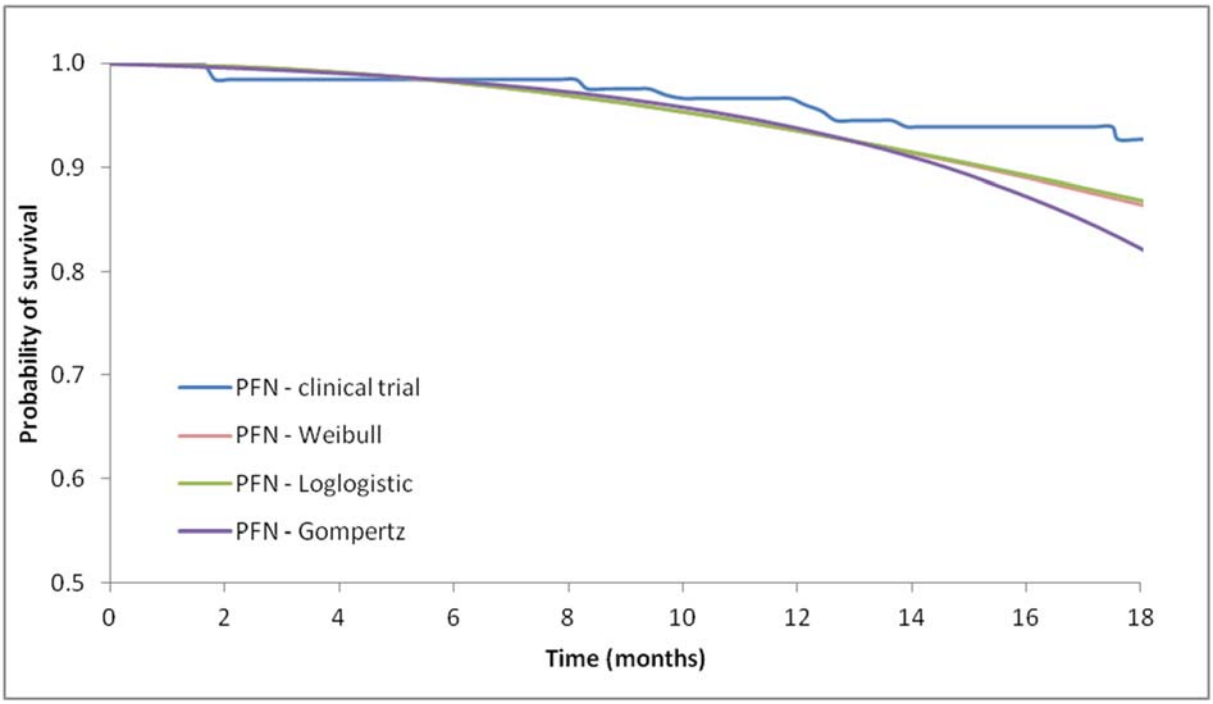


Table 167: Proportion of patients alive at the end of the first year

	Patients alive in the BSC arm	Patients alive in the NDB arm	Patients alive in the PFN arm
Clinical trial (KM)	0.921*	0.941*	0.965**
Loglogistic model	0.914	0.936	0.935
Weibull model	0.914	0.936	0.935
Gompertz model	0.917	0.938	0.938

*Source: (Boehringer Ingelheim, 2014[92])
**Source: (King, 2014[141])], mortality from any cause for the PFN arm, CAPACITY and ASCEND pooled; data from the two CAPACITY studies were censored at 1 year to standardize the follow-up for the three studies.

Validation of exacerbation at the end of one year: model projections vs. clinical trial(s)

Table 168 presents the proportion of patients that are alive and had an exacerbation at the end of the first year. Note that the events were very rare and estimates of exacerbations are most likely imprecise. Given this uncertainty, the model projections closely match the clinical trial data for the first year of the model.

Table 168: Proportion of patients alive with exacerbations at the end of the first year

	BSC arm	NDB arm	PFN arm
Clinical trial*	0.05	0.03	N/A
Model projections	0.07 (95% CI 0.05-0.10)	0.04 (95% CI 0.03-0.06)	0.07 (95% CI 0.05-0.10)

*Pooled TOMORROW and INPULSIS

Validation of FVC%Pred distribution at the end of one year: model projections vs. clinical trial(s)

The FVC%Pred distribution at the end of one year is presented in Table 169 and Figure 62 for the BSC arm and in Table 170 and Figure 63 for the NDB arm. The PFN projections are also presented below (compared to NDB). The model projections closely match the clinical trial data for the first year of the model.

Table 169. Proportion of patients alive and in each FVC%Pred category at the end of one year: BSC arm

FVC%Pred category	≥110	100-109.9	90-99.9	80-89.9	70-79.9	60-69.9	50-59.9	40-49.9
Model projections	3.72%	7.02%	11.41%	18.51%	20.85%	21.41%	14.59%	2.51%
Clinical trial (TOMORROW and INPULSIS)	4.93%	5.91%	8.62%	16.50%	18.72%	22.91%	16.26%	6.16%

Figure 62. Proportion of patients alive and in each FVC%Pred category at the end of one year: BSC arm

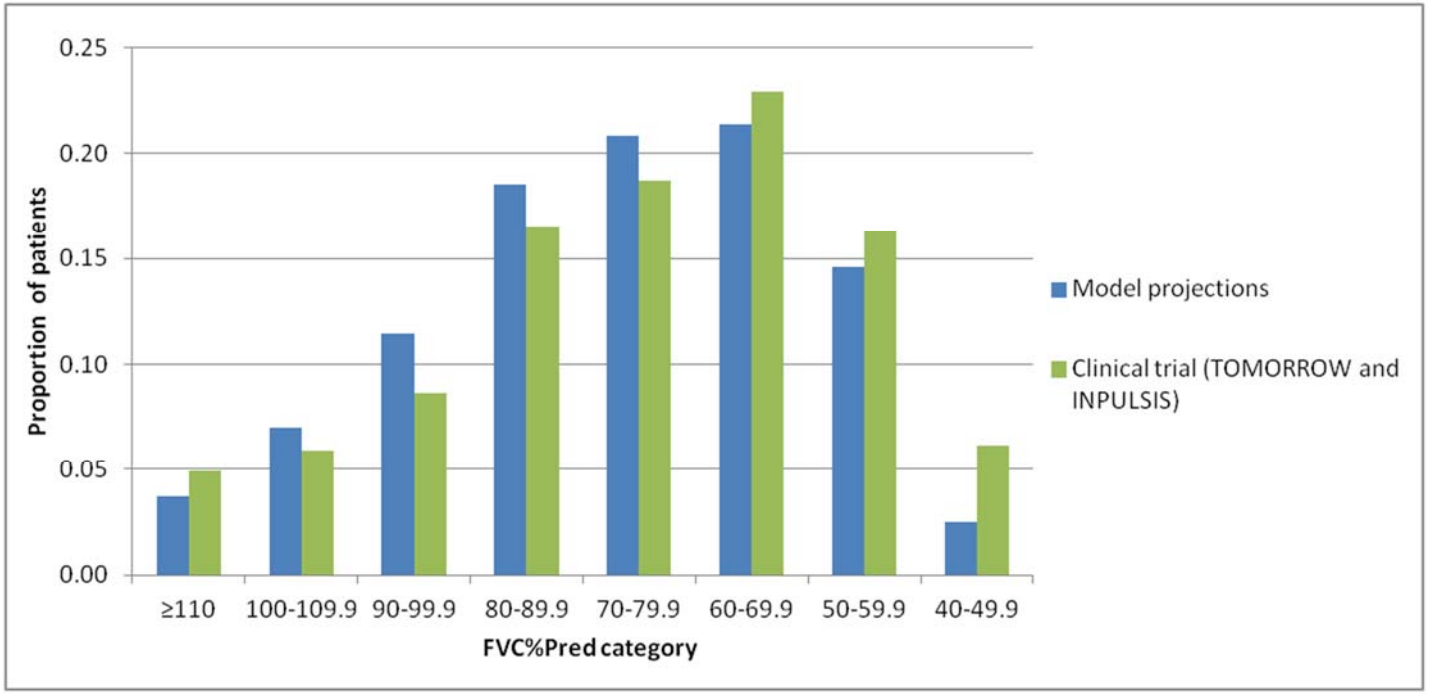


Table 170. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB arm

FVC%Pred category	≥110	100-109.9	90-99.9	80-89.9	70-79.9	60-69.9	50-59.9	40-49.9
Model projections	4.51%	7.30%	11.79%	19.03%	20.66%	21.14%	13.86%	1.72%
Clinical trial (TOMORROW and INPULSIS)	5.38%	6.77%	12.15%	17.53%	21.35%	19.79%	11.63%	5.38%

Figure 63. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB arm

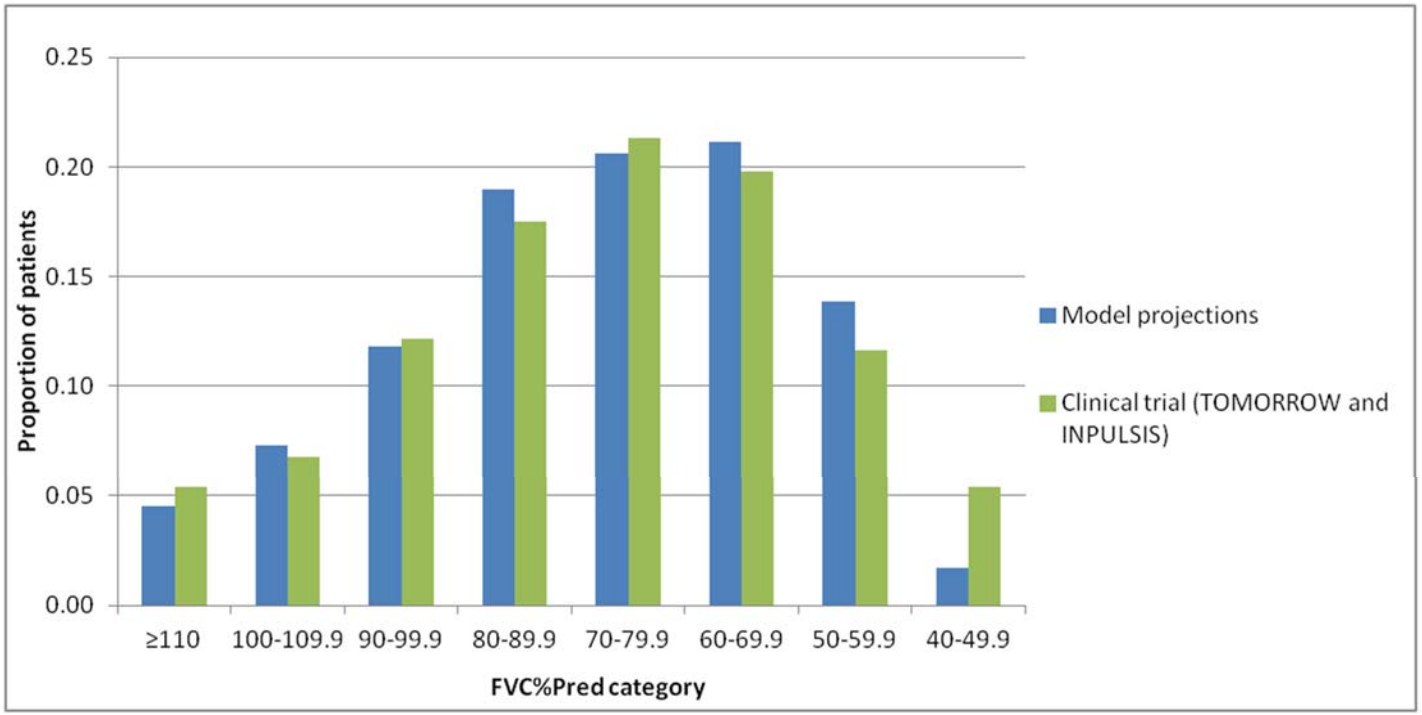
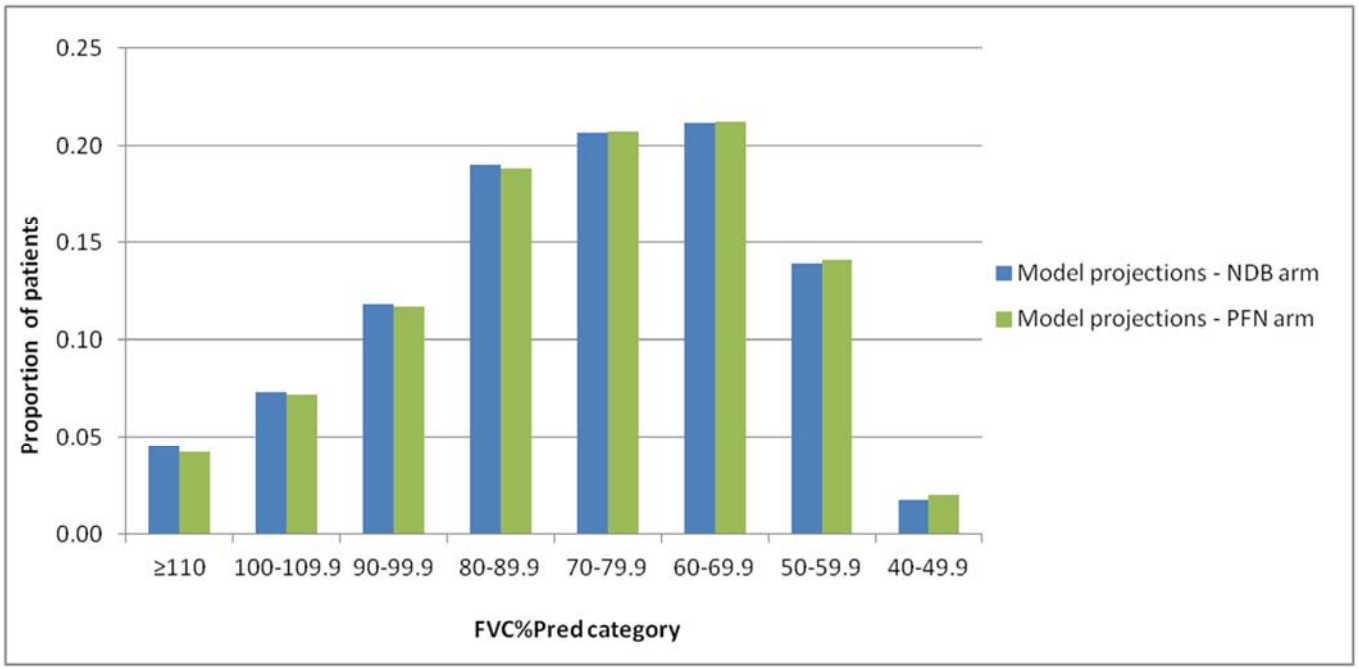


Table 171. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB and PFN

FVC%Pred category	≥110	100-109.9	90-99.9	80-89.9	70-79.9	60-69.9	50-59.9	40-49.9
Model projections NDB arm	4.51%	7.30%	11.79%	19.03%	20.66%	21.14%	13.86%	1.72%
Model projections PFN arm	4.23%	7.22%	11.66%	18.85%	20.73%	21.23%	14.10%	1.98%

Figure 64. Proportion of patients alive and in each FVC%Pred category at the end of one year: NDB and PFN arms (model projections)



5.7.4 Proportion of the cohort in different states over time

Figure 65: Proportion in each state in nintedanib arm

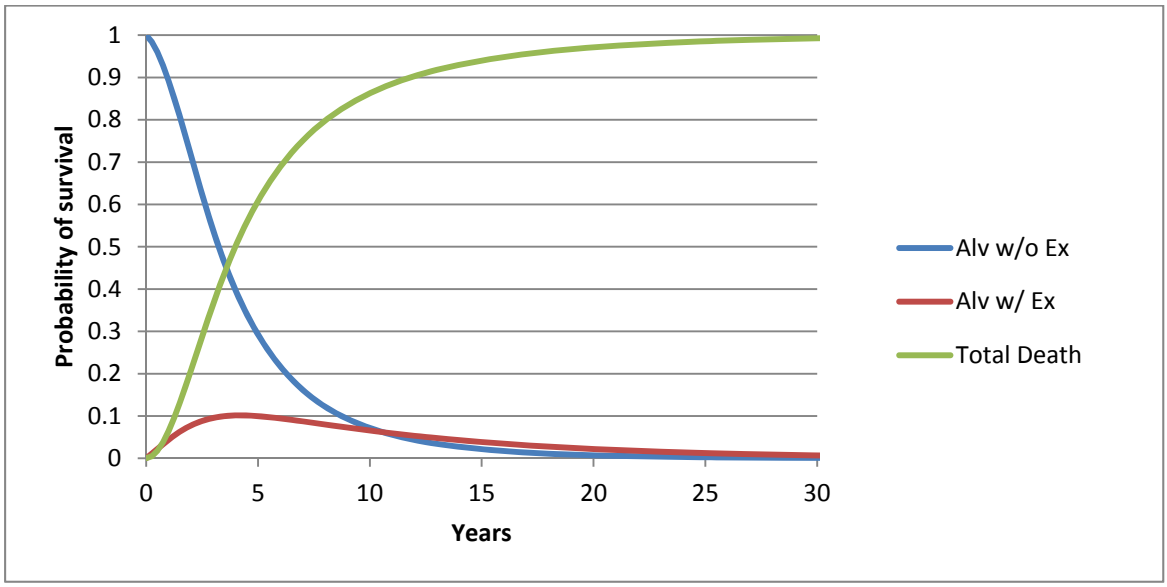


Figure 66: Proportion in each state in pirfenidone arm

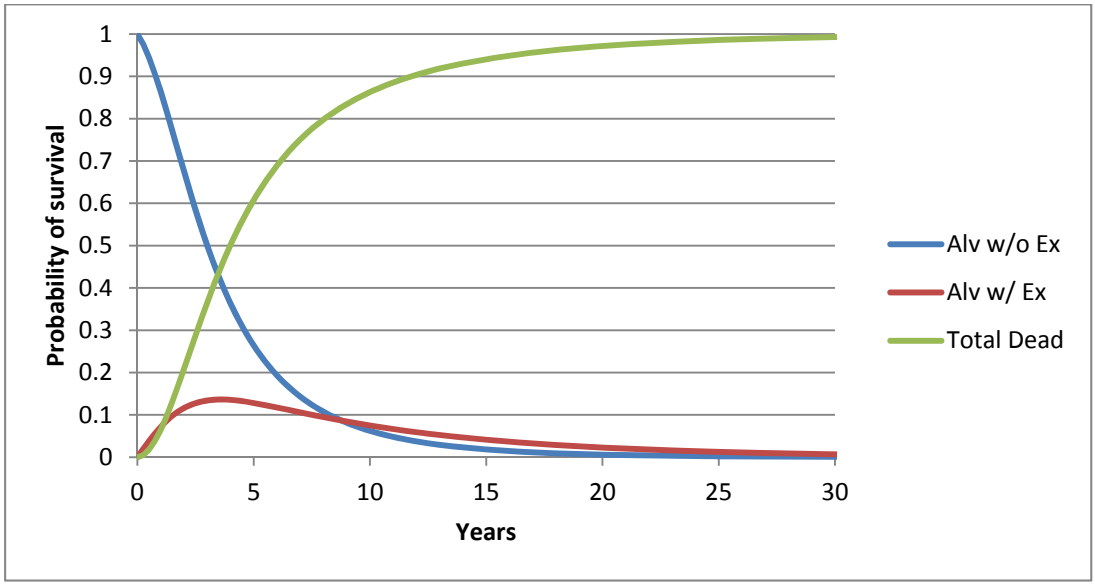
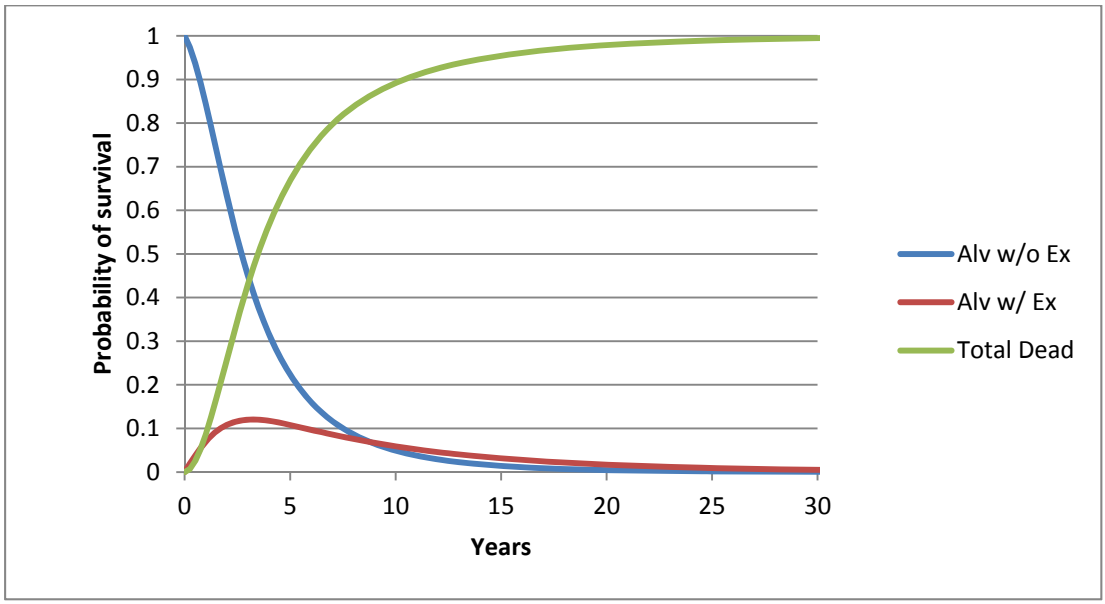


Figure 67: Proportion in each state in best supportive care arm



5.7.5 Description of how QALYs are accrued in the model

The Markov model has a number of states determined by percent predicted of FVC and exacerbation status. These states are outlined in detail in section 5.2.2. Each state within the model is associated with a health-related utility to estimate the QALY over the time horizon of the analysis (section 5.4.1). The cycle length in the model is 3 months and patients can transition between states at each cycle. This component of the QALY is calculated per cycle based on the distribution of the cohort across the health states and the utility associated with being in the health state.

Utility reductions (disutility) due to AEs are applied in the model based on the estimated proportion of patients suffering from AEs in each treatment arm (section 5.4.6).

The model uses a lifetime time horizon in order to fully incorporate the health outcomes of IPF.

5.7.6 Disaggregated results of the base case incremental cost effectiveness analysis

Table 172, Table 173 and Table 174 illustrate the model outputs by clinical outcomes for the BSC, nintedanib, pirfenidone and NAC arms respectively.

Table 172: Model outputs by clinical outcomes for BSC arm

	QALYs	Exacerbations	Total costs (£)
No events	3.3705	0.0000	£23,396.28
Exacerbation events	-0.0852	0.3234	£1,336.83
Serious cardiac events	-0.0120	-	£498.61
Serious GI events	-0.0012	-	£127.37
Photosensitivity reaction	-	-	-
Rash	-	-	-
GI perforation	-	-	-
Base-case	***3.2721***	***0.3234***	***£25,359.09***

Table 173: Model outputs by clinical outcomes for nintedanib arm

	QALYs	Exacerbations	Total costs (£)
No events	3.7693	0.0000	£83,114.03
Exacerbation events	-0.0820	0.2882	£1,191.50
Serious cardiac events	-0.0129	-	£535.96
Serious GI events	-0.0022	-	£230.02
Photosensitivity reaction	-	-	-
Rash	-	-	-
GI perforation	-0.0002	-	£15.99
Base-case	***3.6720***	***0.2882***	***£85,087.49***

Table 174: Model outputs by clinical outcomes for pirfenidone arm

	QALYs	Exacerbations	Total costs (£)
No events	3.7600	0.0000	£84,922.51
Exacerbation events	-0.1016	0.3633	£1,501.72
Serious cardiac events	-0.0151	-	£624.74
Serious GI events	-0.0011	-	£114.82
Photosensitivity reaction	-0.0044	-	£80.41
Rash	-0.0129	-	£234.95
GI perforation	**-**	**-**	**-**
Base-case	**3.6249**	**0.3633**	**£87,479.15**

5.8 *Sensitivity analyses*

Probabilistic sensitivity analysis

5.8.1; 5.8.2 Probabilistic sensitivity analysis inputs

In order to explore uncertainty around the model key variables in the base-case, probabilistic sensitivity analysis (PSA) was performed (1,000 iterations). Table 175 below presents the parameters included in the PSA along with their assumed distributions and standard error or range.

Table 175: Probabilistic sensitivity analysis - parameter inputs

Variable	Distribution	Assumption/sources
Transitions – baseline risk (PBO)		
Overall survival	Multivariate normal	
Exacerbations	Multivariate normal	
Loss of lung function	Multivariate normal	
Discontinuation	Beta (α=105.33, β=1798.9)	Based on SE=0.098, 95% CI (7.190, 7.570)
Transitions – treatment effects (NDB vs. PBO, PFN vs. PBO, OR values)		
Overall survival	Using codas from the WinBugs output (NMA)	
Exacerbations	Using codas from the WinBugs output (NMA)	
Loss of lung function	Using codas from the WinBugs output (NMA)	
Discontinuation	Using codas from the WinBugs output (NMA)	
Adverse events risks		
Serious cardiac event – baseline risk (PBO arm)	Beta (α=23, β=400)	Total N AEs=423, N (cardiac event PBO arm)=23 (Ingelheim, 2014[135])
Serious cardiac event – relative risk (NDB vs. PBO, PFN vs. PBO, OR values)	Using codas from the WinBugs output (NMA)	
Serious GI event – baseline risk (PBO arm)	Beta (α=7, β=416)	Total N AEs=423, N (GI event PBO arm)=7 (Ingelheim, 2014[135])
Serious GI event – relative risk (NDB vs. PBO, PFN vs. PBO, OR values)	Using codas from the WinBugs output (NMA)	
Photosensitivity reaction (PBO arm)	Beta (α=6, β=341)	Total N AEs=347, N (photosensitivity reaction PBO arm)=6 (Noble, 2011[40])
Photosensitivity reaction (PFN arm)	Beta (α=42, β=303)	Total N AEs=345, N (photosensitivity reaction PFN arm)=42 (Noble, 2011[40])
Rash (PBO arm)	Beta (α=40, β=307)	Total N AEs=347, N (rash PBO arm)=40 (Noble, 2011[40]), Table 4]
Rash (PFN arm)	Beta (α=111, β=234)	Total N AEs=345, N (rash PFN arm)=111 (Noble, 2011[40]), Table 4]
GI perforation (PBO arm)	No sampling; 0 events	Total N AEs=508, N (GI perforation PBO arm)=0, FDA NDB report 2014 (FDA highlights of prescribing information, 2014[96])
GI perforation (NDB arm)	Beta (α=2, β=721)	Total N AEs=723, N (GI perforation NDB arm)=2, FDA NDB prescribing information 2014, based on 0.3% events (FDA highlights of prescribing information, 2014[96])
Liver function test cost		
Frequency of liver function test (per cycle)	Gamma(α=15.37, β=0.022)	LLCI= 0.17 (assume every 6 months), HLCI=0.5 (assume every 2 months)
Liver function test cost	Lognormal (mean=1.10, SE=0.219)	LLCI=£1.68, HLCI=£3.97 (NHS Reference Costs 2012/13, DAPS05 Haematology)(Department of Health, 2013[8])
Treatment-related AE costs		
Serious cardiac event cost	Lognormal (mean=7.63, SE=0.165)	LLCI = 1318.39, HLCI = 2521.95, weighted averages from NHS Reference Costs 2012/13(Department of Health, 2013[8])
Serious GI event cost	Lognormal (mean=7.47, SE=0.219)	LLCI =975.95, HLCI=2298.86, weighted averages from NHS Reference Costs 2012/13(Department of Health, 2013[8])
Skin disorder cost	Lognormal (mean=5.92, SE=0.070)	LLCI=313.24, HLCI=412.78, weighted averages from NHS Reference Costs 2012/13(Department of Health, 2013[8])

Variable	Distribution	Assumption/sources
GI perforation cost	Lognormal (mean=7.76, SE=0.218)	LLCI=1336.79, HLCI=3142.00, weighted averages from NHS Reference Costs 2012/13(Department of Health, 2013[8])
Resource use		
Number of hospitalisations	Lognormal (mean=0.12, SE=0.025)	LLCI = 1.07, HLCI = 1.18, based on SE = 0.029 (resource use calculations)
Average duration of each hospitalisation	Lognormal (mean=2.17, SE=0.080)	LLCI = 7.37, HLCI = 10.08, based on SE = 0.692 (resource use calculations)
Proportion in ICU	Beta (α=6.92, β=177.08)	LLCI = 0.01, HLCI = 0.07, based on SE =0.014 (resource use calculations)
Proportion with mechanical ventilation	Beta (α=1.98, β=182.02)	LLCI = -0.004, HLCI = 0.026, based on SE = 0.008 (resource use calculations)
Proportion associated with ER overnight	Beta (α=8.90, β=175.10)	LLCI = 0.02, HLCI = 0.08, based on SE = 0.016 (resource use calculations)
Proportion associated with ambulance use (hospital)	Beta (α=23.74, β=160.26)	LLCI = 0.08, HLCI = 0.18, based on SE = 0.025(resource use calculations)
Average number of ER visits	Lognormal (mean=0.09, SE=0.035)	LLCI = 1.02, HLCI = 1.17, based on SE = 0.038 (resource use calculations)
Proportion associated with ambulance use (ER)	Beta (α=19.66, β=95.34)	LLCI = 0.10, HLCI = 0.24, based on SE = 0.035 (resource use calculations)
Number of GP visits	Lognormal (mean=0.46, SE=0.034)	LLCI = 1.48, HLCI = 1.69, based on SE = 0.054 (resource use calculations)
Number of specialist visits	Lognormal (mean=0.60, SE=0.029)	LLCI = 1.72, HLCI = 1.93, based on SE = 0.052 (resource use calculations)
Number of nurse visits	Lognormal (mean=0.49, SE=0.116)	LLCI = 1.27, HLCI = 2.00, based on SE = 0.186 (resource use calculations)
Number of physiotherapist visits	Lognormal (mean=1.26, SE=0.186)	LLCI = 2.28, HLCI = 4.74, based on SE = 0.628 (resource use calculations)
Number of other visits	Lognormal (mean=0.65, SE=0.087)	LLCI = 1.60, HLCI = 2.24, based on SE = 0.166 (resource use calculations)
Number of chest HRCT	Lognormal (mean=0.09, SE=0.022)	LLCI = 1.05, HLCI = 1.15, based on SE = 0.024 (resource use calculations)
Number of chest x-ray	Lognormal (mean=0.45, SE=0.054)	LLCI = 1.40, HLCI = 1.73, based on SE = 0.085 (resource use calculations)
Number of O2 requirement assessment	Lognormal (mean=0.29, SE=0.110)	LLCI = 1.04, HLCI = 1.61, based on SE = 0.144 (resource use calculations)
Number of bronchoalveolar lavage	Lognormal (mean=0, SE=1.172)	LLCI = 0.02, HLCI = 1.98, based on SE = 0.5 (SE value assume 0.5 due to sampling error, SE=0)
Number of CT - pulmonary angiogram	Lognormal (mean=0.16, SE=0.155)	LLCI = 0.83, HLCI = 1.52, based on SE = 0.176 (resource use calculations)
Number of right heart catheterization	Lognormal (mean=0.08, SE=-0.051)	LLCI = 0.97, HLCI = 1.19, based on SE = 0.055 (resource use calculations)
Number of general diagnostic	Lognormal (mean=0.46, SE=0.041)	LLCI = 1.45, HLCI = 1.71, based on SE = 0.065 (resource use calculations)
Exacerbations - Number of hospitalisations	Lognormal (mean=0.26, SE=0.142)	LLCI = 0.95, HLCI = 1.65, based on SE = 0.18 (resource use calculations)
Exacerbations - Average duration of each hospitalisation	Lognormal (mean=2.79, SE=0.329)	LLCI = 7.05, HLCI = 25.55, based on SE = 4.72 (resource use calculations)
Baseline resource use (FVC%pred 110+→40)	Sampling resource use (beta distribution)	
Oxygen use – cost per cycle	Lognormal (mean=6.04, SE=0.303)	LLCI = 182.84 (inflation to 2012/13 values, NICE Clinical Guideline (table 45 Appendix); assume HLCI=600.77 (LLCI+mean)
Exacerbation cost – acute event	Sampling resource use (beta distribution)	
End of life (EoL) cost	Lognormal (mean=8.27, SE=0.158)	LLCI=2,744, HLCI=5,097; Assume 30% up/down around mean value
Baseline utility values		
FVC%pred 110 and above	Linked to FVC%pred 90	
FVC%pred 100	Linked to FVC%pred 90	

Variable	Distribution	Assumption/sources
FVC%pred 90	Beta (α=1639.86, β=317.09)	SD=0.178, N=458 (INPULSIS baseline utilities data)
FVC%pred 80	Beta (α=2022.89, β=473.06)	SD=0.205, N=684 (INPULSIS baseline utilities data)
FVC%pred 70	Beta (α=2093.81, β=590.69)	SD=0.224, N=788 (INPULSIS baseline utilities data)
FVC%pred 60	Beta (α=1961.53, β=600.10)	SD=0.238, N=809 (INPULSIS baseline utilities data)
FVC%pred 50	Beta (α=1300.20, β=459.98)	SD=0.232, N=490 (INPULSIS baseline utilities data)
FVC%pred 40	Beta (α=222.19, β=112.75)	SD=0.255, N=98 (INPULSIS baseline utilities data)
Investigator exacerbations disutility		
1st month	Beta (α=7.49, β=46.014)	95% CI: (-0.231,-0.049) (INPULSIS exacerbations utilities data)
subsequent month	Beta (α=5.40, β=63.830)	95% CI: (-0.142, -0.015) (INPULSIS exacerbations utilities data)
Adjudication committee exacerbations disutility		
1st month	Beta (α=15.38, β=40.762)	95% CI: (-0.39, -0.157) (INPULSIS exacerbations utilities data)
subsequent month	Beta (α=0.34, β=10.019)	95% CI: (-0.137, 0.07)(INPULSIS exacerbations utilities data)
Treatment-related adverse events disutility		
Serious cardiac event	Beta (α=34.04, β=137.859)	LLCI=-0.2574, HLCI=-0.1386; assume 30% variation around mean value
Serious GI event	Beta (α=0.87, β=11.896)	95% CI: (-0.201, 0.065) (INPULSIS post-hoc analysis of safety data, serious GI events)
Skin disorder	Beta (α=39.10, β=437.755)	LLCI=-0.1066, HLCI=-0.0574; assume 30% variation around mean value
GI perforation	Beta (α=37.53, β=280.518)	LLCI=-0.1534, HLCI=-0.0826; assume 30% variation around mean value

5.8.3 Incremental cost effectiveness results of the probabilistic sensitivity analysis

Table 176 shows that incremental costs are similar for the deterministic and average PSA result. Although the average PSA sample value for incremental QALYs is slightly higher than the deterministic value, Table 177shows that the check row is very close to 1 for both the NDB and PFN arms. The check row is the ratio of the deterministic result vs. the probabilistic average. It can also be seen that the NDB total QALY samples 0.007 higher than its deterministic value, while for PFN it samples 0.01 lower than its deterministic value. This comes to a 0.017 difference, which explains the difference between the incremental average of samples and the incremental deterministic incremental QALY.

Table 176: PSA results for nintedanib vs. pirfenidone

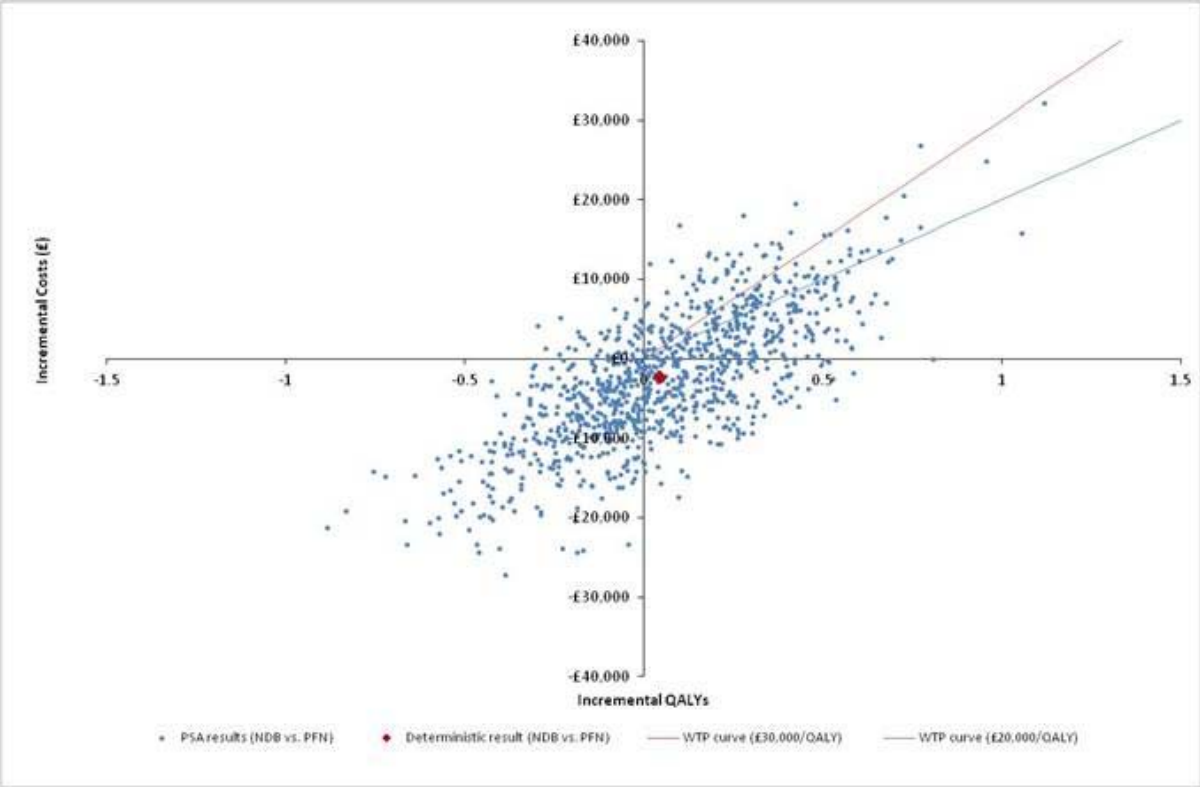
	Incremental cost	Incremental QALY	ICER
Deterministic	[redacted]	0.0471	NDB dominated PFN
Average value from PSA	[redacted]	0.0636	NDB dominates PFN

Table 177: QALY results for nintedanib and pirfenidone

	NDB	PFN	Incremental (NDB – PFN)
Deterministic analysis	3.672	3.625	0.0471
Average of samples	3.679	3.615	0.0636
Variance	0.409	0.387	0.022
Standard error	0.020	0.020	0.001
Lower limit	2.543	2.481	0.061
Upper limit	5.021	4.946	0.075
Check	0.99821	1.00274	0.74065

The result of the PSA (1,000 samples) is presented in [redacted] 68. The scatter-plot indicates that NDB and PFN are nearly equivalent, with samples scattered around the origin.

Figure 68: Incremental cost-effectiveness scatter-plot (nintedanib vs. pirfenidone)



The cost-effectiveness acceptability curve (CEAC) is presented in Figure 69. The CEAC shows that there is a 60% probability that NDB is cost-effective compared with PFN at any threshold value.

Figure 69: CEAC (nintedanib vs. pirfenidone)

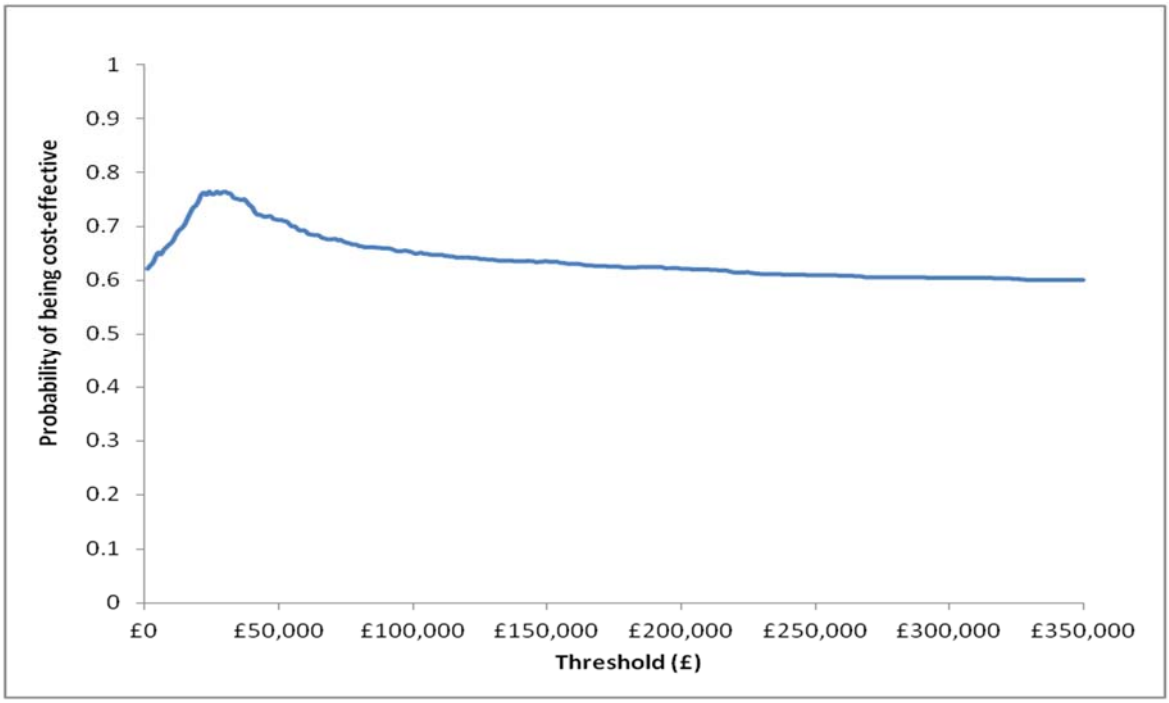


Table 178 shows that the deterministic and average probabilistic results are similar for both the incremental costs and the incremental QALYs.

Table 178: PSA results for nintedanib vs. best supportive care

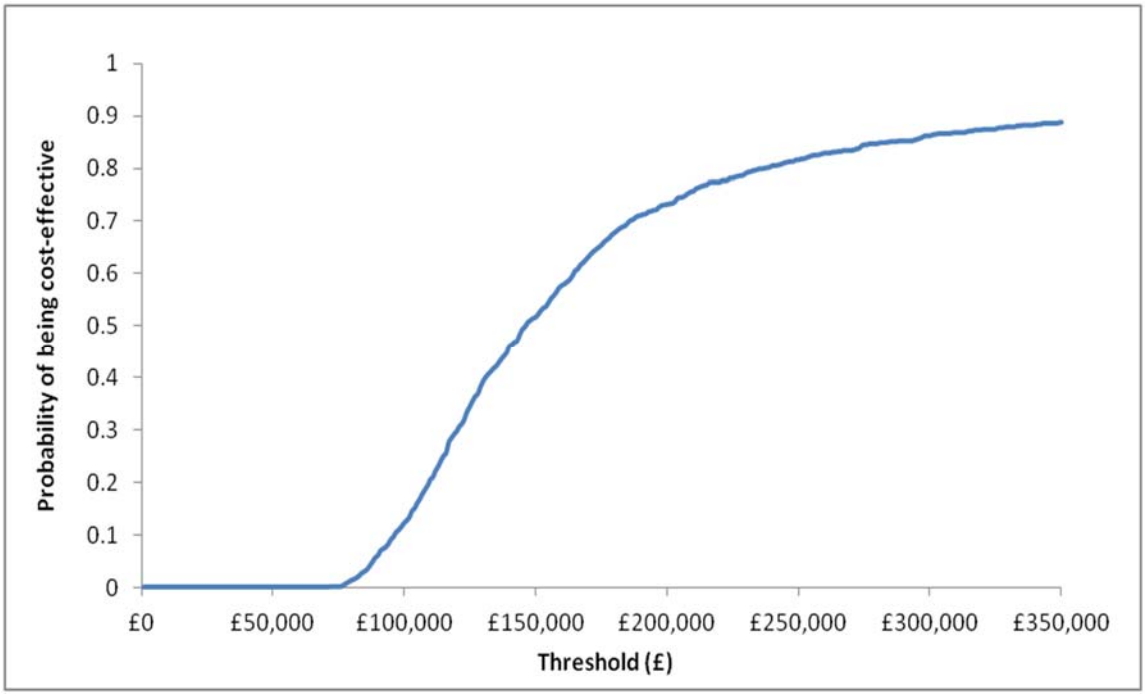
	Incremental cost	Incremental QALY	ICER
Deterministic	£59,728.40	0.4009	£149,361
Average value from PSA	£59,863.32	0.4061	£147,572

The result of the PSA (1,000 samples) is presented in Figure 70.

Figure 70: Incremental cost-effectiveness scatter-plot (nintedanib vs. BSC)

The cost-effectiveness acceptability curve (CEAC) is presented in Figure 71.

Figure 71: CEAC (nintedanib vs. BSC)



PSA results for full incremental analysis of all model arms (BSC, NDB, PFN)

Figure 72: PSA scatterplot (all active treatments vs. BSC)

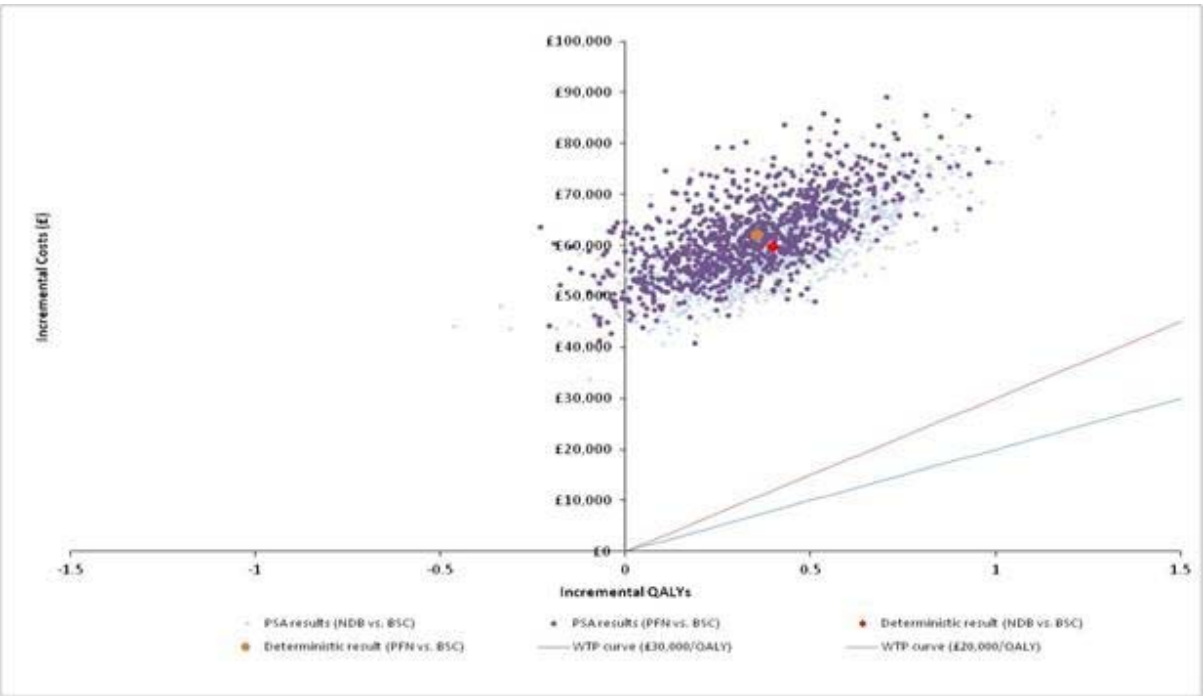
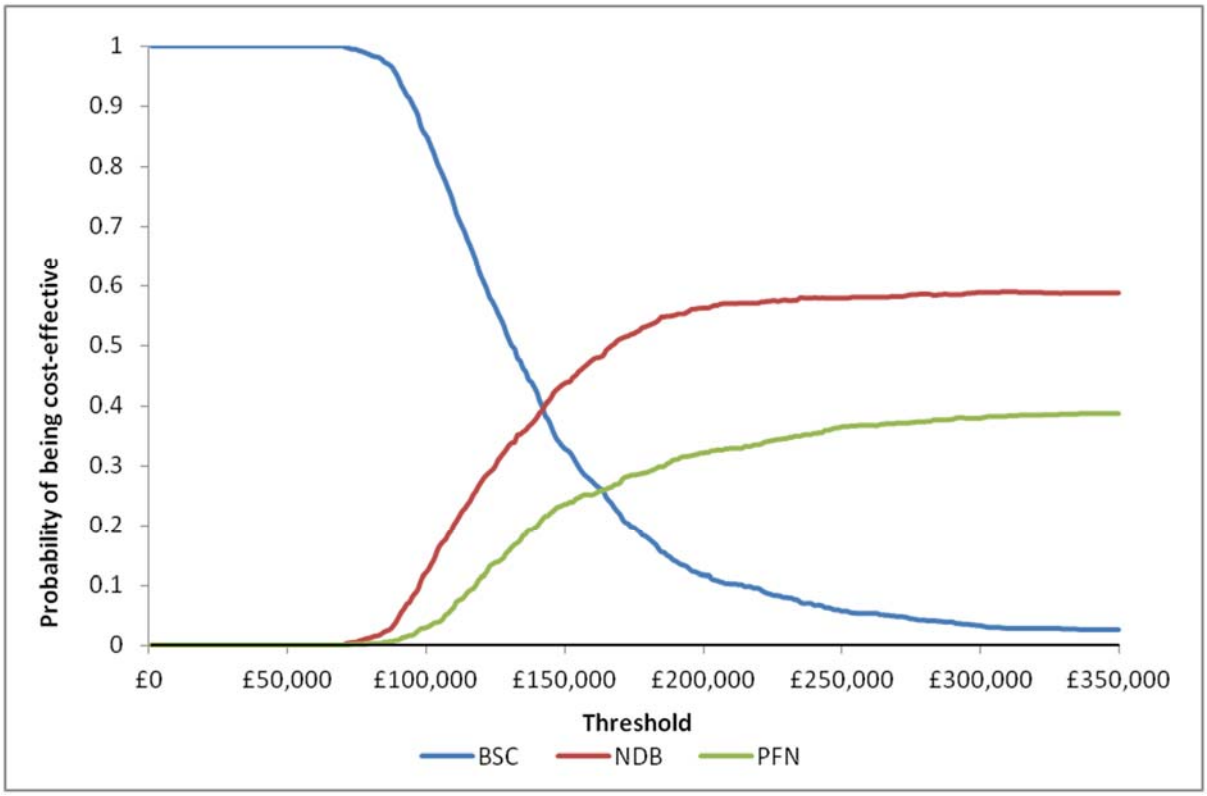


Figure 73: Multiple CEAC



5.85; 5.86; 5.87; 5.8.8; 5.8.9 Deterministic sensitivity analysis and Scenario analysis

One-way sensitivity analysis methodology

The sensitivity of the model results and the drivers of cost-effectiveness were explored with o OWSA. OWSA was performed in two ways:

- Performing tests around the 95% confidence interval values of all model parameters (scenarios 1-14)
- Testing several model assumptions (scenarios 15-47)

Table 179: OWSA scenarios testing parameters between the limits of their 95% CI

Scenario	Parameter varied
1	Mortality probabilities
2	Exacerbation probabilities
3	Progression probabilities
4	Discontinuation probabilities
5	Treatment costs
6	Background follow-up costs
7	Oxygen use costs
8	Exacerbation events costs
9	End of life costs
10	Health state baseline utilities
11	Exacerbation-related utilities
12	Adverse events probabilities
13	Adverse events costs
14	Adverse events-related utilities

Overall survival (scenarios 15-21).

Table 180: OWSA Scenarios for OS

Scenario	Parameter varied
15	Baseline risk: Weibull parametric model
16	Baseline risk: Gompertz parametric model
17	Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)
18	Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)
19	Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi
20	Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)
21	Relative risk: NMA results, scenario 5 excluding the Japanese studies

Acute exacerbations (scenarios 22-28).

Table 181: OWSA scenarios for acute exacerbations

Scenario	Parameter varied
22	Baseline risk: use adjudication committee estimate
23	Baseline risk: exclude recurrent exacerbation risk
24	Relative risk: NMA results, scenario 1, all evidence
25	Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)
26	Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma
27	Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)
28	Relative risk: NMA results, scenario 6 excluding Azuma and Homma

Loss of lung function (scenarios 29-33).

Table 182: OWSA scenarios for loss of lung function

Scenario	Parameter varied
29	Baseline risk: include exacerbation coefficient
30	Relative risk: NMA results, scenario 1, all evidence
31	Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King
32	Relative risk: NMA results, scenario 4, include death (without Noble)
33	Relative risk: NMA results, scenario 5, include death (with Noble)

Safety (scenarios 34-38).

Table 183: OWSA Scenarios for safety

Scenario	Parameter varied
34	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence
35	Relative risk: serious GI events, NMA results, scenario 2 excluding Richeldi 2011
36	SAE disutility value: use alternative value for serious cardiac events
37	SAE disutility value: use alternative value for GI perforation
38	SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value

Discontinuation risk (scenarios 39-43).

Table 184: OWSA scenarios for discontinuation

Scenario	Parameter varied
39	Baseline risk: higher discontinuation rate for NDB and PFN. An InterMune Canadian Registry following patients on IPF reports a discontinuation rate of 22% in 9.5 months [MacQuarrie and Lebel 2014, ATS poster]. This scenario examined the effect of this discontinuation risk for NDB and PFN (6.71% per cycle) on the model results. This scenario is of interest for the NDB vs. PFN comparison only.
40	Relative risk: NMA results, scenario 2,excluding King and Taniguchi
41	Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011
42	Relative risk: NMA results, scenario 4, excluding King
43	Relative risk: NMA results, scenario 5, excluding Taniguchi

FVC%Pred categories (scenarios 44-45).

Table 185: OWSA for FVC%pred categories

Scenario	Parameter varied
44	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point
45	Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point

PFN stopping rule (scenario 46)

This scenario explores the effect of patients discontinuing and losing treatment effect when they observe a decline of FVC%Pred of 10% or more (progressors).

Used the direct evidence for NDB (scenario 47).

In this scenario the relative efficacy of NDB vs. BSC is based on an extrapolation of the NDB clinical trials (Ingelheim, 2014[97]) for the outcomes of OS, acute exacerbations, loss of lung function, and discontinuation, rather than via an OR from the NMA.

The effect of comparing the results of the ASCEND trial (King *et al.*, 2014[41]) to an "ASCEND-like" population from the nintedanib trials was also investigated (scenario 48). Here the population is restricted to the ASCEND selection criteria, which are: IPF diagnosed at least 0.5 years before visit 2, FVC 50-90% predicted, FEV1/FVC ≥ 0.8. The authors state that the ASCEND selection criteria were chosen to "[…] allow enrolment of patients with an increased risk of disease progression" (King *et al.*, 2014[41]). This has been modelled by using a slightly different version of the cost-effectiveness model. Table 186 reports the differences between the main model used to generate the results and the model version used to generate the results for scenario 48.

Table 186: Differences between the ASCEND-like population model and the main model

Model input	ASCEND-like model	Main model
Patient distribution at the start of the model	Based on the INPULSIS restricted population data (ASCEND-like).	Based on the broad INPULSIS population data, no restrictions.
Overall survival	OS parametric models derived for the BSC and NDB arms (informed by the joint PBO and NDB populations, pooled phase II and phase III trial data – TOMORROW and INPULSIS). HR for NDB vs. PFN (from ITC performed on the INPULSIS restricted population and ASCEND population data) applied for the PFN mortality risk.	OS parametric models derived for the BSC arm (informed by the PBO pooled phase II and phase III trial data – TOMORROW and INPULSIS) OR (from NMA) applied to baseline BSC risk to inform the mortality risks of the active treatments (NDB, PFN).
Time to acute exacerbation (investigator reported and adjudication committee reported)*	Time to first acute IPF exacerbation was calculated using the INPULSIS restricted population for both investigator-reported and adjudication committee exacerbations. Exacerbation parametric models were derived for the BSC and NDB arms (informed by the joint PBO and NDB populations, phase III trial data –INPULSIS). RR values for NDB vs. PFN (informed by an ITC performed on the broad INPULSIS and CAPACITY populations, due to lack of exacerbation data in the ASCEND trial), for the investigator-reported, and adjudication committee reported exacerbations, were applied for the PFN exacerbation risks.	Parametric model extrapolations were used for the BSC arm (informed by the PBO phase III trial data - INPULSIS). OR (from NMA) applied to baseline BSC risk to inform the exacerbation risks of the active treatments (NDB, PFN).
Loss of lung function	Two logistic models were used to capture the predictors of lung function decline. Predictors used: current FVC%pred state, treatment allocation (NDB or PBO arm), exacerbation (before and after an exacerbation event) Predictors used: current FVC%pred state, treatment allocation (NDB or PBO arm). RR for NDB vs. PFN (from ITC performed on the INPULSIS restricted population and ASCEND population data) applied for the PFN progression risk	Logistic regression model was used for the BSC arm (informed by the PBO phase III trial data – TOMORROW and INPULSIS). Loss of lung function predictors considered: current FVC%Pred at the start of a 3-month interval and exacerbation.** OR (from NMA) applied to baseline BSC risk to inform the loss of lung function risks of the active treatments (NDB, PFN).
Safety	For comparability with the ASCEND trial data, AEs were selected if they were observed with an incidence of ≥10%. AEs meeting the ≥10% threshold were selected for the INPULSIS restricted population and ASCEND population. The incidence of AEs was calculated using the number of patients experiencing each event throughout the two clinical trials.	The incidence of SAEs (serious cardiac and serious GI events***) was calculated for the PBO arm, using the number of patients experiencing each event throughout the INPULSIS clinical trial. OR (from NMA) applied to baseline BSC risk to inform the SAE risks of the active treatments (NDB, PFN). Clinically important events (flagged by clinicians) were also included in the model as patient numbers for each event, per treatment arm.
Treatment tolerability	PBO and NDB overall discontinuation risk were calculated based on parametric modelling extrapolation of phase III trial data, joint populations of PBO and NDB, INPULSIS population restricted to the ASCEND criteria. PFN discontinuation risk was based on the	PBO overall discontinuation risk was calculated based on parametric modelling extrapolation of phase III trial data (INPULSIS). The discontinuation risks for the active treatments (NDB, PFN) were estimated

Model input	ASCEND-like model	Main model
	number of patients discontinuing treatment throughout the ASCEND clinical trial.	via OR values obtained from the NMA.
AEs disutility values and costs	AEs disutilities values and costs differ because the safety profiles are different between the two models.	

*The base-case model uses the investigator reported exacerbation estimates as base-case (in line with the type of exacerbations used in the NMA exercise), while the ASCEND-like model uses the adjudication committee estimates in the base-case. The impact of the adjudication committee estimate was explored in sensitivity analysis for the base-case model and the results did not change significantly (scenario 22).
**The base-case model uses the loss of lung function logistic model without the exacerbation covariate (which was not statistically significant in the re-analysis of the PBO data only), while the ASCEND-like model uses the exacerbation coefficient in the base-case. The impact of the exacerbation coefficient was explored in sensitivity analysis for the base-case model and the results did not change significantly (scenario 29).
***Selection criteria for the base-case model: AEs with a significant impact on costs and QALYs: assumed to be those that were severe or serious, AEs with an incidence greater or equal to 5%, AEs with an incidence of 1.5 times greater between the two arms.

One-way sensitivity analysis (OWSA) results

OWSA results for NDB vs. PFN

Table 187 reports the results of the deterministic sensitivity analyses performed to identify the drivers of the model (scenarios 1-48). The sections below describe the findings of the sensitivity analysis. Please note that a tornado diagram is not presented for this comparison because dominance cannot be represented in such a graph.

Tests around the 95% confidence interval values of model parameters

Probabilities

NDB dominated PFN for all variations performed on probability values.

Costs

NDB dominated PFN for all variations performed on cost values.

Utilities

NDB dominated PFN for all variations performed on utility values.

Additional sensitivity analyses on model assumptions

Overall survival

NDB dominated PFN for all scenarios performed on overall survival.

Exacerbations

NDB dominated PFN for all scenarios performed on exacerbations.

Loss of lung function

NDB dominated PFN for all scenarios performed on loss of lung function.

Safety

NDB dominated PFN for all scenarios performed on safety.

Overall discontinuation

NDB dominated PFN for all scenarios performed on overall discontinuations.

FVC%Pred categories

NDB dominated PFN for both scenarios performed on the FVC%Pred starting categories.

PFN stopping rule

Imposing a stopping rule (scenario 46) on progressors, i.e. discontinuation and loss of treatment effect for PFN patients that show a loss of lung function of at least 10% in FVC%Pred, meant that patients incurred less costs and accrued more QALYs in the PFN arm. The ICER changes from a negative value in the base-case (NDB dominating PFN) to a positive ICER value of £82,784 per QALY gained.

Use of direct evidence for NDB

This scenario (47) was not applicable to the NDB vs. PFN comparison.

Table 187: Sensitivity analysis results (nintedanib vs. pirfenidone)

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	NDB dominates	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	NDB dominates	NDB dominates
2		Exacerbation probabilities	NDB dominates	NDB dominates
3		Progression probabilities	NDB dominates	NDB dominates
4		Discontinuation probabilities	NDB dominates	NDB dominates
5	Costs	Treatment costs	NDB dominates	NDB dominates
6		Background follow-up costs	NDB dominates	NDB dominates
7		Oxygen use costs	NDB dominates	NDB dominates
8		Exacerbation event costs	NDB dominates	NDB dominates
9		EoL costs	NDB dominates	NDB dominates
10	Utilities	Health state baseline utilities	NDB dominates	NDB dominates
11		Exacerbation-related utilities	NDB dominates	NDB dominates
12	Adverse events	Adverse events probabilities	NDB dominates	NDB dominates
13		Adverse events costs	NDB dominates	NDB dominates
14		Adverse events related utilities	NDB dominates	NDB dominates
Tests on other model parameters				
Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
15	Overall survival	Baseline risk: Weibull parametric model	NDB dominates	
16		Baseline risk: Gompertz parametric model	NDB dominates	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	NDB dominates	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	NDB dominates	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	NDB dominates	

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	NDB dominates	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	NDB dominates	
22	Exacerbation	Baseline risk: use adjudication committee estimate	NDB dominates	
23		Baseline risk: exclude recurrent exacerbation risk	NDB dominates	
24		Relative risk: NMA results, scenario 1, all evidence	NDB dominates	
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	NDB dominates	
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	NDB dominates	
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	NDB dominates	
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	NDB dominates	
29	Loss of lung function	Baseline risk: include exacerbation coefficient	NDB dominates	
30		Relative risk: NMA results, scenario 1, all evidence	NDB dominates	
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	NDB dominates	
32		Relative risk: NMA results, scenario 4, include death (without Noble)	NDB dominates	
33		Relative risk: NMA results, scenario 5, include death (with Noble)	NDB dominates	
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	NDB dominates	
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	NDB dominates	
36		SAE disutility value: use alternative value for serious cardiac events	NDB dominates	
37		SAE disutility value: use alternative value for GI perforation	NDB dominates	
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	NDB dominates	
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	NDB dominates	
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	NDB dominates	
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	NDB dominates	
42		Relative risk: NMA results, scenario 4, excluding King	NDB dominates	
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	NDB dominates	
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	NDB dominates	
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as	NDB dominates	

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
		starting point		
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	£82,784.36	
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	N/A for this comparison	

ASCEND-like population: NDB vs. PFN (scenario 48)

ASCEND-like population: NDB vs. PFN

This sensitivity analysis (scenario 48) was performed to investigate the cost-effectiveness of nintedanib vs. pirfenidone in a restricted population representative of the ASCEND trial selection criteria. The restricting criteria used in creating the subgroup are: IPF diagnosed at least 0.5 years before visit 2, FVC 50-90% predicted, $FEV_1/FVC \geq 0.8$. In their publication, the investigators of the ASCEND trial state that the ASCEND selection criteria were chosen to "[…] allow enrolment of patients with an increased risk of disease progression" (King *et al.*, 2014[41]).

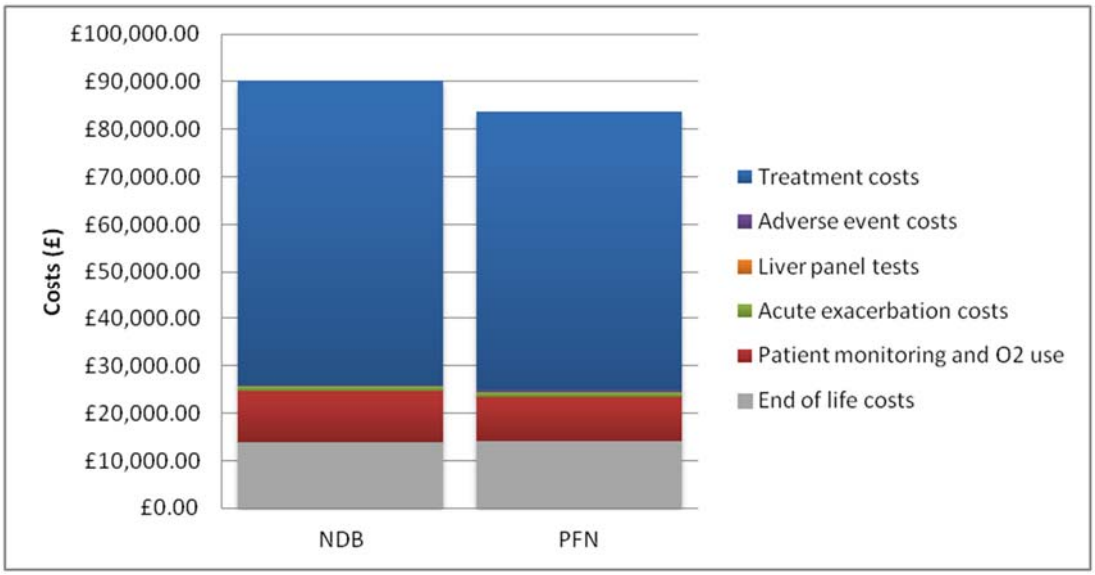
The base-case deterministic results of NDB vs. PFN for this sensitivity analysis are presented in Table 188. The total cost for the NDB treatment strategy is estimated to be £64,388, and £58,803 for the PFN one. The breakdown of the total cost is reported in Table 188 below. With an incremental total cost of £6,538 and an estimated incremental QALY gained of 0.49, the estimated ICER for NDB vs. PFN is £13,459 per QALY gained. The incremental LYs gained with NDB are estimated to be 0.53 (6 months). The estimated number of acute exacerbation events avoided with NDB is 0.052.

Table 188: Deterministic results for nintedanib vs. pirfenidone

	PFN	NDB	Incremental
Treatment costs	£58,803.29	£64,387.68	£5,584.39
Adverse event costs	£361.74	£256.00	-£105.75
Liver panel tests	£9.01	£9.87	£0.86
Patient monitoring and O2 use	£9,276.29	£10,770.37	£1,494.08
Acute exacerbation costs	£1,142.57	£929.36	-£213.20
End of life costs	£14,094.61	£13,871.76	-£222.85
Total costs	£83,687.52	£90,225.03	£6,537.52
Total QALYs	2.5024	2.9881	0.4857
ICER (per QALY)			£13,459.17

Figure 74 presents the analysis of total costs over the entire model duration.

Figure 74: Base-case total costs (nintedanib and pirfenidone)



Impact of Discount Rate Variation (Pirfenidone PAS)

As Boehringer Ingelheim does not have access to the net price of pirfenidone used in practice, scenario analyses were performed by reducing the list price for pirfenidone in 5% increments from a 5% discount to a 95% discount, and with the nintedanib discount implemented.

Table 189: Impact of pirfenidone discount rate on the ICER, nintedanib at list price

	Discount Applied to Drug Cost						
Pirfenidone	**Base-case (0% discount)**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**
Cost per pack	£2,151.10	£2043.55	£1935.99	£1828.44	£1720.88	£1613.33	£1505.77
ICER	Dominates	£13,663.45	£78,108.24	£142,553.03	£206,997.81	£271,442.60	£335,887.39
Pirfenidone	**35%**	**40%**	**45%**	**50%**	**55%**	**60%**	**65%**
Cost per pack	£1398.22	£1290.66	£1183.11	£1075.55	£968.00	£860.44	£752.89
ICER	£400,332.17	£464,776.96	£529,221.75	£593,666.54	£658,111.32	£722,556.11	£787,000.90

	Discount Applied to Drug Cost						
Pirfenidone	**70%**	**75%**	**80%**	**85%**	**90%**	**95%**	
Cost per pack	£645.33	£537.78	£430.22	£322.67	£215.11	£107.56	
ICER	£851,445.68	£915,890.47	£980,335.26	£1,044,780.04	£1,109,224.83	£1,173,669.62	

OWSA results for NDB vs. BSC

Table 191 reports the results of the deterministic sensitivity analyses performed to identify the drivers of the model. The main drivers of the model are presented in Figure 78 in the form of a tornado diagram. The sections below describe the findings of the sensitivity analysis.

Tests around the 95% confidence interval values of model parameters

Probabilities

The impact of the modelled probabilities are explored in scenarios 1 (mortality), 2 (exacerbations), 3 (progression), 4 (discontinuation), and 12 (adverse events probabilities).

The mortality risk estimates were very strong drivers of the model results; the ICER ranged from £87,000 to BSC dominating the NDB treatment. This was mainly due to the effect of mortality on the total and incremental QALYs. Given that the base case QALYs gained were relatively low, any small change in the QALYs could have a big impact on the ICER. When the mortality risk test adopted the low limit of the 95% confidence interval of the NMA OR values used for OS, the difference in incremental QALYs between the two strategies increased, and the ICER decreased by about £62,000 per QALY gained. Note that the mortality parameter was removed from Figure 78 because one of the results included dominance.

Varying the exacerbation probabilities had a low impact on the overall ICER; it was reduced by about £3,000 for the low value of the 95% CI for the exacerbation OR value, while for the high exacerbation OR value the ICER was increased by about £8,000 per QALY gained.

The variation of the progression risks had a relatively low impact on the overall results; the ICER was reduced by approximately £6,000 for the low progression OR value, while for the high progression OR value, the ICER was increased by about £9,000.

The ICER was reduced by approximately £9,000 for the low discontinuation OR value, while for the high discontinuation OR value, the ICER increased by about £14,000.

Finally, the AEs probabilities did not have a considerable effect on the ICER (variation of £2,000-£5,000).

Costs

The impact of alternative costs was explored in scenarios 5 (treatment costs), 6 (background follow-up costs), 7 (oxygen use costs), 8 (EoL costs), and 13 (adverse events costs). The biggest impact to the ICER was observed in scenario 5 where the treatment cost of NDB was varied ±5% (±£7,000 variation in the ICER value). Variation of the other cost inputs had no significant effect on the value of the ICER.

Utilities

Variation of the utilities parameters was explored in scenarios 10 (baseline utilities), 11 (exacerbation related), and 14 (AE-related). The changes performed on the utilities values had no significant effect on the ICER value (no more than ±£3,000 variation around the base-case value).

Additional sensitivity analyses on model assumptions

Overall survival

Changing the baseline OS risk from loglogistic (base-case) to a Weibull parametric model (scenario 15) had a significant impact on the result; the ICER increased by approximately £91,000. Similarly, changing the baseline risk to a Gompertz parametric model (scenario 16), affected the ICER even further, by about £320,000 per QALY gained.

The rest of the sensitivity scenarios on OS (scenarios 17-21) had no significant impact on the ICER.

Exacerbations

The scenarios performed on the exacerbation risk (scenarios 22-28) had no considerable effect on the model results.

Loss of lung function

The scenarios performed on the loss of lung function risk (scenarios 29-33) had no considerable effect on the model results.

Safety

The scenarios performed on the safety parameters (scenarios 34-38) had no considerable effect on the model results.

Overall discontinuation

The scenarios performed on the overall discontinuation parameters (scenarios 39-43) had no considerable effect on the model results.

FVC%Pred categories

The scenarios performed on the starting FVC%Pred categories (scenarios 44-45) had no considerable effect on the model results.

Use of direct evidence for NDB

This analysis involved changing the efficacy (OS, exacerbation, loss of lung function) and discontinuation extrapolation (scenario 47) to be based on the direct evidence of the clinical trial data rather than the NMA outcomes. This change had a significant impact on the model results. The ICER increased by £45,000 per QALY gained when using the extrapolated data based on the "joint" clinical trial population survival analysis. This effect was investigated further and an attempt to isolate the parameter(s) that led to this was made. To this end, each parameter was changed in the following sequence: progression, exacerbation, discontinuation, and mortality. The first three parameters did not affect the ICER substantially; however, when adding mortality (modelled using survival analysis for both the NDB and PFN arms), the ICER increased considerably.

The projected survival for NDB was compared further when this was a product of the direct or indirect evidence. Survival over 10 years is presented below in Figure 75 for loglogistic, Figure 76 for Weibull, and Figure 77 for Gompertz. The indirect evidence slightly overestimated survival comparing to the direct evidence. When the loglogistic model was used (longest survival out of the three OS parametric models), the overestimation increased (see Figure 75).

Therefore, the change in the results observed in this scenario was mainly explained by the increased survival estimated by using an OR vs. PBO for NDB instead of the extrapolation derived from the clinical trial.

Figure 75: Comparison of nintedanib survival for the loglogistic model

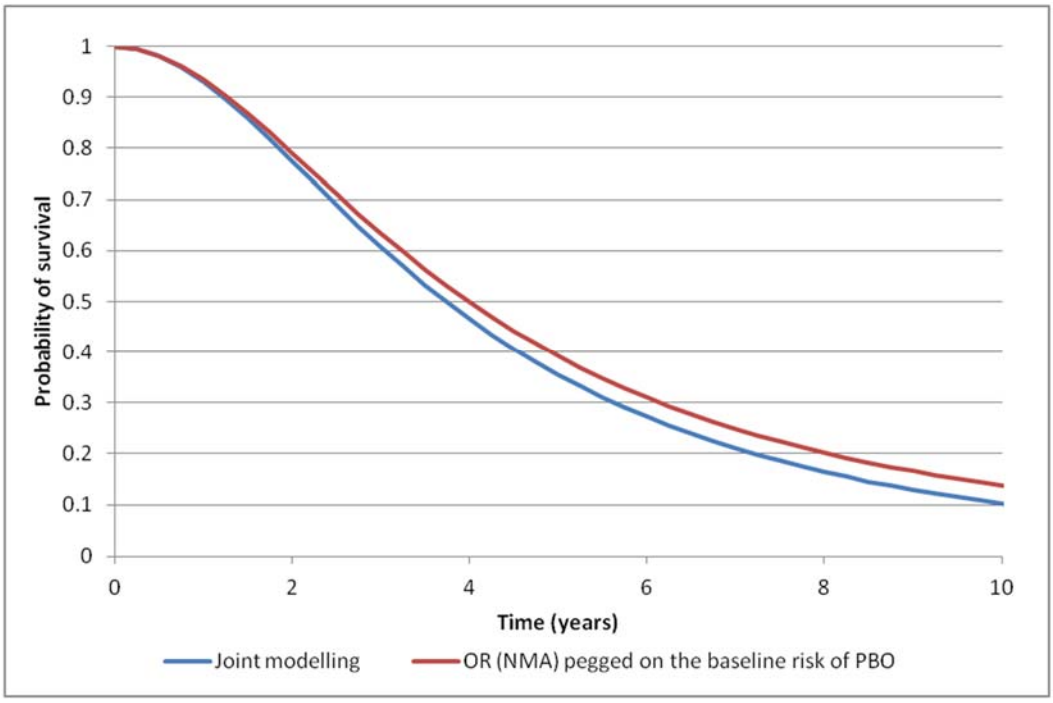


Figure 76: Comparison of nintedanib survival for the Weibull model

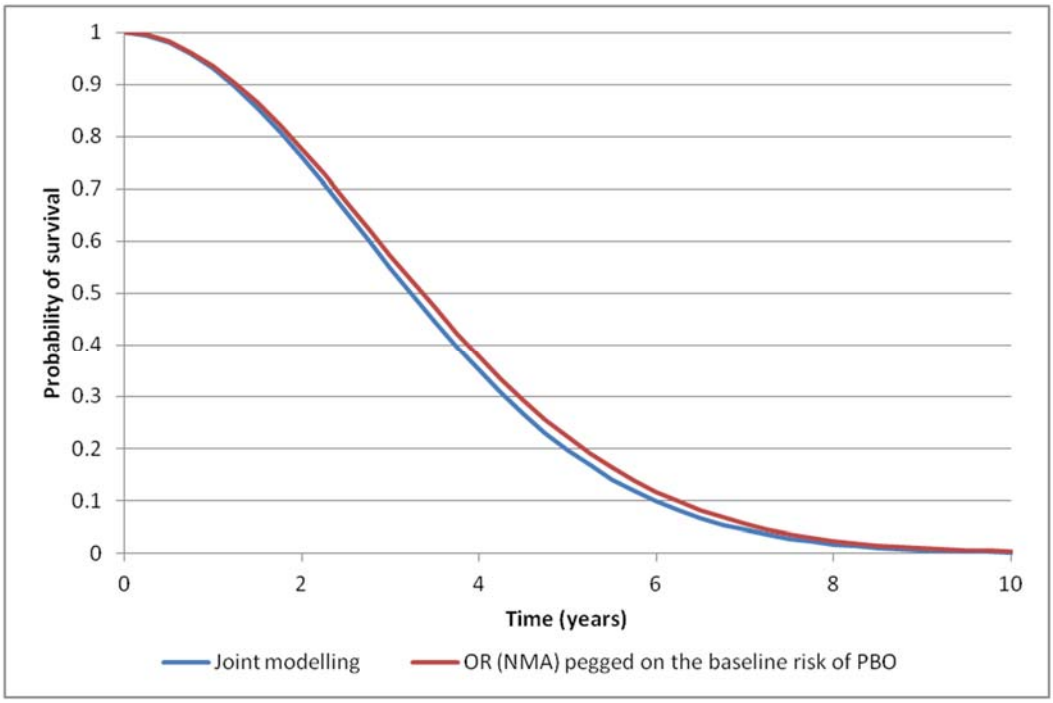


Figure 77: Comparison of nintedanib survival for the Gompertz model

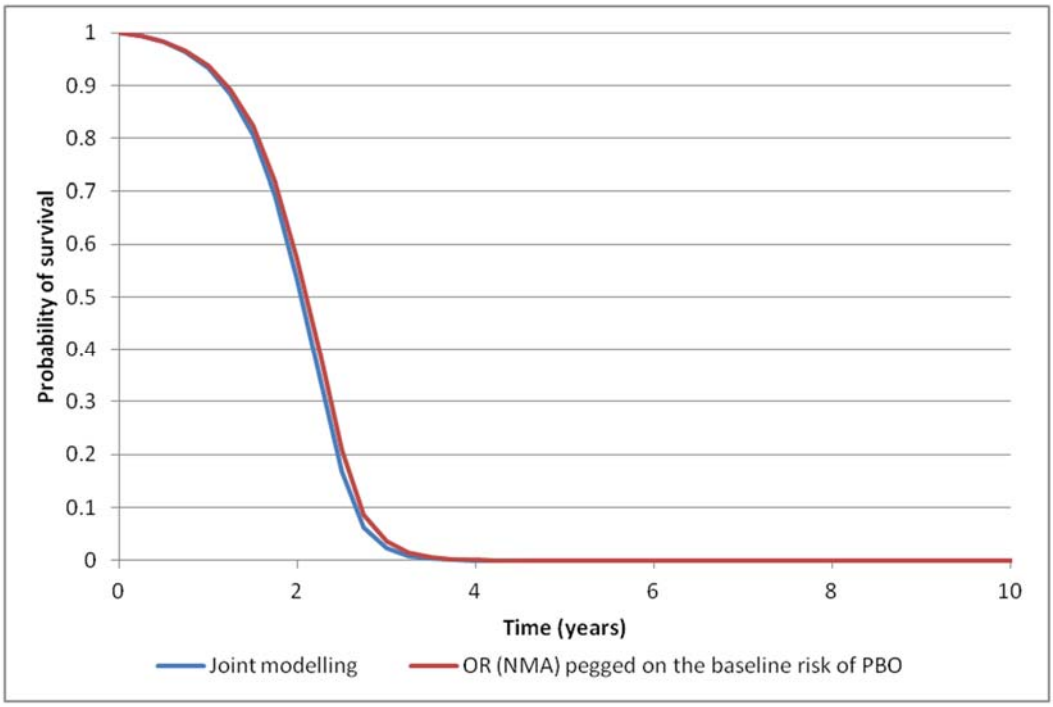


Table 190: Comparison of average survival time (nintedanib arm)

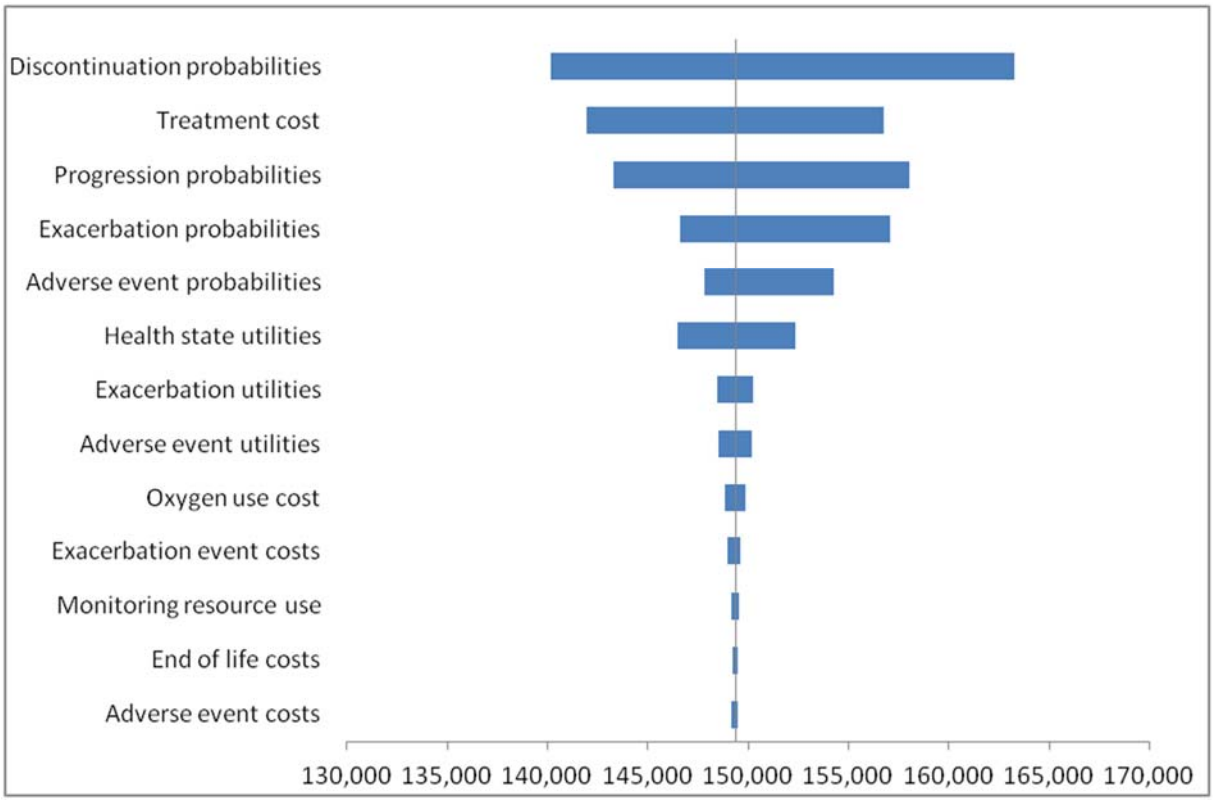
Model	Loglogistic	Weibull	Gompertz
Average survival time under joint population model	5.0 years	3.6 years	2.1 years
Average survival time under survival modelling via OR (NMA)	5.75 years	3.7 years	2.2 years

Table 191: Sensitivity analysis results (nintedanib vs. BSC)

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	*£149,360.88*	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	£87,246.22	BSC dominates
2		Exacerbation probabilities	£146,591.27	£157,062.84
3		Progression probabilities	£143,278.82	£158,034.61
4		Discontinuation probabilities	£140,192.71	£163,249.01
5	Costs	Treatment costs	£141,962.86	£156,764.70
6		Background follow-up costs	£149,178.14	£149,553.56
7		Oxygen use costs	£148,844.04	£149,877.68
8		Exacerbation event costs	£149,606.92	£149,004.11
9		EoL costs	£149,509.73	£149,211.99
10	Utilities	Health state baseline utilities	£152,328.36	£146,506.75
11		Exacerbation-related utilities	£148,485.13	£150,239.30
12	Adverse events	Adverse events probabilities	£147,851.27	£154,255.00
13		Adverse events costs	£149,196.65	£149,476.14
14		Adverse events related utilities	£150,217.68	£148,513.74
Tests on other model parameters				
Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
15	Overall survival	Baseline risk: Weibull parametric model	£240,092.88	
16		Baseline risk: Gompertz parametric model	£468,834.61	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	£151,738.22	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	£150,274.70	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	£146,205.31	
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	£150,274.70	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	£150,274.70	
22	Exacerbation	Baseline risk: use adjudication committee estimate	£149,528.72	
23		Baseline risk: exclude recurrent exacerbation risk	£149,498.31	
24		Relative risk: NMA results, scenario 1, all evidence	£149,380.35	
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	£149,380.35	
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	£150,547.65	
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	£149,380.35	
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	£149,380.35	

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
29	Loss of lung function	Baseline risk: include exacerbation coefficient	£146,628.44	
30		Relative risk: NMA results, scenario 1, all evidence	£149,597.13	
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	£149,061.52	
32		Relative risk: NMA results, scenario 4, include death (without Noble)	£149,597.13	
33		Relative risk: NMA results, scenario 5, include death (with Noble)	£149,597.13	
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	£148,884.56	
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	£149,173.86	
36		SAE disutility value: use alternative value for serious cardiac events	£149,193.00	
37		SAE disutility value: use alternative value for GI perforation	£149,301.91	
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	£150,127.78	
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	N/A for this comparison	
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	£149,429.48	
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	£148,914.35	
42		Relative risk: NMA results, scenario 4, excluding King	£149,429.48	
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	£149,257.93	
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	£149,818.88	
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	£148,851.34	
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	N/A for this comparison	
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	£194,458.59	

Figure 78: Tornado of nintedanib vs. BSC excluding mortality probabilities



5.8.10 Summary of sensitivity analyses results

See sections 5.8.1 to 5.8.9 for detailed description of sensitivity results in the text.

In the comparison between nintedanib and pirfenidone, nintedanib dominates pirfenidone in all one-way sensitivity analyses except when the stopping rule for pirfenidone is applied (patients discontinue treatment if they decline by >10% FVC in one year), for which the ICER is £82,784.36. The reason for this is that the reduction in costs of stopping pirfenidone is relatively greater than the reduction in QALYs, thus increasing pirfenidone's cost effectiveness relative to nintedanib.

In the comparison between nintedanib and best supportive care, the mortality risk estimates were very strong drivers of the model results; the ICER ranged from £87,000 to BSC dominating the NDB treatment. This was mainly due to the effect of mortality on the total and incremental QALYs. Given that the base case QALYs gained were relatively low, any small change in the QALYs could have a big impact on the ICER. When the mortality risk test adopted the low limit of the 95% confidence interval of the NMA OR values used for OS, the difference in incremental QALYs between the two strategies increased, and the ICER decreased by about £62,000 per QALY gained. In addition, changing the baseline OS risk from loglogistic (base-case) to a Weibull parametric model (scenario 15) had a significant impact on the result; the ICER increased by approximately £91,000. Similarly, changing the baseline risk to a Gompertz parametric model (scenario 16), affected the ICER even further, by about £320,000 per QALY gained.

5.9 *Subgroup analysis*

No subgroup analysis was undertaken.

5.10 *Validation*

Validation of de novo cost-effectiveness analysis

A number of steps were taken to ensure that the analysis was validated. These include:

- External review by leading UK clinical experts (see Section 5.2.2)
  - The model structure was discussed with leading clinicians. This clinical validation serves to ensure that the model adheres to the clinical course of the disease and is reflective of current clinical practice.
  - Sensitivity analysis outlined in Section 5.8.
- Validation by model developers
  - Apart from the advisory board with the UK clinical experts (discussed in Section 5.2.2), a senior modeller within the model developers organisation (with no involvement in the nintedanib model's development) performed a detailed QA check on the model.
- Validation by manufacturer

- This involved increasing and decreasing various parameters or changing assumptions in the model and then monitoring the impact on outputs. If the outputs were unexpected, further checks were made to determine whether this was the result of an error in the model.

5.11 Interpretation and conclusions of economic evidence

Results were generated from both pairwise comparisons with pirfenidone, and BSC and multiple, incremental comparisons. The results were generated in this way to demonstrate the cost effectiveness of nintedanib compared with pirfenidone, which is currently the only active treatment accepted for use in England. This is therefore the most relevant result when considering the cost effectiveness of nintedanib. Nintedanib at list price and at discounted price dominated pirfenidone (more QALYs and lower costs) at its list price, and this result was consistent in the majority of sensitivity analyses. Scenario analyses investigating theoretical discounts for pirfenidone, demonstrated nintedanib without PAS implemented dominated pirfenidone with a 0% discount of pirfenidone; at a 5% pirfenidone discount, nintedanib's ICER vs. pirfenidone was £13,663.45. Nintedanib and pirfenidone were largely equivalent in estimated costs and QALYs, and the small incremental QALY gain leads the results to be sensitive to a change in incremental costs.

In the pairwise comparison with BSC, nintedanib's (discounted price) ICER was [redacted].

In the comparison with BSC, the strongest driver was the calculation of patient mortality. The choice of a survival model for the baseline mortality risk (loglogistic, Weibull or Gompertz), played an important role in the cost-effectiveness results.

In the full incremental cost-effectiveness comparison, PFN was again dominated by NDB.

The data used to inform the model used in the submission is derived from the INPULSIS (Richeldi, 2014[4]) and TOMORROW (Richeldi *et al.*, 2011[1]) trials which enrolled patients matching the population for which this submission is made, and is reflective of the English population. Given that INPULSIS and TOMORROW only had two treatment arms (nintedanib and placebo), indirect evidence for pirfenidone and NAC has been generated from a network meta-analysis designed to incorporate patients relevant to the population for which this submission is made. Extensive scenario analyses were carried out to ensure the analysis was robust.

The HRQL data used in the model for nintedanib and BSC are taken from the phase III trials (Ingelheim, 2014[97]) and are therefore particularly relevant to the target population; it is the HRQL experienced by patients in relevant health states receiving the treatment being investigated.

Resource use within the model was derived from the INPULSIS trials (Ingelheim, 2014[97]) and thus directly relates to the population for which this submission is made.

Unit costs of procedures and hospital stays within the model are derived from the National Schedule of Reference Costs (Department of Health, 2013[8]) and The Personal Social Services Research unit (Curtis L, 2013[137], Curtis L, 2010[138]), and are thus applicable to the English population. The methodology followed for costing each of the inputs provided a sufficient level of granularity to inform cost parameters of the English population within the cost effectiveness model.

6 Assessment of factors relevant to the NHS and other parties

6.1 The purpose of this section is to provide an analysis of any factors relevant to the NHS and other parties that may fall outside the remit of the assessments of clinical and cost effectiveness. This will allow subsequent evaluation of the budget impact analysis. Such factors might include issues relating to service organisation and provision, resource allocation and equity, societal or ethical issues, plus any impact on patients or carers.

Provide the information specified in sections 6.2–6.10.

6.2 State how many people are eligible for treatment in England. Present results for the full marketing authorisation or CE marking and for any subgroups considered. Also present results for the subsequent 5 years.

The estimated epidemiology of IPF and treatment rates with current modalities in England from 2015 to 2019 are presented in Table 192.

The population estimates for England are obtained from the Quality and Outcomes Framework 2013-14 (Health and Social Care information Centre (HSCIC), [142]). The prevalence of 0.0230% for IPF in England is derived from a BI forecasting guidance document developed in 2014 and this is in line with the estimation by Navatnaram *et al.* and the one used in NICE Single technology appraisal (STA) TA282 (NICE, 2014[10]).

Internal market share assumed that the treatment rate for IPF is 26.4% (Boehringer Ingelheim, 2014[136]). Based on market research for pirfenidone in Germany, which indicated a good adherence rate but also a high treatment discontinuation rate, it is estimated that the compliance rate for IPF treatment is 62% (Boehringer Ingelheim, 2014[136]). The pirfenidone NICE STA assumed that 40% of patients were not eligible for treatment due to contraindication or because the patient did not wish to receive therapy, therefore 60% of patients are eligible for treatment. As a result, a total of 1,272 patients are expected to be eligible for nintedanib treatment. Since there is no population growth assumed, the eligible population remains constant from 2015 to 2019.

Table 192: Estimated number of patients eligible for nintedanib treatment

	England	
	Proportion	Number of patients
Population		56,324,887
Patients with IPF	0.0230%	12,955
Treatment Rate	26.4%	3,420
Compliance Rate	62.0%	2,120
Treatment eligibility	60.0%	1,272

6.3 Explain any assumptions that were made about current treatment options and uptake of technologies.

Current treatment options are pirfenidone or best supportive care and there is no treatment discontinuation assumed. After the introduction of nintedanib, it was assumed that the breakdown of treatment uptake will follow internal market share assumption, as discussed in section 6.4.

6.4 When relevant, explain any assumptions that were made about market share in England.

With the share of 1,272 patients eligible for nintedanib treatment, a proportion uptake is envisaged under two scenarios: the existing treatment scenario (without nintedanib) and a new treatment scenario (with nintedanib in the expected mix of treatments).

In the existing treatment scenario, it is assumed that 70% of patients will receive pirfenidone and the remaining 30% will be managed by best supportive care (Table 193). Once nintedanib is introduced into the market, internal market share assumptions predict that the market share of nintedanib will increase from 0% in Year 1 to 40% in Year 5 (Table 194).

Table 193: Existing treatment scenario (without nintedanib)

Treatment	Year 1	Year 2	Year 3	Year 4	Year 5
NDB	0%	0%	0%	0%	0%
BSC	30%	30%	30%	30%	30%
PFN	70%	70%	70%	70%	70%

Table 194: New treatment scenario (with nintedanib in the expected mix of treatments)

Treatment	Year 1	Year 2	Year 3	Year 4	Year 5
NDB	0%	14%	28%	35%	40%
BSC	30%	26%	22%	20%	18%
PFN	70%	60%	50%	46%	42%

6.5 In addition to technology costs, please consider other significant costs associated with treatment that may be of interest to commissioners (for example, administration costs, monitoring costs and the costs of managing adverse reactions).

No additional costs are included in this budget impact analysis as the only test needed for nintedanib treatment is liver function test, which is also required for treatment with pirfenidone.

6.6 State what unit costs were assumed and how they were calculated. If unit costs used in health economic modelling were not based on national reference costs or the payment-by-results tariff, explain how a cost for the activity was calculated.

The drug acquisition and intervention costs in this section are identical to those assumed in the cost-effectiveness evaluation in Section 5. Table 195 presents a summary of the drug costs per day for each comparator.

Table 195: Costs of drug and intervention

	Cost of drug/intervention per day
Nintedanib	£71.10
Pirfenidone	£71.10
Best supportive care	£0

6.7 If there were any estimates of resource savings, explain what they were and when they are likely to be made.

There are no additional estimates of resource savings.

6.8 State the estimated annual budget impact on the NHS in England.

The annual budget impact after introducing nintedanib for the next five years is summarised in Table 196. As a patient access scheme is also proposed for nintedanib, the annual budget impact for the next five years taking account into the nintedanib discount are summarised in Table 197.

Table 196: Summary of budget impact (based on list price)

	Year 1	Year 2	Year 3	Year 4	Year 5
Total Cost for existing treatment scenario (without nintedanib)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total Cost for new treatment scenario (nintedanib in the mix of treatment)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Net Budget Impact	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 197: Summary of budget impact (based on discounted price under PAS)

	Year 1	Year 2	Year 3	Year 4	Year 5
Total Cost for existing treatment scenario (without nintedanib)	£21,708,418	£38,444,188	£50,310,110	£58,390,738	£63,837,493
Total Cost for new treatment scenario (nintedanib in the mix of treatment)	£21,708,418	£37,329,243	£47,035,168	£52,656,832	£55,674,231
Net Budget Impact	-	-£1,114,945	-£3,274,942	-£5,733,906	-£8,163,262

6.9 Identify any other opportunities for resource savings or redirection of resources that it has not been possible to quantify.

There are no additional resource savings or redirection of resources expected.

6.10 Highlight the main limitations within the budget impact analysis.

The main limitation is that the discount offered by the manufacturer of pirfenidone under the patient access scheme is confidential. As such, the list price of pirfenidone is used for this budget impact analysis.

7 References

1. Richeldi L, Costabel U, Selman M, Kim D, Hansell D, Nicholson A, et al. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. New Engl J Med. 2011;365(12):1079-87.
2. Hilberg F, Roth GJ, Krssak M, Kautschitsch S, Sommergruber W, Tontsch-Grunt U, et al. BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. Cancer research. 2008;68(12):4774-82.
3. Wollin L, Maillet I, Quesniaux V, Holweg A, Ryffel B. Antifibrotic and anti-inflammatory activity of the tyrosine kinase inhibitor nintedanib in experimental models of lung fibrosis. The Journal of pharmacology and experimental therapeutics. 2014;349(2):209-20.
4. Richeldi LdB, R. M. Raghu, G. Azuma, A. Brown, K. K. Costabel, U. Cottin, V. Flaherty, K. R. Hansell, D. M. Inoue, Y. Kim, D. S. Kolb, M. Nicholson, A. G. Noble, P. W. Selman, M. Taniguchi, H. Brun, M. Le Maulf, F. Girard, M. Stowasser, S. Schlenker-Herceg, R. Disse, B. Collard, H. R. Inpulsis Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. New Engl J Med. 2014;370(22):2071-82.
5. Brun M, Sroka-Saidi K, Girard M, Marzin K. A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF) (1199.32). [Clinical trial report]. In press 2014.
6. Brun M, Sroka-Saidi K, Girard M, Marzin K. A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF) (1199.34). [Clinical trial report]. In press 2014.
7. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med. 2011;184(12):1382-9.
8. Department of Health. National Health Service Reference Costs. 2013.
9. Electronic Medicines Compendium (eMC). Pirfenidone Accessed May 2014. Available from: https://www.medicines.org.uk/emc/.
10. Idiopathic Pulmonary Fibrosis - Pirfenidone review decision https://www.nice.org.uk/guidance/ta282/documents/idiopathic-pulmonary-fibrosis-pirfenidone-review-decision-october-2014 [press release]. 2014.
11. Kolb M RL, Kimura T, et al. Effect of baseline FVC on decline in lung function with nintedanib in patients with IPF: results from the INPULSIS trials. . Submitted to American Thoracic Society2015.
12. Richeldi L, Bois RD, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy And Safety Of Nintedanib In Patients With Idiopathic Pulmonary Fibrosis: Results Of Two 52-Week, Phase III, Randomized, Placebo-Controlled Trials (INPULSIS). A95 SKYFALL: LATE BREAKING CLINICAL TRIALS IN IDIOPATHIC PULMONARY FIBROSIS. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2014. p. A6603-A.
13. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824.

14. Swigris J, Fairclough D, Morrison M, Make B, Kozora E, Brown K, et al. Benefits of pulmonary rehabilitation in idiopathic pulmonary fibrosis. Respir Care. 2011;56(6):783-9.
15. Rafii R, Juarez MM, Albertson TE, Chan AL. A review of current and novel therapies for idiopathic pulmonary fibrosis. Journal of thoracic disease. 2013;5(1):48-73.
16. Brun M, Luedtke D, Kluglich M, Juhel N, Roulot S. A 52 week, double blind, randomized, placebo-controlled trial evaluating the effect of BIBF 1120 administered at oral doses of 50 mg qd, 50 mg bid, 100 mg bid and 150 mg bid on Forced Vital Capacity decline during one year, in patients with Idiopathic Pulmonary Fibrosis, with optional active treatment extension until last patient out. [Clinical trial report]. In press 2012.
17. NICE. Pirfenidone for treating idiopathic pulmonary fibrosis. NICE technology appraisal guidance 282. 2013.
18. NICE. Idiopathic pulmonary fibrosis: The diagnosis and management of suspected idiopathic pulmonary fibrosis. NICE clinical guideline 163. 2013.
19. Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. Orphanet J Rare Dis. 2008;3:8.
20. Navaratnam V, Fleming KM, West J, Smith CJ, Jenkins RG, Fogarty A, et al. The rising incidence of idiopathic pulmonary fibrosis in the U.K. Thorax. 2011;66(6):462-7.
21. NICE. Quality Standard QS79. 2013.
22. NICE. Pathway for Idiopathic Pulmonary Fibrosis.: http://pathways.nice.org.uk/pathways/idiopathic-pulmonary-fibrosis. 2013.
23. Society BT. Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. Issued 2008. https://www.brit-thoracic.org.uk/document-library/clinical-information/interstitial-lung-disease/ild-guidelines/bts-interstitial-lung-disease-guideline/.
24. Dissemination CfRa. CRD's guidance for undertaking reviews in health care. Centre for Review Dissemination [Internet]. 2009. Available from: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf.
25. Agarwal R, Jindal SK. Acute exacerbation of idiopathic pulmonary fibrosis: A systematic review. Eur J Intern Med. 2008;19(4):227-35.
26. Richeldi L. Assessing the treatment effect from multiple trials in idiopathic pulmonary fibrosis. European Respiratory Review. 2012;21(124):147-51.
27. Spagnolo PDG, C. Luppi, F. Cerri, S. Balduzzi, S. Walters, E. H. D'Amico, R. Richeldi, L. Non-steroid agents for idiopathic pulmonary fibrosis. Cochrane database of systematic reviews (Online). 2010;9:CD003134.
28. A 12 Month, Double Blind, Randomized, Placebo-controlled Trial Evaluating the Effect of BIBF 1120 Administered at Oral Doses of 50 mg qd, 50 mg Bid, 100 mg Bid and 150 mg Bid on Forced Vital Capacity Decline During One Year, in Patients With Idiopathic Pulmonary Fibrosis, With Optional Active Treatment Extension Until Last Patient Out [NCT00514683] [Internet]. 2007 [cited 2014 08 15]. Available from: http://ClinicalTrials.gov/show/NCT00514683.
29. Richeldi LB, K. K. Costabel, U. Flaherty, K. R. Kim, D. Noble, P. W. Raghu, G. Selman, M. Gupta, A. Brun, M. Kluglich, M. Juhe, N. Du Bois, R. M. The oral triple kinase inhibitor BIBF 1120 reduces decline in lung function in patients with Idiopathic Pulmonary Fibrosis (IPF): Results from the tomorrow study. Am J Respir Crit Care Med. 2011;183 (1 MeetingAbstracts).

30. Richeldi LB, K. K. Costabel, U. Flaherty, K. R. Kim, D. S. Noble, P. W. Treatment With BIBF 1120 Reduces Acute Exacerbations And Improves Quality Of Life In Patients With IPF: Results From The Tomorrow Study [Abstract]. Am J Respir Crit Care Med [Internet]. 2011; 183(Meeting Abstracts):[A3809 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/329/CN-00797329/frame.html.
31. Rangarajan S, Kurundkar A, Zhang J, Thannickal VJ. Non-Canonical Autophagy And The Anti-Fibrotic Effects Of BIBF-1120 In IPF. D17 STOP, DROP, AND DIE: PROTEOSTASIS AND AUTOPHAGY IN THE LUNG. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2014. p. A5395-A.
32. A 52 Weeks, Double Blind, Randomized, Placebo-controlled Trial Evaluating the Effect of Oral BIBF 1120, 150 mg Twice Daily, on Annual Forced Vital Capacity Decline, in Patients With Idiopathic Pulmonary Fibrosis (IPF) [NCT01335477] [Internet]. 2011 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01335477.
33. A 52 Weeks, Double Blind, Randomized, Placebo-controlled Trial Evaluating the Effect of Oral BIBF 1120, 150 mg Twice Daily, on Annual Forced Vital Capacity Decline, in Patients With Idiopathic Pulmonary Fibrosis (IPF) [NCT01335464] [Internet]. 2011 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01335464.
34. Inoue YA, A. Taniguchi, H. Ogura, T. Tadayasu, Y. Fujimoto, T. Klueglich, M. Sugiyama, Y. Nukiwa, T. The pharmacokinetics of BIBF 1120 alone or in combination with pirfenidone in Japanese patients with idiopathic pulmonary fibrosis (IPF). Respirology Conference: 16th Congress of the Asian Pacific Society of Respirology Shanghai China Conference Start: 20111103 Conference End: 20111106 Conference Publication: (varpagings) [Internet]. 2011; 16(Suppl 2):[318 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/787/CN-00833787/frame.html.
35. A Double-blind, Randomised, Placebo-controlled (Within a Dose Group) Study to Evaluate Safety and Pharmacokinetics of Multiple Rising Doses of BIBF 1120 at 50 mg Bid (14 Days), 100 mg Bid (14 Days), and 150 mg Bid (28 Days) p.o., on Top of Standard Medical Care With Stratification According to Pirfenidone Use, in Japanese Patients With Idiopathic Pulmonary Fibrosis [NCT01136174] [Internet]. 2010 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01136174.
36. A Phase II Open Label, Roll Over Study of the Long Term Tolerability, Safety and Efficacy of Oral BIBF 1120 in Patients With Idiopathic Pulmonary Fibrosis [NCT01170065] [Internet]. 2010 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01170065.
37. A Double Blind Randomized Placebo Controlled Trial Evaluating the Effect of Oral Nintedanib 150 mg Twice Daily on High Resolution Computerized Tomography Quantitative Lung Fibrosis Score, Lung Function, Six Minute Walk Test Distance and St. George's Respiratory Questionnaire After Twelve Months of Treatment in Patients With Idiopathic Pulmonary Fibrosis With Continued Evaluations Over a Period of up to Eighteen Months [NCT01979952] [Internet]. 2013 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01979952.
38. A Phase II Open Label, Follow up Study to Investigate the Long Term Tolerability and Safety of Oral BIBF 1120 on Top of Pirfenidone in Japanese Patients With

Idiopathic Pulmonary Fibrosis [NCT01417156] [Internet]. 2011 [cited 2014 08 15]. Available from: http://ClinicalTrials.gov/show/NCT01417156.
39. An Open-label Extension Trial of the Long Term Safety of Oral BIBF 1120 in Patients With Idiopathic Pulmonary Fibrosis (IPF) [NCT01619085] [Internet]. 2012 [cited 2014 08 15]. Available from: http://clinicaltrials.gov/ct2/show/record/NCT01619085.
40. Noble PWA, C. Bradford, W. Z. Costabel, U. Glassberg, M. K. Kardatzke, D. King Jr, T. E. Lancaster, L. Sahn, S. A. Szwarcberg, J. Valeyre, D. Du Bois, R. M. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): Two randomised trials. The Lancet. 2011;377(9779):1760-9.
41. King TE, Bradford WZ, Castro-Bernadini S, Fagan EA, Glaspole I, Glassberg MK, et al. The ASCEND Study: A Randomized, Double-Blind, Placebo Controlled Trial Of Pirfenidone In Patients With Idiopathic Pulmonary Fibrosis (IPF). A95 SKYFALL: LATE BREAKING CLINICAL TRIALS IN IDIOPATHIC PULMONARY FIBROSIS. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2014. p. A6602-A.
42. Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. Thorax. 2013;68(4):309-10.
43. Alexeeff SE, Litonjua AA, Suh H, Sparrow D, Vokonas PS, Schwartz J. Ozone exposure and lung function: effect modified by obesity and airways hyperresponsiveness in the VA normative aging study. Chest. 2007;132(6):1890-7.
44. Griffith KA, Sherrill DL, Siegel EM, Manolio TA, Bonekat HW, Enright PL. Predictors of loss of lung function in the elderly: the Cardiovascular Health Study. Am J Respir Crit Care Med. 2001;163(1):61-8.
45. Costabel UB, E. Cottin, V. Dewint, P. Egan, J. J. J. Ferguson, J. Groves, R. Hellstrom, P. M. Kreuter, M. Maher, T. M. Molina-Molina, M. Nordlind, K. Sarafidis, A. Vancheri, C. Pirfenidone in idiopathic pulmonary fibrosis: Expert panel discussion on the management of drug-related adverse events. Advances in Therapy. 2014;31(4):375-91.
46. Cottin V TH, Richeldi L, Collard HR, Kay M, Hansell DM, Bailes Z, Schlenker-Herceg R, Raghu G;. Effect of baseline emphysema on reduction in FVC decline with nintedanib in the INPULSIS™ trials. ICLAF 2014 2014.
47. Sakae HS, Sakamoto;Keishi, Sugino;Yoko, Muramatsu;Kazutoshi, Isobe;Yujiro, Takai. Efficacy Of Pirfenidone Combined With N-Acetylcysteine Therapy In Patients With Advanced Idiopathic Pulmonary Fibrosis. A38 DIAMONDS ARE FOREVER BUT NEW TREATMENTS FOR INTERSTITIAL LUNG DISEASE CAN'T WAIT. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2014. p. A1417-A.
48. Azuma A, Nukiwa T, Tsuboi E, Suga M, Abe S, Nakata K, et al. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med [Internet]. 2005; 171(9):[1040-7 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/186/CN-00511186/frame.html http://www.atsjournals.org/doi/pdf/10.1164/rccm.200404-571OC.
49. Taniguchi HE, M. Kondoh, Y. Ogura, T. Azuma, A. Suga, M. Taguchi, Y. Takahashi, H. Nakata, K. Sato, A. Takeuchi, M. Raghu, G. Kudoh, S. Nukiwa, T. Pirfenidone in idiopathic pulmonary fibrosis. The European respiratory journal [Internet]. 2010; 35(4):[821-9 pp.]. Available from:

http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/639/CN-00751639/frame.html http://erj.ersjournals.com/content/35/4/821.full.pdf.
50. King TEB, W. Z. Castro-Bernardini, S. Fagan, E. A. Glaspole, I. Glassberg, M. K. Gorina, E. Hopkins, P. M. Kardatzke, D. Lancaster, L. Lederer, D. J. Nathan, S. D. Pereira, C. A. Sahn, S. A. Sussman, R. Swigris, J. J. Noble, P. W. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. The New England journal of medicine [Internet]. 2014; 370(22):[2083-92 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/630/CN-00991630/frame.html.
51. Research IPFC, Raghu G, Anstrom KJ, King TE, Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. New Engl J Med [Internet]. 2012; 366(21):[1968-77 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/398/CN-00814398/frame.html http://www.nejm.org/doi/pdf/10.1056/NEJMoa1113354.
52. Martinez FJ, de Andrade JA, Anstrom KJ, King TE, Jr., Raghu G, Idiopathic Pulmonary Fibrosis Clinical Research Network. Randomized trial of acetylcysteine in idiopathic pulmonary fibrosis. New Engl J Med. 2014;370(22):2093-101.
53. Tomioka HK, Y. Imanaka, K. Hashimoto, K. Ohnishi, H. Tada, K. Sakamoto, H. Iwasaki, H. A pilot study of aerosolized N-acetylcysteine for idiopathic pulmonary fibrosis. Respirology (Carlton, Vic) [Internet]. 2005; 10(4):[449-55 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/069/CN-00552069/frame.html http://onlinelibrary.wiley.com/doi/10.1111/j.1440-1843.2005.00725.x/abstract.
54. Homma SA, A. Taniguchi, H. Ogura, T. Mochiduki, Y. Sugiyama, Y. Nakata, K. Yoshimura, K. Takeuchi, M. Kudoh, S. Efficacy of inhaled N-acetylcysteine monotherapy in patients with early stage idiopathic pulmonary fibrosis. Respirology (Carlton, Vic) [Internet]. 2012; 17(3):[467-77 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/028/CN-00834028/frame.html http://onlinelibrary.wiley.com/store/10.1111/j.1440-1843.2012.02132.x/asset/j.1440-1843.2012.02132.x.pdf?v=1&t=i07yoawm&s=6c778ee54e64fa7d1930df61672c172b7002f6e3.
55. Taniguchi HK, Y. Ebina, M. Azuma, A. Ogura, T. Taguchi, Y. Suga, M. Takahashi, H. Nakata, K. Sato, A. Sugiyama, Y. Kudoh, S. Nukiwa, T. The clinical significance of 5% change in vital capacity in patients with idiopathic pulmonary fibrosis: extended analysis of the pirfenidone trial. Respiratory research [Internet]. 2011; 12:[93 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/783/CN-00798783/frame.html http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160381/pdf/1465-9921-12-93.pdf.
56. Azuma A, Taguchi Y, Ogura T, Ebina M, Taniguchi H, Kondoh Y, et al. Exploratory analysis of a phase III trial of pirfenidone identifies a subpopulation of patients with idiopathic pulmonary fibrosis as benefiting from treatment. Respiratory research [Internet]. 2011; 12:[143 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/864/CN-00833864/frame.html http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3216874/pdf/1465-9921-12-143.pdf.
57. Loveman E, Copley VR, Scott DA, Colquitt JL, Clegg AJ, O'Reilly KMA. Comparing new treatments for idiopathic pulmonary fibrosis – a network meta-analysis. BMC Pulmonary Medicine. 2015;15(37):e-pub ahead of print.

58. Martinez T, Pereira C, Dos Santos M, Ciconelli R, Guimaraes S, Martinez J. Evaluation of the short-form 36-item questionnaire to measure health- related quality of life in patients with idiopathic pulmonary fibrosis. Chest. 2000;117(6):1627-32.
59. EMA. Public summary of opinion on orphan designation. Nintedanib for the treatment of idiopathic pulmonary fibrosis. 2013.
60. Griffiths AL, Wurzel DF, Robinson PJ, Carzino R, Massie J. Australian epidemic strain pseudomonas (AES-1) declines further in a cohort segregated cystic fibrosis clinic. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society. 2012;11(1):49-52.
61. Collard HR, Moore BB, Flaherty KR, Brown KK, Kaner RJ, King TE, Jr., et al. Acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2007;176(7):636-43.
62. Pirfenidone new drug. Prescrire International. 2013;22(138):117-9.
63. Homma SAA, A. Taniguchi, H. Ogura, T. Mochizuki, Y. Sugiyama, Y. A Randomized, Double-blind, Multicentre, Controlled Trial Of Inhaled N-Acetylcysteine In Patients With The Early Stage Of Idiopathic Pulmonary Fibrosis In Japan [Abstract]. Am J Respir Crit Care Med [Internet]. 2010; 181(Meeting Abstracts):[A3986 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/501/CN-00755501/frame.html.
64. Homma SM, A. Sakamoto, S. An assessment of clinical efficacy of inhaled N-acetylcysteine in idiopathic pulmonary fibrosis. [Japanese]. Japanese Journal of Chest Diseases. 2005;64(2):99-105.
65. Cochrane Bias Methods Group. Assessing risk of bias in clinical studies. https://bmg.cochrane.org/assessing-risk-bias-included-studies. [Accessed October 2014].
66. Orphanos GS, Ioannidis GN, Ardavanis AG. Cardiotoxicity induced by tyrosine kinase inhibitors. Acta oncologica. 2009;48(7):964-70.
67. Gribbin J, Hubbard RB, Le Jeune I, Smith CJ, West J, Tata LJ. Incidence and mortality of idiopathic pulmonary fibrosis and sarcoidosis in the UK. Thorax. 2006;61(11):980-5.
68. Richeldi LC, V. Flaherty, K. R. Kolb, M. Inoue, Y. Raghu, G. Taniguchi, H. Hansell, D. M. Nicholson, A. G. Maulf, F. Stowasser, S. Collard, H. R. Design of the INPULSIS trials: Two phase 3 trials of nintedanib in patients with idiopathic pulmonary fibrosis. Respir Med [Internet]. 2014; 108(7):[1023-30 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/021/CN-00996021/frame.html http://www.resmedjournal.com/article/S0954-6111(14)00156-5/pdf.
69. Collard HR, King TE, Jr., Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2003;168(5):538-42.
70. Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, et al. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. Am J Respir Crit Care Med. 2003;168(5):531-7.
71. Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. Am J Respir Crit Care Med. 2003;168(5):543-8.
72. Mallick S. Outcome of patients with idiopathic pulmonary fibrosis (IPF) ventilated in intensive care unit. Respir Med. 2008;102(10):1355-9.

73. Economics filter [Internet]. York (UK): InterTASC Information Specialists' Subgroup [Internet]. 2009.
74. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press [Internet]. 2005.
75. Loveman E, Copley VR, Colquitt JL, Scott DA, Clegg AJ, Jones J, et al. The effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: systematic review, network meta-analysis and health economic evaluation. BMC pharmacology & toxicology. 2014;15:63.
76. National Institute for Health and Care Excellence. Diagnosis and management of suspected idiopathic pulmonary fibrosis. Clinical Guideline. Methods, evidence and recommendations. Idiopathic pulmonary fibrosis: full guideline draft. January 2013.
77. Maher T JG. Personal Communication. 2014.
78. Alakhras M DP, Nadrous HF, Collazo-Clavell M, Ryu JH. Body mass index and mortality in patients with idiopathic pulmonary fibrosis. Chest. 2007;131(5):1448-53.
79. Antoniou KM HD, Rubens MB, Marten K, Desai SR, Siafakas NM, Nicholson AG, du Bois RM, Wells AU. Idiopathic pulmonary fibrosis: outcome in relation to smoking status. Am J Respir Crit Care. 2008;177(2):190-4.
80. Hamada K, Nagai S, Tanaka S, Handa T, Shigematsu M, Nagao T, et al. Significance of pulmonary arterial pressure and diffusion capacity of the lung as prognosticator in patients with idiopathic pulmonary fibrosis. Chest. 2007;131(3):650-6.
81. Lederer DJ, Arcasoy SM, Wilt JS, D'Ovidio F, Sonett JR, Kawut SM. Six-minute-walk distance predicts waiting list survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2006;174(6):659-64.
82. Lederer DJ AS, Barr RG, Wilt JS, Bagiella E, D'Ovidio F, Sonett JR, Kawut SM. Racial and ethnic disparities in idiopathic pulmonary fibrosis: A UNOS/OPTN database analysis. Am J Transplant. 2006;6(10):2436-42.
83. Lopes AJ CD, Mogami R, Lanzillotti RS, Melo PL, Jansen JM. Severity classification for idiopathic pulmonary fibrosis by using fuzzy logic. Clinics (Sao Paulo). 2011;66(6):1015-9.
84. Manali ED, Stathopoulos GT, Kollintza A, Kalomenidis I, Emili JM, Sotiropoulou C, et al. The Medical Research Council chronic dyspnea score predicts the survival of patients with idiopathic pulmonary fibrosis. Respir Med. 2008;102(4):586-92.
85. Nathan SD, Shlobin OA, Weir N, Ahmad S, Kaldjob JM, Battle E, et al. Long-term course and prognosis of idiopathic pulmonary fibrosis in the new millennium. Chest. 2011;140(1):221-9.
86. Zappala CJ LP, Nicholson AG, Colby TV, Cramer D, Renzoni EA, Hansell DM, du Bois RM, Wells AU. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. Eur Respir J. 2010;35(4):830-6.
87. Ley B RC, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, Poletti V, Buccioli M, Elicker BM, Jones KD, King TE Jr, Collard HR. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156(10):684-91.
88. Du Bois RMW, D.;Albera, C.;Bradford, W. Z.;Costabel, U.;Kartashov, A.;Lancaster, L.;Noble, P. W.;Raghu, G.;Sahn, S. A.;Szwarcberg, J.;Thomeer, M.;Valeyre, D.;King Jr, T. E. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;184(4):459-66.

89. NICE. Pirfenidone for treating idiopathic pulmonary fibrosis. 2013.
90. Chaudhuri N LC. A quarter of IPF patients not eligible for pirfenidone treatment due to the NICE criteria significantly decline over time. Thorax. 2014;69(Suppl 2):A218.
91. Boehringer Ingelheim. Advisory board minutes. 2014.
92. Boehringer Ingelheim. INPULSIS patient data. 2014.
93. Kondoh Y, Taniguchi H, Katsuta T, Kataoka K, Kimura T, Nishiyama O, et al. Risk factors of acute exacerbation of idiopathic pulmonary fibrosis. Sarcoidosis, vasculitis, and diffuse lung diseases : official journal of WASOG / World Association of Sarcoidosis and Other Granulomatous Disorders. 2010;27(2):103-10.
94. Raghu G, Brown KK, Bradford WZ, Starko K, Noble PW, Schwartz DA, et al. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. The New England journal of medicine. 2004;350(2):125-33.
95. Douglas WW, Ryu JH, Swensen SJ, Offord KP, Schroeder DR, Caron GM, et al. Colchicine versus prednisone in the treatment of idiopathic pulmonary fibrosis. A randomized prospective study. Members of the Lung Study Group. Am J Respir Crit Care Med. 1998;158(1):220-5.
96. FDA highlights of prescribing information. OFEV (nintedanib) capsules for oral use. Initial US Approval 2014. 2014.
97. Ingelheim B. data on file - phase III trial (trial no. 1199.32 and 1199.34) post-hoc analysis. 2014.
98. Papaioannou D, Brazier J, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. In: Excellence NIfHaC, editor TSD 92010. 2011.
99. Paisley S, BA., Mansinkai S. Chapter 12: Health-related Quality of Life Studies. Etext on Health Technology Assessment (HTA) Information Resources. 2005.
100. King TE, Jr., Brown KK, Raghu G, du Bois RM, Lynch DA, Martinez F, et al. BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2011;184(1):92-9.
101. Zisman D, Schwarz M, Anstrom K, Collard H, Flaherty K, Hunninghake G. A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. New Engl J Med. 2010;363(7):620-8.
102. Mishra A, Bhattacharya P, Paul S, Paul R, Swarnakar S. An alternative therapy for idiopathic pulmonary fibrosis by doxycycline through matrix metalloproteinase inhibition. Lung India. 2011;28(3):174-9.
103. King J, Behr J, Brown K, Du Bois R, Lancaster L, De Andrade J, et al. BUILD-1: A randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2008;177(1):75-81.
104. Lubin M, Chen H, Elicker B, Jones K, Collard H, Lee J. A comparison of health-related quality of life in idiopathic pulmonary fibrosis and chronic hypersensitivity pneumonitis. Chest. 2014;145(6):1333-8.
105. Lee R, Kelly E, Nolan G, Keane M, McNicholas W. Comparison of sleep and exercise gas exchange in patients with stable idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2010;181(1 MeetingAbstracts).
106. Zimmermann C, Carvalho C, Silveira K, Yamaguti W, Moderno E, Salge J, et al. Comparison of two questionnaires which measure the health-related quality of life of idiopathic pulmonary fibrosis patients. Braz J Med Biol Res. 2007;40(2):179-87.
107. Mermigkis C, Bouloukaki I, Antoniou K, Mermigkis D, Psathakis K, Giannarakis I, et al. CPAP therapy in patients with idiopathic pulmonary fibrosis and obstructive

sleep apnea: Does it offer a better quality of life and sleep? Sleep Breathing. 2013;17(4):1137-43.
108. Peng S, Li Z, Kang J, Hou X. Cross-sectional and longitudinal construct validity of the Saint George's Respiratory Questionnaire in patients with IPF. Respirology. 2008;13(6):871-9.
109. Swigris J, Wilson S, Green K, Sprunger D, Brown K, Wamboldt F. Development of the ATAQ-IPF: A tool to assess quality of life in IPF. Health Qual Life Outcomes. 2010;8.
110. Swigris JJS, A. L.;Gould, M. K.;Wilson, S. R. Patients' perspectives on how idiopathic pulmonary fibrosis affects the quality of their lives. Health Qual Life Outcomes. 2005;3:61.
111. Kozu R, Senjyu H, Jenkins S, Mukae H, Sakamoto N, Kohno S. Differences in response to pulmonary rehabilitation in idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease. Respiration. 2011b;81(3):196-205.
112. Baddini Martinez J, Martinez T, Lovetro Galhardo F, De Castro Pereira C. Dyspnea scales as a measure of health-related quality of life in patients with idiopathic pulmonary fibrosis. Med Sci Monit. 2002;8(6):CR405-CR10.
113. Kozu R, Jenkins S, Senjyu H. Effect of disability level on response to pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. Respirology. 2011c;16(8):1196-202.
114. Ozalevli S, Karaali H, Ilgin D, Ucan E. Effect of home-based pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. Multidiscip Resp Med. 2010;5(1):31-7.
115. Mishima T, Ogawa T, Watanabe F, Hirasawa J, Furukawa T, Tanaka Y, et al. Effects of in-patient pulmonary rehabilitation for advanced idiopathic pulmonary fibrosis. Respirology. 2013;18:136.
116. Nishiyama O, Kondoh Y, Kimura T, Kato K, Kataoka K, Ogawa T, et al. Effects of pulmonary rehabilitation in patients with idiopathic pulmonary fibrosis. Respirology. 2008;13(3):394-9.
117. Tzanakis N, Samiou M, Lambiri I, Antoniou K, Siafakas N, Bouros D. Evaluation of health-related quality-of-life and dyspnea scales in patients with idiopathic pulmonary fibrosis. Correlation with pulmonary function tests. Eur J Intern Med. 2005;16(2):105-12.
118. Han M, Swigris J, Liu L, Bartholmai B, Murray S, Giardino N, et al. Gender influences health-related Quality of Life in IPF. Respir Med. 2010;104(5):724-30.
119. Verma G, Marras T, Chowdhury N, Singer L. Health-related quality of life and 6 min walk distance in patients with idiopathic pulmonary fibrosis. Can Respir J. 2011;18(5):283-7.
120. Nishiyama O, Taniguchi H, Kondoh Y, Kimura T, Kataoka K, Nishimura K, et al. Health-related quality of life does not predict mortality in idiopathic pulmonary fibrosis. Sarcoidosis Vasc Diffuse Lung Dis. 2012;29(2):113-8.
121. Furukawa T, Taniguchi H, Kondoh Y, Kimura T, Kataoka K, Matsuda T, et al. Health-related quality of life impacts on survival in idiopathic pulmonary fibrosis. Respirology. 2013;18:32.
122. Antoniou K, Nicholson A, Dimadi M, Malagari K, Latsi P, Rapti A, et al. Long-term clinical effects of interferon gamma-1b and colchicine in idiopathic pulmonary fibrosis. Eur Respir J. 2006;28(3):496-504.
123. Jones RM, Hilldrup S, Hope-Gill BD, Eccles R, Harrison NK. Mechanical induction of cough in Idiopathic Pulmonary Fibrosis. Cough. 2011;7:2.

124. Key AL, Holt K, Hamilton A, Smith JA, Earis JE. Objective cough frequency in Idiopathic Pulmonary Fibrosis. Cough. 2010;6:4.
125. Tzouvelekis A, Paspaliaris V, Koliakos G, Ntolios P, Bouros E, Oikonomou A, et al. A prospective, non-randomized, no placebo-controlled, phase Ib clinical trial to study the safety of the adipose derived stromal cells-stromal vascular fraction in idiopathic pulmonary fibrosis. J Transl Med. 2013;11(1).
126. Jastrzebski D, Kozielski J, Banas A, Cebula T, Gumola A, Ziora D, et al. Quality of life during one-year observation of patients with idiopathic pulmonary fibrosis awaiting lung transplantation. J Physiol Pharmacol. 2005;56(SUPPL. 4):99-105.
127. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, et al. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. Respir Med. 2010;104(2):296-304.
128. Raghu G, Brown KK, Costabel U, Cottin V, du Bois RM, Lasky JA, et al. Treatment of idiopathic pulmonary fibrosis with etanercept: an exploratory, placebo-controlled trial. Am J Respir Crit Care Med. 2008;178(9):948-55.
129. Elfferich M, De Fries J, Drent M. Type D or 'distressed' personality in sarcoidosis and idiopathic pulmonary fibrosis. Sarcoidosis Vasc Diffuse Lung Dis. 2011;28(1):65-71.
130. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health. 2010;13(5):509-18.
131. Leontiadis GI, Sharma VK, Howden CW. Proton pump inhibitor therapy for peptic ulcer bleeding: Cochrane collaboration meta-analysis of randomized controlled trials. Mayo Clinic proceedings. 2007;82(3):286-96.
132. Cooper K SH, Carroll C, Papaioannou D, Booth A, Logan RF, et al. Chemoprevention of colorectal cancer: systematic review and economic evaluation. Health Technol Assess. 2010;14(32).
133. Parfrey HL, C.; Gibbons, M. A.; Armstrong, E.; Harris, E.; Frank, R.; Sharp, C.; Percival, F. Healthcare utilisation by patients with idiopathic pulmonary fibrosis; observations from the UK pirfenidone named patient programme. Thorax. 2013;68:A164-A5.
134. MIMS (Monthly Index of Medical Specialties). Printed version, published by Haymarket Business Media. March-May 2014.
135. Ingelheim B. data on file - phase III trial (trial no. 1199.32 and 1199.34) Summary of Clinical Safety 2014. Nintedanib (BIBF 1120). 2014.
136. Boehringer Ingelheim. Forecasting guidance document. 2014.
137. Curtis L. Unit Costs of Health and Social Care. 2013.
138. Curtis L. Unit Costs of Health and Social Care. 2010.
139. Hatziandreu E AF, Daly A, in conjuction with the National Audit Office. The potential cost savings of greater use of home- and hospice-based end of life care in England. Rand Europe. 2008.
140. Ingelheim B. data on file - phase III trial (trial no 1199.32 and 1199.34) Summary of Clinical Efficacy. Nintedanib (BIBF 1120). 2014.
141. King TE, Jr. Albera, C. Bradford, W. Z. Costabel, U. du Bois, R. M. Leff, J. A. Nathan, S. D. Sahn, S. A. Valeyre, D. Noble, P. W. All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials. Am J Respir Crit Care Med. 2014;189(7):825-31.
142. Health and Social Care information Centre (HSCIC). Quality and Outcomes Framework (QOF) for 2013 - 2014. Available at: http://www.hscic.gov.uk/catalogue/PUB15751/qof-1314-prev-ach-exc-region-nation.xlsx. 2014. .

8 Appendices

Appendix A.docx

Appendix B.docx

Appendix C.xlsx

Appendix D.docx

Appendix E.pdf

Appendix F.pdf

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Technology appraisals

Patient access scheme submission template

October 2009

1 Introduction

The 2009 Pharmaceutical Price Regulation Scheme (PPRS) (www.dh.gov.uk/en/Healthcare/Medicinespharmacyandindustry/Pharmaceuticalpriceregulationscheme/2009PPRS) is a non-contractual scheme between the Department of Health and the Association of the British Pharmaceutical Industry. The purpose of the 2009 PPRS is to ensure that safe and cost-effective medicines are available on reasonable terms to the NHS in England and Wales. One of the features of the 2009 PPRS is to improve patients' access to medicines at prices that better reflect their value through patient access schemes.

Patient access schemes are arrangements which may be used on an exceptional basis for the acquisition of medicines for the NHS in England and Wales. Patient access schemes propose either a discount or rebate that may be linked to the number, type or response of patients, or a change in the list price of a medicine linked to the collection of new evidence (outcomes). These schemes help to improve the cost effectiveness of a medicine and therefore allow the National Institute for Health and Care Excellence (NICE) to recommend treatments which it would otherwise not have found to be cost effective. More information on the framework for patient access schemes is provided in the 2009 PPRS (www.dh.gov.uk/en/Healthcare/Medicinespharmacyandindustry/Pharmaceuticalpriceregulationscheme/2009PPRS.

Patient access schemes are proposed by a pharmaceutical company and agreed with the Department of Health, with input from the Patient Access Schemes Liaison Unit (PASLU) within the Centre for Health Technology Evaluation at NICE.

2 Instructions for manufacturers and sponsors

This document is the patient access scheme submission template for technology appraisals. If manufacturers and sponsors want the National Institute for Health and Care Excellence (NICE) to consider a patient access scheme as part of a technology appraisal, they should use this template. NICE can only consider a patient access scheme after formal referral from the Department of Health.

The template contains the information NICE requires to assess the impact of a patient access scheme on the clinical and cost effectiveness of a technology, in the context of a technology appraisal, and explains the way in which background information (evidence) should be presented. If you are unable to follow this format, you must state your reasons clearly. You should insert 'N/A' against sections that you do not consider relevant, and give a reason for this response.

Please refer to the following documents when completing the template:

- 'Guide to the methods of technology appraisal' (http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9)
- 'Specification for manufacturer/sponsor submission of evidence' (http://www.nice.org.uk/aboutnice/howwework/devnicetech/singletechnologyappraisalsubmissiontemplates.jsp) and
- Pharmaceutical Price Regulation Scheme 2009 (www.dh.gov.uk/en/Healthcare/Medicinespharmacyandindustry/Pharmaceuticalpriceregulationscheme/2009PPRS).

For further details on the technology appraisal process, please see NICE's 'Guide to the single technology appraisal (STA) process' and 'Guide to the multiple technology appraisal (MTA) process' (http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/technology_appraisal_process_guides.jsp). The

‘Specification for manufacturer/sponsor submission of evidence’ provides details on disclosure of information and equality issues.

Make the submission as brief and informative as possible. Only mark information as confidential when absolutely necessary. Sufficient information must be publicly available for stakeholders to comment on the full content of the technology appraisal, including details of the proposed patient access scheme. Send submissions electronically to NICE in Word or a compatible format, not as a PDF file.

Appendices may be used to include additional information that is considered relevant to the submission. Do not include information in the appendices that has been requested in the template. Appendices should be clearly referenced in the main submission.

When making a patient access scheme submission, include:

- an updated version of the checklist of confidential information, if necessary
- an economic model with the patient access scheme incorporated, in accordance with the ‘Guide to the methods of technology appraisal’ (http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9).

If you are submitting the patient access scheme at the end of the appraisal process, you should update the economic model to reflect the assumptions that the Appraisal Committee considered to be most plausible. No other changes should be made to the model.

3 Details of the patient access scheme

3.1 Please give the name of the technology and the disease area to which the patient access scheme applies.

Ofev (nintedanib) for the treatment of adult patients with Idiopathic Pulmonary Fibrosis (IPF). This patient access scheme also applies to nintedanib under the brand name of Vargatef; nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent non-small cell lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy.

3.2 Please outline the rationale for developing the patient access scheme.

The patient access scheme has been developed in order to support the cost effectiveness case for Ofev.

3.3 Please describe the type of patient access scheme, as defined by the PPRS.

The scheme is a commercial in confidence simple discount patient access scheme (PAS). A confidential discount will be applied to the list price of Ofev.

3.4 Please provide specific details of the patient population to which the patient access scheme applies. Does the scheme apply to the whole licensed population or only to a specific subgroup (for example, type of tumour, location of tumour)? If so:

- How is the subgroup defined?
- If certain criteria have been used to select patients, why have these have been chosen?
- How are the criteria measured and why have the measures been chosen?

The PAS for Ofev applies to the entire licensed population. This patient access scheme also applies to nintedanib under the brand name of Vargatef;

nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent non-small cell lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy.

3.5 Please provide details of when the scheme will apply to the population specified in 3.4. Is the scheme dependent on certain criteria, for example, degree of response, response by a certain time point, number of injections? If so:

- Why have the criteria been chosen?
- How are the criteria measured and why have the measures been chosen.

The scheme is not dependent on any criteria and the discounted price will be reflected on all original invoices for the product.

3.6 What proportion of the patient population (specified in 3.4) is expected to meet the scheme criteria (specified in 3.5)?

100%

3.7 Please explain in detail the financial aspects of the scheme. How will any rebates be calculated and paid?

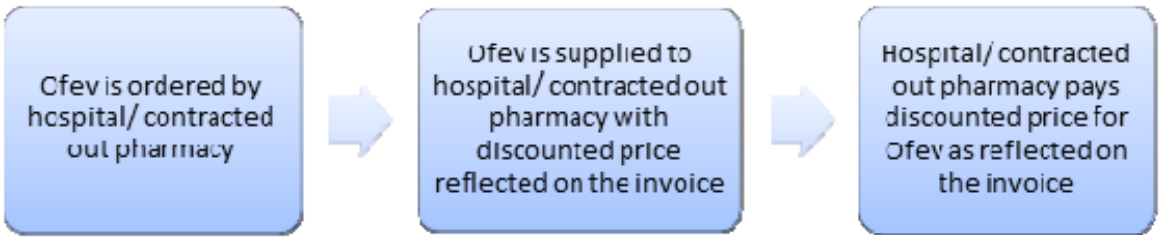
A fixed net price (which will not vary with any change to the UK list price) will apply to all packs of Ofev (nintedanib). The approved discounted price in the scheme will be the price paid by the NHS at the point of sale so there is no requirement for the calculation of rebates.

3.8 Please provide details of how the scheme will be administered. Please specify whether any additional information will need to be collected, explaining when this will be done and by whom.

NHS organisations will be required to sign a Confidential Disclosure Agreement to take part in the scheme. There are no associated administrative processes required with the scheme as stock for the product will be ordered in

the usual way and the approved discounted price will be paid at the point of sale by the NHS and will be reflected on all original invoices.

3.9 Please provide a flow diagram that clearly shows how the scheme will operate. Any funding flows must be clearly demonstrated.



3.10 Please provide details of the duration of the scheme.

The scheme will remain in place until NICE next reviews the guidance on the product.

3.11 Are there any equity or equalities issues relating to the scheme, taking into account current legislation and, if applicable, any concerns identified during the course of the appraisal? If so, how have these been addressed?

No issues have been identified by Boehringer Ingelheim Ltd in this regard.

3.12 If available, please list any scheme agreement forms, patient registration forms, pharmacy claim forms/rebate forms, guides for pharmacists and physicians and patient information documents. Please include copies in the appendices.

A Confidential Disclosure Agreement will need to be signed by NHS stakeholders before the discounted price can be shared.

3.13 In the exceptional case that you are submitting an outcome-based scheme, as defined by the PPRS, please also refer to appendix B.

4 Cost effectiveness

4.1 If the population to whom the scheme applies (as described in sections 3.4 and 3.5) has not been presented in the main manufacturer/sponsor submission of evidence for the technology appraisal (for example, the population is different as there has been a change in clinical outcomes or a new continuation rule), please (re-)submit the relevant sections from the 'Specification for manufacturer/sponsor submission of evidence' (particularly sections 5.5, 6.7 and 6.9). You should complete those sections both with and without the patient access scheme. You must also complete the rest of this template.

N/A

4.2 If you are submitting the patient access scheme at the end of the technology appraisal process, you should update the economic model to reflect the assumptions that the Appraisal Committee considered to be most plausible. No other changes should be made to the model.

N/A. The patient access scheme is submitted at the same time as the main submission.

4.3 Please provide details of how the patient access scheme has been incorporated into the economic model. If applicable, please also provide details of any changes made to the model to reflect the assumptions that the Appraisal Committee considered most plausible.

In the "Costs" tab, cell D8 allows the user to enter the discounted daily cost for nintedanib (NDB). As the patient access scheme is submitted at the same time as the main submission, there is no change made to the model at the time of preparing this document. The daily cost for nintedanib is changed to [redacted] in order to implement the PAS.

4.4 Please provide the clinical effectiveness data resulting from the evidence synthesis and used in the economic model which includes the patient access scheme.

The patient access scheme does not change the clinical data; the clinical data is the same as in the main submission document.

4.5 Please list any costs associated with the implementation and operation of the patient access scheme (for example, additional pharmacy time for stock management or rebate calculations). A suggested format is presented in table 1. Please give the reference source of these costs. Please refer to section 6.5 of the 'Specification for manufacturer/sponsor submission of evidence'.

N/A

Table 1 Costs associated with the implementation and operation of the patient access scheme (PAS)

	Calculation of cost	Reference source
Stock management		
Administration of claim forms		
Staff training		
Other costs…		
…		
…		
Total implementation/ operation costs		

4.6 Please provide details of any additional treatment-related costs incurred by implementing the patient access scheme. A suggested format is presented in table 2. The costs should be provided for the intervention both with and without the patient access scheme. Please give the reference source of these costs.

N/A

Table 2 Additional treatment-related costs for the intervention both with and without the patient access scheme (PAS)

	Intervention without PAS		Intervention with PAS		Reference source
	Unit cost (£)	**Total cost e.g. per cycle, per patient (£)**	**Unit cost (£)**	**Total cost e.g. per cycle, per patient (£)**	
Interventions					
Monitoring tests					
Diagnostic tests					
Appointments					
Other costs…					
…					
…					
Total treatment-related costs					

Summary results

Base-case analysis

4.7 Please present in separate tables the cost-effectiveness results as follows.[1]

- the results for the intervention without the patient access scheme
- the results for the intervention with the patient access scheme.

A suggested format is shown below (table 3).

[1] For outcome-based schemes, please see section 5.2.8 in appendix B.

4.7.1. Base-case analysis results - without patient access scheme discount

Table 3 Base-case cost-effectiveness results (Nintedanib versus pirfenidone) - without PAS

	Nintedanib	Pirfenidone
Intervention cost (£)	59,037.76	60,707.68
Other costs (£)	Adverse events: 781.97 Liver panel tests: 9.05 Patient monitoring and O2 use: 10,825.15 Acute exacerbation: 1,191.50 End of life: 13,242.07 Total: 26,049.73	Adverse events: 1,054.92 Liver panel tests: 9.31 Patient monitoring and O2 use: 10,953.78 Acute exacerbation: 1,501.72 End of life: 13,251.74 Total: 26,771.47
Total costs (£)	**85,087.49**	**87,479.15**
Difference in total costs (£)	N/A	***-2,391.65***
LYG	4.8593	4.8558
LYG difference	N/A	*0.0036*
QALYs	3.6720	3.6249
QALY difference	N/A	*0.0471*
ICER (£)	N/A	**Nintedanib dominates**

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Table 4 Base-case cost-effectiveness results (nintedanib versus best supportive care) - without PAS

	Nintedanib	Best supportive care (BSC)
Intervention cost (£)	59,037.76	0
Other costs (£)	Adverse events: 781.97 Liver panel tests: 9.05 Patient monitoring and O2 use: 10,825.15 Acute exacerbation: 1,191.50 End of life: 13,242.07 Total: 26,049.73	Adverse events: 625.98 Liver panel tests: 0 Patient monitoring and O2 use: 9,955.76 Acute exacerbation: 1,336.83 End of life: 13,440.51 Total: 25,359.09
Total costs (£)	**85,087.49**	**25,359.09**
Difference in total costs (£)	N/A	***59,728.40***
LYG	4.8593	4.3602
LYG difference	N/A	*0.4992*
QALYs	3.6720	3.2721
QALY difference	N/A	*0.3999*
ICER (£)	N/A	**149,360.88**

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Table 5: Full incremental cost-effectiveness analysis for all comparators – without PAS

Treatment	Total costs	Total QALYs	ICER
BSC	£25,359.09	3.2721	
PFN	£87,479.15	3.6249	Dominated by NDB
NDB	£85,087.49	3.6720	£149,360.88 per QALY gained

4.7.2. Base-case analysis results - with patient access scheme discount

Table 6 Base-case cost-effectiveness results for nintedanib versus pirfenidone - with PAS

	Nintedanib	Pirfenidone
Intervention cost (£)	[redacted]	[redacted]
Other costs (£)	[redacted]	[redacted]
Total costs (£)	[redacted]	[redacted]
Difference in total costs (£)	N/A	[redacted]
LYG	4.8593	4.8558
LYG difference	N/A	*0.0036*
QALYs	3.6720	3.6249
QALY difference	N/A	*0.0471*
ICER (£)	N/A	[redacted]

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Table 7 Base-case cost-effectiveness results for nintedanib versus best supportive care - with PAS

	Nintedanib	Best supportive care (BSC)
Intervention cost (£)	[redacted]	[redacted]
Other costs (£)	[redacted]	[redacted]
Total costs (£)	[redacted]	[redacted]
Difference in total costs (£)	N/A	[redacted]
LYG	4.8593	4.3602
LYG difference	N/A	*0.4992*
QALYs	3.6720	3.2721
QALY difference	N/A	*0.3999*
ICER (£)	N/A	[redacted]

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Table 8: Full incremental cost-effectiveness analysis for all comparators – with PAS

Treatment	Total costs	Total QALYs	ICER
BSC	£25,359.09	3.2721	
PFN	[redacted]	3.6249	[redacted]
NDB	£[redacted]	3.6720	[redacted]

4.8 Please present in separate tables the incremental results as follows. [2]

- the results for the intervention without the patient access scheme

[2] For outcome-based schemes, please see section 5.2.9 in appendix B.

- the results for the intervention with the patient access scheme.

List the interventions and comparator(s) from least to most expensive. Present the incremental cost-effectiveness ratios (ICERs) in comparison with baseline (usually standard care), and the incremental analysis ranking technologies in terms of dominance and extended dominance. A suggested format is presented in table 4.

Table 9 Base-case incremental results without PAS

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (QALYs)	ICER (£) incremental (QALYs)
BSC	25,359.09	4.3602	3.2721	-	-	-	-	-
Nintedanib	85,087.49	4.8593	3.6720	59,728.40	0.4992	0.3999	149,360.88	149,360.88
Pirfenidone	87,479.15	4.8558	3.6249	2,391.65	-0.0036	-0.0471	176,077.26	Dominated

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Table 10 Base-case incremental results with PAS

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) versus baseline (QALYs)	ICER (£) incremental (QALYs)
BSC	25,359.09	4.3602	3.2721	-	-	-	-	-
Nintedanib	[redacted]	4.8593	3.6720	[redacted]	0.4992	0.3999	[redacted]	[redacted]
Pirfenidone	[redacted]	4.8558	3.6249	[redacted]	-0.0036	-0.0471	[redacted]	[redacted]

LYG: life-year gained; QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio.

Sensitivity analyses

4.9 Please present deterministic sensitivity analysis results as described for the main manufacturer/sponsor submission of evidence for the technology appraisal. Consider using tornado diagrams.

The tornado diagrams for this section are run in the same way as described in the main submission document.

4.9.1. Deterministic sensitivity analysis results for nintedanib versus pirfenidone

1. Without patient access scheme discount

Deterministic sensitivity analysis results show that nintedanib dominates pirfenidone. Please note that a tornado diagram is not presented for this comparison because dominance cannot be represented in such a graph.

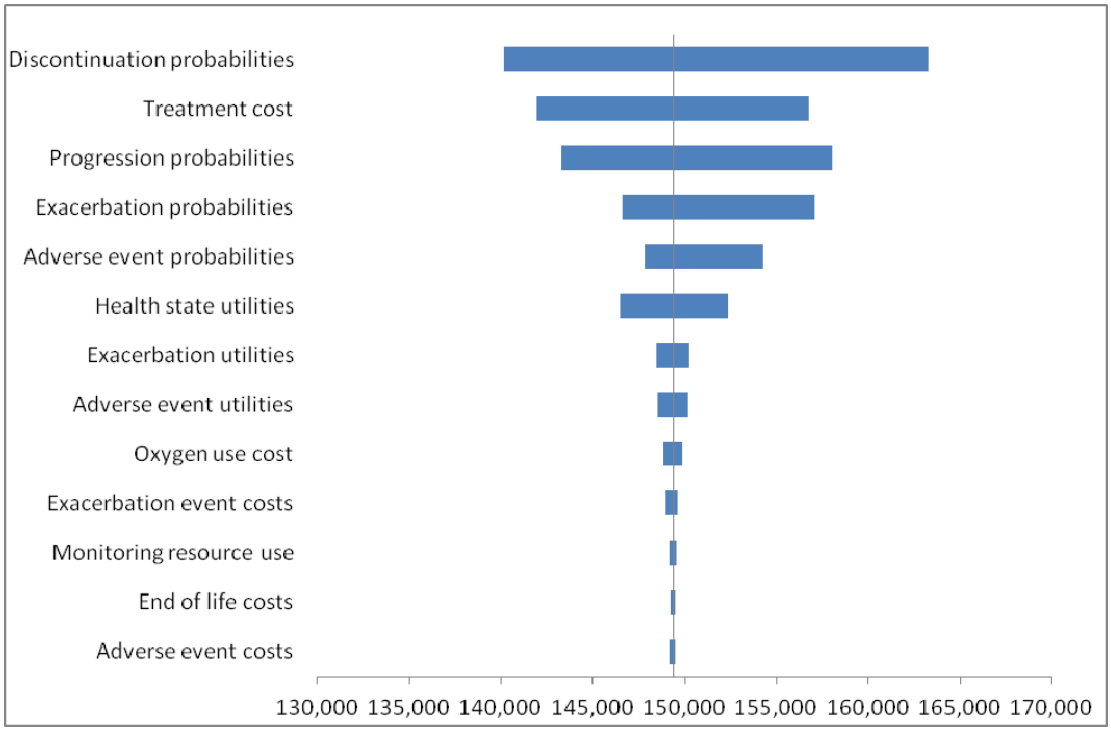
2. With patient access scheme discount

Deterministic sensitivity analysis results show that nintedanib dominates pirfenidone. Please note that a tornado diagram is not presented for this comparison because dominance cannot be represented in such a graph.

4.9.2. Deterministic sensitivity analysis results for nintedanib versus best supportive care (BSC)

1. Without patient access scheme discount

Figure 1: Tornado diagram of nintedanib versus BSC excluding the effect of mortality probabilities; without PAS



2. With patient access scheme discount

[redacted]

4.10 Please present any probabilistic sensitivity analysis results, and include scatter plots and cost-effectiveness acceptability curves.

4.10.1. Probabilistic sensitivity analysis results for nintedanib versus pirfenidone

1. Without patient access scheme discount

Table 11: Comparison of ICERs obtained from deterministic and probabilistic sensitivity analyses for nintedanib versus pirfenidone without PAS

	Incremental cost	Incremental QALY	ICER
Deterministic Values	-£2,391.65	0.0471	Nintedanib dominates
Average value for PSA	-£2,369.86	0.0600	Nintedanib dominates

Figure 3: Incremental cost-effectiveness scatter-plot (nintedanib versus pirfenidone) without PAS

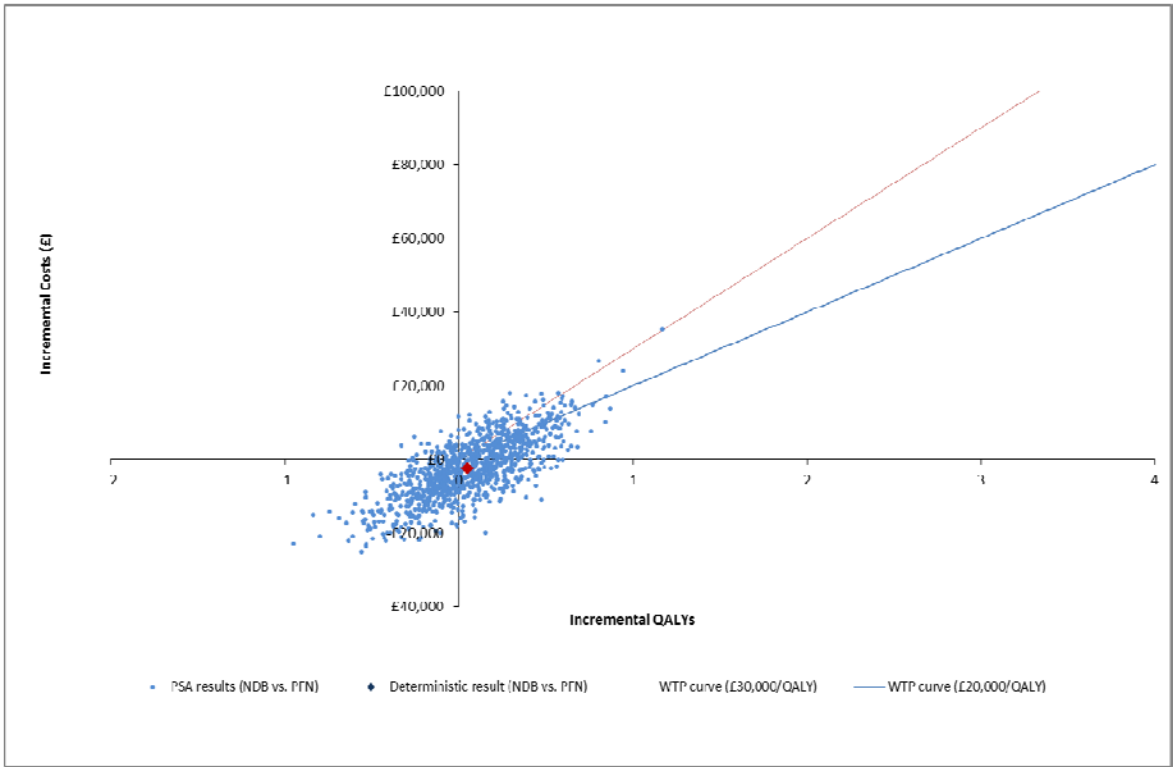
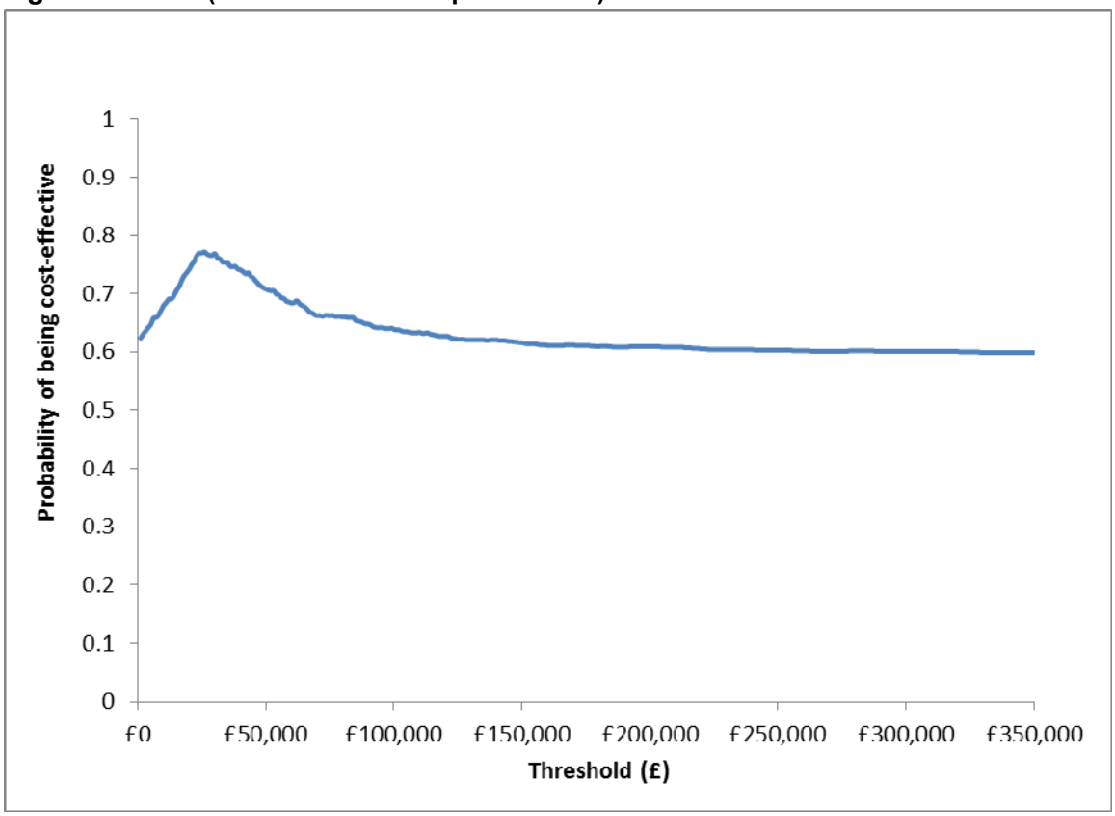


Figure 4: CEAC (nintedanib versus pirfenidone) without PAS



2. With patient access scheme discount

Table 12: Comparison of ICERs obtained from deterministic and probabilistic sensitivity analyses for nintedanib versus pirfenidone with PAS

	Incremental cost	Incremental QALY	ICER
Deterministic Values	[redacted]	0.0471	[redacted]
Average value for PSA		0.0640	

[redacted]

4.10.2. Probabilistic sensitivity analysis results for nintedanib versus best supportive care

Table 13: Comparison of ICERs obtained from deterministic and probabilistic sensitivity analyses for nintedanib versus best supportive care without PAS.

	Incremental cost	Incremental QALY	ICER
Deterministic Values	£59,728.40	0.3999	£149,360.88
Average value for	£59,568.76	0.4020	£148,180.99

PSA			

Figure 7: Incremental cost-effectiveness scatter-plot (nintedanib versus BSC) without PAS

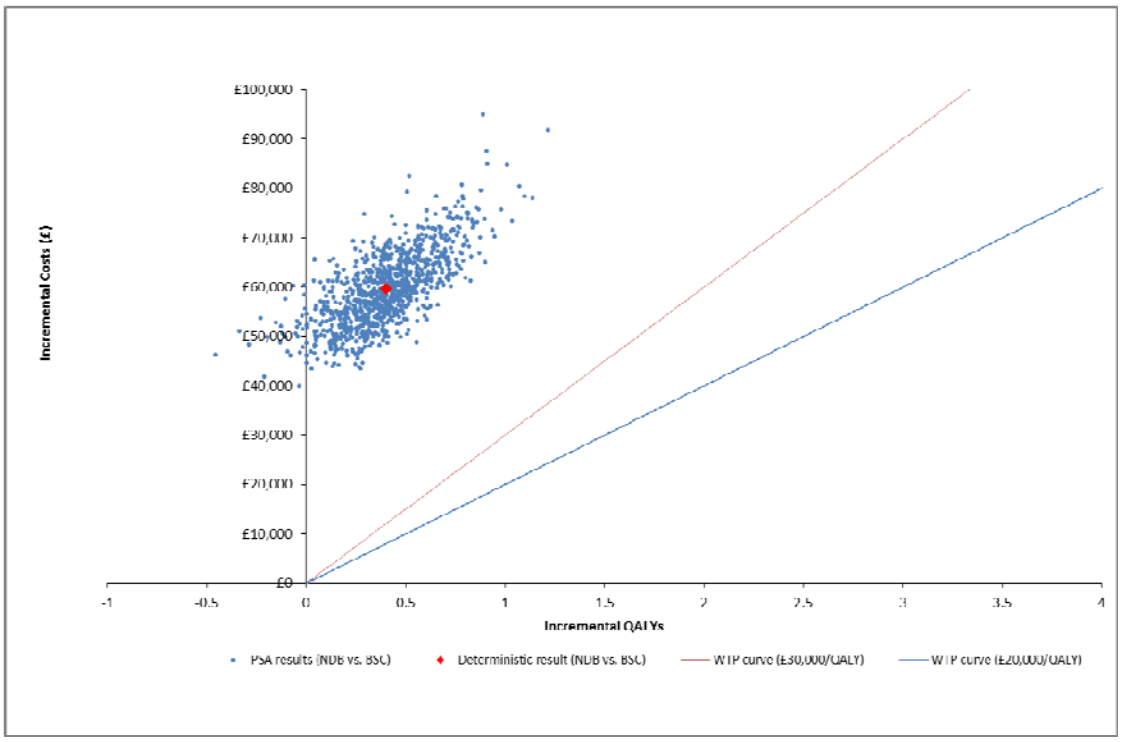
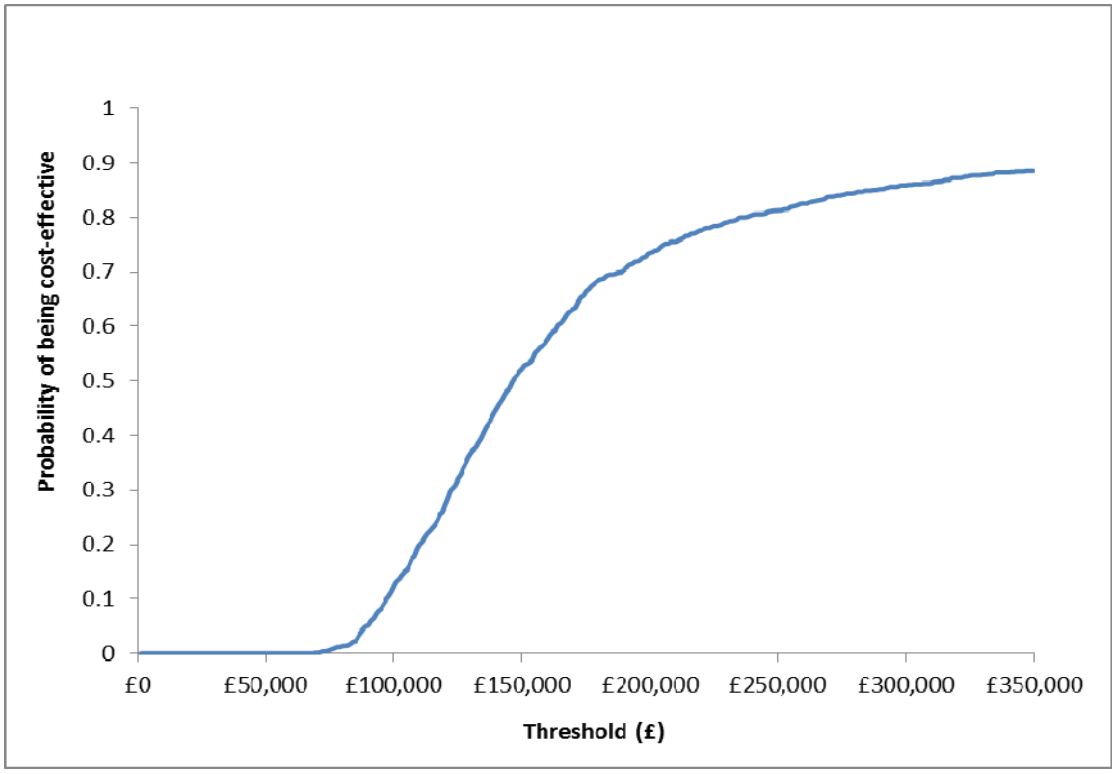


Figure 8: CEAC (nintedanib versus BSC) – without PAS



2. With patient access scheme discount

Table 14: Comparison of ICERs obtained from deterministic and probabilistic sensitivity analyses for nintedanib versus best supportive care with PAS

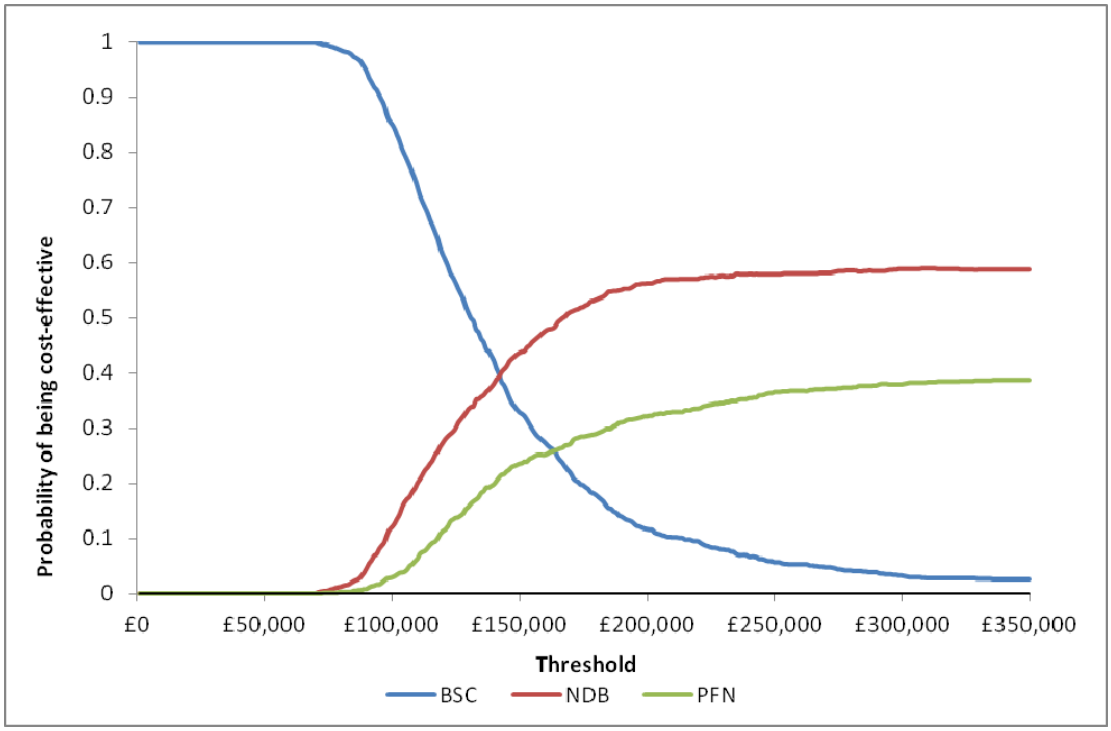
	Incremental cost	Incremental QALY	ICER
Deterministic Values	[redacted]	0.3999	[redacted]
Average value for PSA	[redacted]	0.4070	[redacted]

[redacted]

[redacted]

4.10.3. Cost-effectiveness acceptability curves and scatter-plots comparing multiple interventions

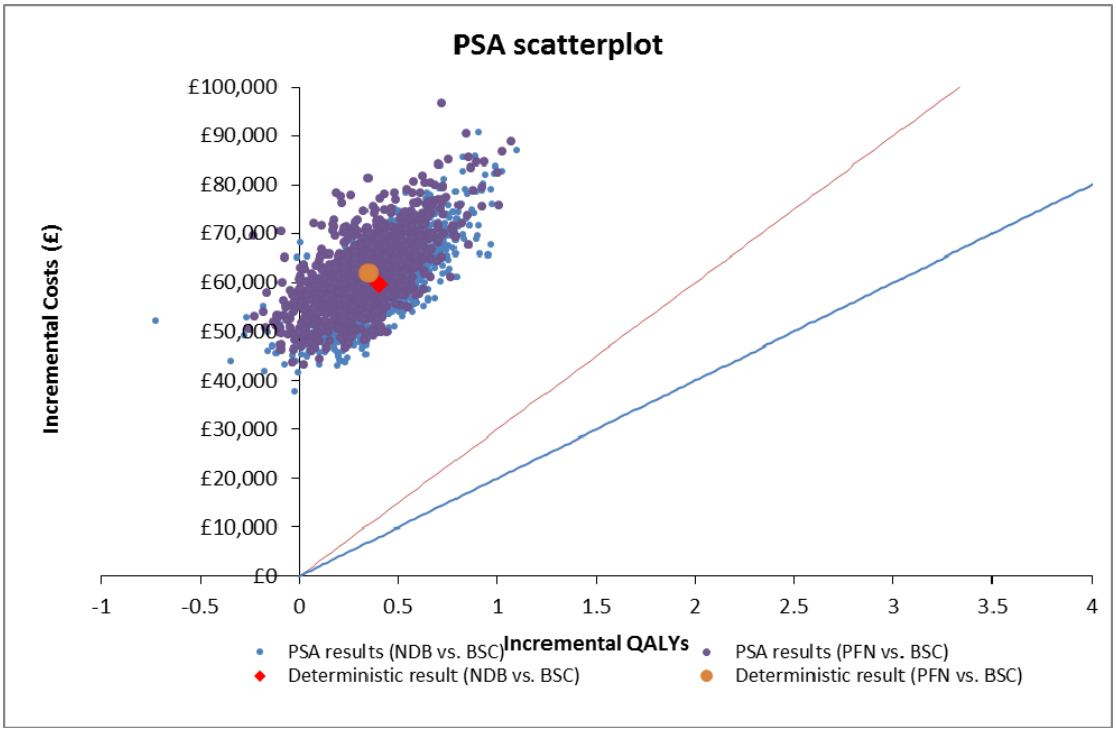
Figure 11: Multiple CEACs - without patient access scheme



BSC – best supportive care; NDB – Nintedanib; PFN - pirfenidone

[redacted]

Figure 13: Probabilistic sensitivity analysis scatter plot (all active treatment versus best supportive care); nintedanib without PAS



NDB: nintedanib, PFN: pirfenidone, BSC: best supportive care.

[redacted]

NDB: nintedanib, PFN: pirfenidone, BSC: best supportive care.

4.11 Please present scenario analysis results as described for the main manufacturer/sponsor submission of evidence for the technology appraisal.

The same scenarios as described in the main submission (Section 5.8.5 – 5.8.9) were explored in this section for nintedanib versus pirfenidone treatment comparison and nintedanib versus best supportive care.

4.11.1 Nintedanib versus pirfenidone

1. Without patient access scheme discount

Table 15: Sensitivity analysis results (nintedanib versus pirfenidone) – without PAS

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	NDB dominates	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	NDB dominates	NDB dominates
2		Exacerbation probabilities	NDB dominates	NDB dominates
3		Progression probabilities	NDB dominates	NDB dominates
4		Discontinuation probabilities	NDB dominates	NDB dominates
5	Costs	Treatment costs	NDB dominates	NDB dominates
6		Background follow-up costs	NDB dominates	NDB dominates
7		Oxygen use costs	NDB dominates	NDB dominates
8		Exacerbation event costs	NDB dominates	NDB dominates
9		EoL costs	NDB dominates	NDB dominates
10	Utilities	Health state baseline utilities	NDB dominates	NDB dominates
11		Exacerbation-related utilities	NDB dominates	NDB dominates
12	Adverse events	Adverse events probabilities	NDB dominates	NDB dominates
13		Adverse events costs	NDB dominates	NDB dominates
14		Adverse events related utilities	NDB dominates	NDB dominates
Tests on other model parameters				
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**	
15	Overall survival	Baseline risk: Weibull parametric model	NDB dominates	
16		Baseline risk: Gompertz parametric model	NDB dominates	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	NDB dominates	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	NDB dominates	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	NDB dominates	
20		Relative risk: NMA results, scenario 4	NDB dominates	

		excluding Azuma (outcome definition)	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	NDB dominates
22	Exacerbation	Baseline risk: use adjudication committee estimate	NDB dominates
23		Baseline risk: exclude recurrent exacerbation risk	NDB dominates
24		Relative risk: NMA results, scenario 1, all evidence	NDB dominates
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	NDB dominates
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	NDB dominates
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	NDB dominates
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	NDB dominates
29	Loss of lung function	Baseline risk: include exacerbation coefficient	NDB dominates
30		Relative risk: NMA results, scenario 1, all evidence	NDB dominates
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	NDB dominates
32		Relative risk: NMA results, scenario 4, include death (without Noble)	NDB dominates
33		Relative risk: NMA results, scenario 5, include death (with Noble)	NDB dominates
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	NDB dominates
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	NDB dominates
36		SAE disutility value: use alternative value for serious cardiac events	NDB dominates
37		SAE disutility value: use alternative value for GI perforation	NDB dominates
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	NDB dominates
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	NDB dominates
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	NDB dominates
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	NDB dominates
42		Relative risk: NMA results, scenario 4, excluding King	NDB dominates
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	NDB dominates
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	NDB dominates
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	NDB dominates
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	£82, 784.36

47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	N/A for this comparison
48	Ascend-like population	Use a more narrow patient population (likely to progress more rapidly) than INPULSIS trials	£13,459.17

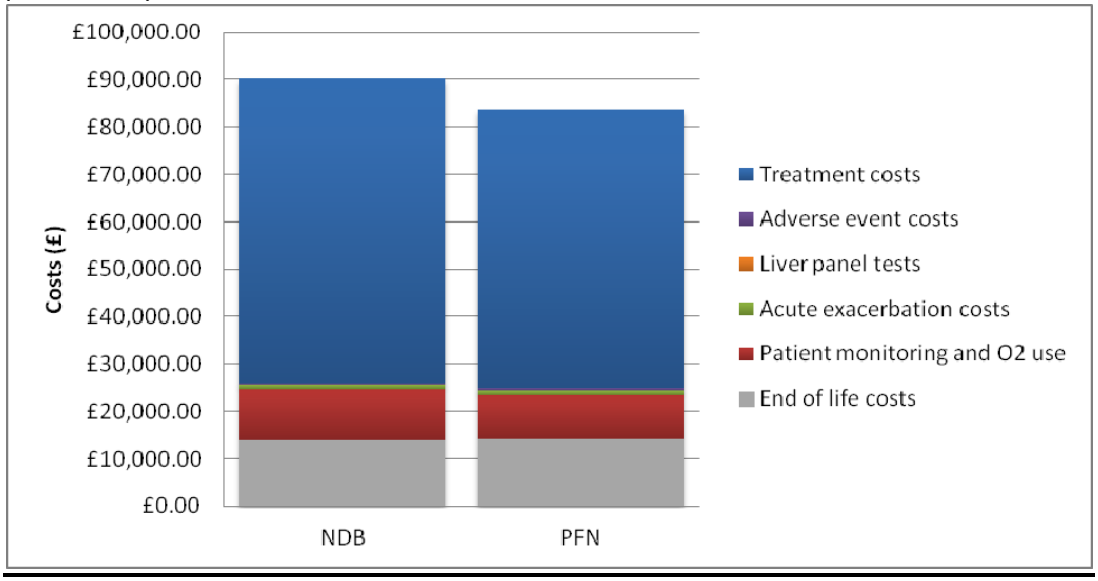
ASCEND-like population: NDB vs. PFN (Scenario 48)

This sensitivity analysis (scenario 48) was performed to investigate the cost-effectiveness of nintedanib versus pirfenidone in a restricted population representative of the ASCEND trial selection criteria. The restricting criteria used in creating the subgroup are: FVC 50-90% predicted, DL_{CO} 30-90% predicted, $FEV_1/FVC \geq 0.8$. The ASCEND selection criteria were chosen to allow enrolment of patients with an increased risk of disease progression.

Table 16: Deterministic results for nintedanib versus pirfenidone in Ascend-like population – without PAS

	PFN	NDB	Incremental
Treatment costs	£58,803.29	£64,387.68	£5,584.39
Adverse event costs	£361.74	£256.00	-£105.75
Liver panel tests	£9.01	£9.87	£0.86
Patient monitoring and O2 use	£9,276.29	£10,770.37	£1,494.08
Acute exacerbation costs	£1,142.57	£929.36	-£213.20
End of life costs	£14,094.61	£13,871.76	-£222.85
Total costs	**£83,687.52**	**£90,225.03**	**£6,537.52**
Total QALYs	**2.5024**	**2.9881**	**0.4857**
ICER (per QALY)			**£13,459.17**

Figure 15: Base-case total costs (nintedanib and pirfenidone) in Ascend-like population (Scenario 48) – without PAS



2. With patient access scheme discount

Table 17: Sensitivity analysis results (Nintedanib versus Pirfenidone) – with PAS

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	NDB dominates	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	[redacted]	[redacted]
2		Exacerbation probabilities	[redacted]	[redacted]
3		Progression probabilities	[redacted]	[redacted]
4		Discontinuation probabilities	[redacted]	[redacted]
5	Costs	Treatment costs	[redacted]	[redacted]
6		Background follow-up costs	[redacted]	[redacted]
7		Oxygen use costs	[redacted]	[redacted]
8		Exacerbation event costs	[redacted]	[redacted]
9		EoL costs	[redacted]	[redacted]
10	Utilities	Health state baseline utilities	[redacted]	[redacted]
11		Exacerbation-related utilities	[redacted]	[redacted]
12	Adverse events	Adverse events probabilities	[redacted]	[redacted]
13		Adverse events costs	[redacted]	[redacted]
14		Adverse events related utilities	[redacted]	[redacted]
Tests on other model parameters				
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**	
15	Overall survival	Baseline risk: Weibull parametric model	[redacted]	
16		Baseline risk: Gompertz parametric model	[redacted]	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	[redacted]	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	[redacted]	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	[redacted]	
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	[redacted]	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	[redacted]	
22	Exacerbation	Baseline risk: use adjudication committee estimate	[redacted]	
23		Baseline risk: exclude recurrent exacerbation risk	[redacted]	
24		Relative risk: NMA results, scenario 1, all evidence	[redacted]	
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	[redacted]	
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	[redacted]	
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	[redacted]	
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	[redacted]	
29	Loss of lung function	Baseline risk: include exacerbation coefficient	[redacted]	
30		Relative risk: NMA results, scenario 1, all evidence	[redacted]	
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	[redacted]	

32		Relative risk: NMA results, scenario 4, include death (without Noble)	[redacted]
33		Relative risk: NMA results, scenario 5, include death (with Noble)	[redacted]
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	[redacted]
35		Relative risk: serious GI events, NMA results, scenario 2 excluding Richeldi 2011	[redacted]
36		SAE disutility value: use alternative value for serious cardiac events	[redacted]
37		SAE disutility value: use alternative value for GI perforation	[redacted]
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	[redacted]
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	[redacted]
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	[redacted]
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	[redacted]
42		Relative risk: NMA results, scenario 4, excluding King	[redacted]
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	[redacted]
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	[redacted]
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	[redacted]
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	[redacted]
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	N/A for this comparison
48	Ascend-like population	Use a more narrow patient population (likely to progress more rapidly) than INPULSIS trials	[redacted]

ASCEND-like population: NDB vs. PFN (scenario 48)

Table 18: Deterministic results for nintedanib vs. pirfenidone in Ascend-like population – with PAS

	PFN	NDB	Incremental
Treatment costs	£58,803.29	[redacted]	[redacted]
Adverse event costs	£361.74	[redacted]	[redacted]
Liver panel tests	£9.01	[redacted]	[redacted]
Patient monitoring and O2 use	£9,276.29	[redacted]	[redacted]
Acute exacerbation costs	£1,142.57	[redacted]	[redacted]
End of life costs	£14,094.61	[redacted]	[redacted]
Total costs	**£83,687.52**	[redacted]	[redacted]
Total QALYs	**2.5024**	**2.9881**	**0.4857**
ICER (per QALY)			[redacted]

[redacted]

4.11.2. Nintedanib versus Best Supportive Care

1. Without patient access scheme discount

Table 19: Sensitivity analysis results (Nintedanib versus best supportive care) – without PAS

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	*£149,360.88*	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	£87,246.22	BSC dominates
2		Exacerbation probabilities	£146,591.27	£157,062.84
3		Progression probabilities	£143,278.82	£158,034.61
4		Discontinuation probabilities	£140,192.71	£163,249.01
5	Costs	Treatment costs	£141,962.86	£156,764.70
6		Background follow-up costs	£149,178.14	£149,553.56
7		Oxygen use costs	£148,844.04	£149,877.68
8		Exacerbation event costs	£149,606.92	£149,004.11
9		EoL costs	£149,509.73	£149,211.99
10	Utilities	Health state baseline utilities	£152,328.36	£146,506.75
11		Exacerbation-related utilities	£148,485.13	£150,239.30
12	Adverse events	Adverse events probabilities	£147,851.27	£154,255.00
13		Adverse events costs	£149,196.65	£149,476.14
14		Adverse events related utilities	£150,217.68	£148,513.74
Tests on other model parameters				
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**	
15	Overall survival	Baseline risk: Weibull parametric model	£240,092.88	
16		Baseline risk: Gompertz parametric model	£468,834.61	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	£151,738.22	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	£150,274.70	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	£146,205.31	
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	£150,274.70	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	£150,274.70	
22	Exacerbation	Baseline risk: use adjudication committee estimate	£149,528.72	
23		Baseline risk: exclude recurrent exacerbation risk	£149,498.31	
24		Relative risk: NMA results, scenario 1, all evidence	£149,380.35	
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	£149,380.35	
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	£150,547.65	
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	£149,380.35	
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	£149,380.35	

29	Loss of lung function	Baseline risk: include exacerbation coefficient	£146,628.44
30		Relative risk: NMA results, scenario 1, all evidence	£149,597.13
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	£149,061.52
32		Relative risk: NMA results, scenario 4, include death (without Noble)	£149,597.13
33		Relative risk: NMA results, scenario 5, include death (with Noble)	£149,597.13
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	£148,884.56
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	£149,173.86
36		SAE disutility value: use alternative value for serious cardiac events	£149,193.00
37		SAE disutility value: use alternative value for GI perforation	£149,301.91
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	£150,127.78
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	N/A for this comparison
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	£149,429.48
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	£148,914.35
42		Relative risk: NMA results, scenario 4, excluding King	£149,429.48
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	£149,257.93
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	£149,818.88
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	£148,851.34
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	N/A for this comparison

47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	£194,458.59
48	Ascend-like population	Use a more narrow patient population (likely to progress more rapidly) than INPULSIS trials	N/A for this comparison

2. With patient access scheme discount

Table 20: Sensitivity analysis results (Nintedanib versus best supportive care) – with PAS

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)	
			Low value of the parameter	High value of the parameter
Base case		*N/A*	[redacted]	
Tests around the 95% confidence interval values of model parameters				
1	Probabilities	Mortality probabilities	[redacted]	[redacted]
2		Exacerbation probabilities	[redacted]	[redacted]
3		Progression probabilities	[redacted]	[redacted]
4		Discontinuation probabilities	[redacted]	[redacted]
5	Costs	Treatment costs	[redacted]	[redacted]
6		Background follow-up costs	[redacted]	[redacted]
7		Oxygen use costs	[redacted]	[redacted]
8		Exacerbation event costs	[redacted]	[redacted]
9		EoL costs	[redacted]	[redacted]
10	Utilities	Health state baseline utilities	[redacted]	[redacted]
11		Exacerbation-related utilities	[redacted]	[redacted]
12	Adverse events	Adverse events probabilities	[redacted]	[redacted]
13		Adverse events costs	[redacted]	[redacted]
14		Adverse events related utilities	[redacted]	[redacted]
Tests on other model parameters				
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**	
15	Overall survival	Baseline risk: Weibull parametric model	[redacted]	
16		Baseline risk: Gompertz parametric model	[redacted]	
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	[redacted]	
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	[redacted]	
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	[redacted]	
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	[redacted]	
21		Relative risk: NMA results, scenario 5 excluding Japanese studies	[redacted]	

		(European studies only)	
22	Exacerbation	Baseline risk: use adjudication committee estimate	[redacted]
23		Baseline risk: exclude recurrent exacerbation risk	[redacted]
24		Relative risk: NMA results, scenario 1, all evidence	[redacted]
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	[redacted]
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	[redacted]
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	[redacted]
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	[redacted]
29	Loss of lung function	Baseline risk: include exacerbation coefficient	[redacted]
30		Relative risk: NMA results, scenario 1, all evidence	[redacted]
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	[redacted]
32		Relative risk: NMA results, scenario 4, include death (without Noble)	[redacted]
33		Relative risk: NMA results, scenario 5, include death (with Noble)	[redacted]
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	[redacted]
35		Relative risk: serious GI events, NMA results, scenario 2 excluding Richeldi 2011	[redacted]
36		SAE disutility value: use alternative value for serious cardiac events	[redacted]
37		SAE disutility value: use alternative value for GI perforation	[redacted]
38		SAE disutility value: use extreme value for all SAEs: maximum disutility – serious cardiac events value	[redacted]
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	[redacted]
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	[redacted]
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	[redacted]
42		Relative risk: NMA results, scenario 4, excluding King	[redacted]
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	[redacted]
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	[redacted]
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	[redacted]
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	[redacted]
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	[redacted]
48	Ascend-like	Use a more narrow patient population	[redacted]

	population	(likely to progress more rapidly) than INPULSIS trials	

4.11.3. Impact of Discount Rate Variation (Pirfenidone PAS)

As Boehringer Ingelheim does not have access to the net price of pirfenidone used in practise, scenario analyses were performed by reducing the list price for pirfenidone in 5% increments from a 5% discount to a 95% discount. Results are shown in Table 21 and Table 22.

Table 21: Impact of pirfenidone discount rate on ICER; nintedanib without PAS

	Discount Applied to Pirfenidone Drug Cost						
Pirfenidone	**Base-case (0% discount)**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**
Cost per pack	£2,151.10	£2043.55	£1935.99	£1828.44	£1720.88	£1613.33	£1505.77
ICER	Dominates	£13,663.45	£78,108.24	£142,553.03	£206,997.81	£271,442.60	£335,887.39
Pirfenidone	**35%**	**40%**	**45%**	**50%**	**55%**	**60%**	**65%**
Cost per pack	£1398.22	£1290.66	£1183.11	£1075.55	£968.00	£860.44	£752.89
ICER	£400,332.17	£464,776.96	£529,221.75	£593,666.54	£658,111.32	£722,556.11	£787,000.90
Pirfenidone	**70%**	**75%**	**80%**	**85%**	**90%**	**95%**	
Cost per pack	£645.33	£537.78	£430.22	£322.67	£215.11	£107.56	
ICER	£851,445.68	£915,890.47	£980,335.26	£1,044,780.04	£1,109,224.83	£1,173,669.62	

Table 22: Impact of pirfenidone discount rate on ICER; nintedanib with PAS

	Discount Applied to Pirfenidone Drug Cost						
Pirfenidone	**Base-case (0% discount)**	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**
Cost per pack	£2,151.10	£2043.55	£1935.99	£1828.44	£1720.88	£1613.33	£1505.77
ICER	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pirfenidone	**35%**	**40%**	**45%**	**50%**	**55%**	**60%**	**65%**
Cost per pack	£1398.22	£1290.66	£1183.11	£1075.55	£968.00	£860.44	£752.89
ICER	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pirfenidone	**70%**	**75%**	**80%**	**85%**	**90%**	**95%**	
Cost per pack	£645.33	£537.78	£430.22	£322.67	£215.11	£107.56	
ICER	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	

4.12 If any of the criteria on which the patient access scheme depends are clinical variable (for example, choice of response measure, level of response, duration of treatment), sensitivity analyses around the individual criteria should be provided, so that the Appraisal Committee can determine which criteria are the most appropriate to use.

N/A

Impact of patient access scheme on ICERs

4.13 For financially based schemes, please present the results showing the impact of the patient access scheme on the ICERs for the base-case and any scenario analyses. A suggested format is shown below (see table 5). If you are submitting the patient access scheme at the end of the appraisal process, you must include the scenario with the assumptions that the Appraisal Committee considered to be most plausible.

Table 23 Results showing the impact of patient access scheme on ICERs

	ICER for intervention (nintedanib) versus:			
	Pirfenidone		**Best Supportive Care**	
	Without PAS	**With PAS**	**Without PAS**	**With PAS**
Base-case deterministic values	Nintedanib dominates	[redacted]	£149,360.88	[redacted]
Average for PSA	Nintedanib dominates	[redacted]	£148,180.99	[redacted]

PAS - patient access scheme; PSA – probabilistic sensitivity analysis

Table 24: Results showing the impact of patient access scheme on ICERs for different scenarios (nintedanib versus pirfenidone)

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)			
			Low value of the parameter	High value of the parameter	Low value of the parameter	High value of the parameter
			Without PAS discount		With PAS discount	
Base case		*N/A*	NDB dominates		[redacted]	
Tests around the 95% confidence interval values of model parameters						
1	Probabilities	Mortality probabilities	NDB dominates	NDB dominates	[redacted]	[redacted]
2		Exacerbation probabilities	NDB dominates	NDB dominates	[redacted]	[redacted]
3		Progression probabilities	NDB dominates	NDB dominates	[redacted]	[redacted]
4		Discontinuation probabilities	NDB dominates	NDB dominates	[redacted]	[redacted]
5	Costs	Treatment costs	NDB dominates	NDB dominates	[redacted]	[redacted]
6		Background follow-up costs	NDB dominates	NDB dominates	[redacted]	[redacted]
7		Oxygen use costs	NDB dominates	NDB dominates	[redacted]	[redacted]
8		Exacerbation event costs	NDB dominates	NDB dominates	[redacted]	[redacted]
9		EoL costs	NDB dominates	NDB dominates	[redacted]	[redacted]
10	Utilities	Health state baseline utilities	NDB dominates	NDB dominates	[redacted]	[redacted]
11		Exacerbation-related utilities	NDB dominates	NDB dominates	[redacted]	[redacted]
12	Adverse events	Adverse events probabilities	NDB dominates	NDB dominates	[redacted]	[redacted]
13		Adverse events costs	NDB dominates	NDB dominates	[redacted]	[redacted]
14		Adverse events related utilities	NDB dominates	NDB dominates	[redacted]	[redacted]
Tests on other model parameters						
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**			
			Without PAS discount		With PAS discount	
15	Overall survival	Baseline risk: Weibull parametric model	NDB dominates		[redacted]	
16		Baseline risk: Gompertz parametric model	NDB dominates		[redacted]	
17		Baseline risk: do not allow progression from FVC40-	NDB dominates		[redacted]	

		49.9%Pred to FVC30-39.9%Pred (death)		
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	NDB dominates	[redacted]
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	NDB dominates	[redacted]
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	NDB dominates	[redacted]
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	NDB dominates	[redacted]
22	Exacerbation	Baseline risk: use adjudication committee estimate	NDB dominates	[redacted]
23		Baseline risk: exclude recurrent exacerbation risk	NDB dominates	[redacted]
24		Relative risk: NMA results, scenario 1, all evidence	NDB dominates	[redacted]
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	NDB dominates	[redacted]
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	NDB dominates	[redacted]
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	NDB dominates	[redacted]
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	NDB dominates	[redacted]
29	Loss of lung function	Baseline risk: include exacerbation coefficient	NDB dominates	[redacted]
30		Relative risk: NMA results, scenario 1, all evidence	NDB dominates	[redacted]
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	NDB dominates	[redacted]
32		Relative risk: NMA results, scenario 4, include death (without Noble)	NDB dominates	[redacted]
33		Relative risk: NMA results, scenario 5, include death (with Noble)	NDB dominates	[redacted]
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	NDB dominates	[redacted]
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	NDB dominates	[redacted]
36		SAE disutility value: use alternative value for serious cardiac events	NDB dominates	[redacted]
37		SAE disutility value: use alternative value for GI perforation	NDB dominates	[redacted]
38		SAE disutility value: use extreme value for all SAEs:	NDB dominates	[redacted]

		maximum disutility - serious cardiac events value		
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	NDB dominates	[redacted]
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	NDB dominates	[redacted]
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	NDB dominates	[redacted]
42		Relative risk: NMA results, scenario 4, excluding King	NDB dominates	[redacted]
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	NDB dominates	[redacted]
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	NDB dominates	[redacted]
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	NDB dominates	[redacted]
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	£82, 784.36	[redacted]
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	N/A for this comparison	N/A for this comparison
48	Ascend-like population	Use a more narrow patient population (likely to progress more rapidly) than INPULSIS trials	£13,459.17	[redacted]

Table 25: Results showing the impact of patient access scheme on ICERs for different scenarios (nintedanib versus best supportive care)

Scenario	Parameter	Description of parameter varied	ICER (IC/IQALYs)			
			Low value of the parameter	High value of the parameter	Low value of the parameter	High value of the parameter
			Without PAS discount		With PAS discount	
Base case		*N/A*	*£149,360.88*		[redacted]	
Tests around the 95% confidence interval values of model parameters						
1	Probabilities	Mortality probabilities	£87,246.22	BSC dominates	[redacted]	[redacted]
2		Exacerbation probabilities	£146,591.27	£157,062.84	[redacted]	[redacted]
3		Progression probabilities	£143,278.82	£158,034.61	[redacted]	[redacted]
4		Discontinuation probabilities	£140,192.71	£163,249.01	[redacted]	[redacted]
5	Costs	Treatment costs	£141,962.86	£156,764.70	[redacted]	[redacted]
6		Background follow-up costs	£149,178.14	£149,553.56	[redacted]	[redacted]
7		Oxygen use costs	£148,844.04	£149,877.68	[redacted]	[redacted]
8		Exacerbation event costs	£149,606.92	£149,004.11	[redacted]	[redacted]
9		EoL costs	£149,509.73	£149,211.99	[redacted]	[redacted]
10	Utilities	Health state baseline utilities	£152,328.36	£146,506.75	[redacted]	[redacted]
11		Exacerbation-related utilities	£148,485.13	£150,239.30	[redacted]	[redacted]
12	Adverse events	Adverse events probabilities	£147,851.27	£154,255.00	[redacted]	[redacted]
13		Adverse events costs	£149,196.65	£149,476.14	[redacted]	[redacted]
14		Adverse events related utilities	£150,217.68	£148,513.74	[redacted]	[redacted]
Tests on other model parameters						
Scenario	**Parameter**	**Description of parameter varied**	**ICER (IC/IQALYs)**			
			Without PAS discount		With PAS discount	
15	Overall	Baseline risk: Weibull parametric model	£240,092.88		[redacted]	

16	survival	Baseline risk: Gompertz parametric model	£468,834.61	[redacted]
17		Baseline risk: do not allow progression from FVC40-49.9%Pred to FVC30-39.9%Pred (death)	£151,738.22	[redacted]
18		Relative risk: NMA results, scenario 2, excluding King, Azuma and Taniguchi (patient characteristics)	£150,274.70	[redacted]
19		Relative risk: NMA results, scenario 3 excluding Richeldi 2011, King, Azuma and Taniguchi	£146,205.31	[redacted]
20		Relative risk: NMA results, scenario 4 excluding Azuma (outcome definition)	£150,274.70	[redacted]
21		Relative risk: NMA results, scenario 5 excluding Japanese studies (European studies only)	£150,274.70	[redacted]
22	Exacerbation	Baseline risk: use adjudication committee estimate	£149,528.72	[redacted]
23		Baseline risk: exclude recurrent exacerbation risk	£149,498.31	[redacted]
24		Relative risk: NMA results, scenario 1, all evidence	£149,380.35	[redacted]
25		Relative risk: NMA results, scenario 2 excluding Homma (route of administration and study bias)	£149,380.35	[redacted]
26		Relative risk: NMA results, scenario 4 excluding Richeldi 2011, Azuma, Taniguchi and Homma	£150,547.65	[redacted]
27		Relative risk: NMA results, scenario 5 excluding Azuma (heterogeneity)	£149,380.35	[redacted]
28		Relative risk: NMA results, scenario 6 excluding Azuma and Homma	£149,380.35	[redacted]
29	Loss of lung function	Baseline risk: include exacerbation coefficient	£146,628.44	[redacted]
30		Relative risk: NMA results, scenario 1, all evidence	£149,597.13	[redacted]
31		Relative risk: NMA results, scenario 3, excluding Richeldi 2011 and King	£149,061.52	[redacted]
32		Relative risk: NMA results, scenario 4, include death (without Noble)	£149,597.13	[redacted]
33		Relative risk: NMA results, scenario 5, include death (with Noble)	£149,597.13	[redacted]
34	Safety	Relative risk: serious cardiac events, NMA results, scenario 1, all evidence	£148,884.56	[redacted]
35		Relative risk: serious GI events, NMA results, scenario excluding Richeldi 2011	£149,173.86	[redacted]
36		SAE disutility value: use alternative value for serious cardiac events	£149,193.00	[redacted]
37		SAE disutility value: use alternative value for GI	£149,301.91	[redacted]

		perforation		
38		SAE disutility value: use extreme value for all SAEs: maximum disutility - serious cardiac events value	£150,127.78	[redacted]
39	Overall discontinuation	Baseline risk: discontinuation rate of 22% in 9.5 months for NDB and PFN (6.71% per cycle)	N/A for this comparison	N/A for this comparison
40		Relative risk: NMA results, scenario 2,excluding King and Taniguchi	£149,429.48	[redacted]
41		Relative risk: NMA results, scenario 3, excluding King, Taniguchi, and Richeldi 2011	£148,914.35	[redacted]
42		Relative risk: NMA results, scenario 4, excluding King	£149,429.48	[redacted]
43		Relative risk: NMA results, scenario 5, excluding Taniguchi	£149,257.93	[redacted]
44	FVC%Pred values	Use the lowest value of each FVC%Pred category (e.g. 50 for the 50-59.99 FVC%Pred category) as starting point	£149,818.88	[redacted]
45		Use the highest value of each FVC%Pred category (e.g. 59.9 for the 50-59.9 FVC%Pred category) as starting point	£148,851.34	[redacted]
46	PFN stopping rule	Progressors in the PFN arm discontinue and lose treatment effect	N/A for this comparison	N/A for this comparison
47	Direct evidence for NDB	Use clinical trial data for both NDB and PBO for all efficacy parameters (OS, exacerbations, loss of lung function) and discontinuation	£194,458.59	[redacted]
48	Ascend-like population	Use a more narrow patient population (likely to progress more rapidly) than INPULSIS trials	N/A for this comparison	N/A for this comparison

Appendices

4.14 Appendix A: Additional documents

4.14.1 If available, please include copies of patient access scheme agreement forms, patient registration forms, pharmacy claim forms/rebate forms, guides for pharmacists and physicians, patient information documents.

Response

4.15 Appendix B: Details of outcome-based schemes

4.15.1 If you are submitting a proven value: price increase scheme, as defined in the PPRS, please provide the following information:

- the current price of the intervention
- the proposed higher price of the intervention, which will be supported by the collection of new evidence
- a suggested date for when NICE should consider the additional evidence.

Response

4.15.2 If you are submitting an expected value: rebate scheme, as defined in the PPRS, please provide the following details:

- the current price of the intervention (the price that will be supported by the collection of new evidence)
- the planned lower price of the intervention in the event that the additional evidence does not support the current price
- a suggested date for when NICE should consider the additional evidence.

Response

4.15.3 If you are submitting a risk-sharing scheme, as defined in the PPRS, please provide the following details:

- the current price of the intervention (the price that will be supported by the collection of new evidence)
- the proposed relationship between future price changes and the evidence to be collected.

Response

4.15.4 For outcome-based schemes, as defined in the PPRS, please provide the full details of the new information (evidence) planned to be collected, who will collect it and who will carry the cost associated with this planned data collection. Details of the new information (evidence) may include:

- design of the new study
- patient population of the new study
- outcomes of the new study
- expected duration of data collection
- planned statistical analysis, definition of study groups and reporting (including uncertainty)
- expected results of the new study
- planned evidence synthesis/pooling of data (if applicable)
- expected results of the evidence synthesis/pooling of data (if applicable).

Response

4.15.5 If you are submitting a risk-sharing scheme, please specify the period between the time points when the additional evidence will be considered.

Response

4.15.6 Please provide the clinical effectiveness data resulting from the evidence synthesis and used in the economic modelling of the patient access scheme at the different time points when the additional evidence is to be considered.

Response

4.15.7 Please provide the other data used in the economic modelling of the patient access scheme at the different time points when the additional evidence is to be considered. These data could include cost/resource use, health-related quality of life and utilities.

Response

4.15.8 Please present the cost-effectiveness results as follows.

- For proven value: price increase schemes, please summarise in separate tables:
  - the results based on current evidence and current price
  - the anticipated results based on the expected new evidence and the proposed higher price.
- For expected value: rebate schemes, please summarise in separate tables:
  - the results based on the expected new evidence and the current price (which will be supported by the additional evidence collection)
  - the results based on the current evidence and the lower price (if the new evidence is not forthcoming).
- For risk-sharing schemes, please summarise in separate tables:
  - the results based on current evidence and current price
  - the results based on the expected new evidence and the current price (which will be supported by the additional evidence collection)
  - the results based on the current evidence and the lower price (if the new evidence is not forthcoming)
  - the anticipated results based on the expected new evidence and the proposed higher price.

A suggested format is shown in table 3, section 4.7.

4.15.9 Please present in separate tables the incremental results for the different scenarios as described above in section 5.2.8 for the type of outcome-based scheme being submitted.

List the interventions and comparator(s) from least to most expensive. Present the incremental cost-effectiveness ratios (ICERs) in comparison with baseline (usually standard care), and the incremental analysis ranking technologies in terms of dominance and extended dominance. A suggested format is presented in table 4, section 4.8.