CONFIDENTIAL

External Assessment Group Report commissioned by the NIHR Evidence Synthesis Programme on behalf of NICE

Nintedanib for treating idiopathic pulmonary fibrosis in people with a forced vital capacity above 80% predicted (part-review of technology appraisal guidance 379) [ID4062] Post-factual accuracy check (FAC) version

Produced by	Southampton Health Technology Assessments Centre (SHTAC)
Authors	Keith Cooper, Senior Research Fellow, Health Economics Lorna Hazell, Senior Research Assistant, Evidence Synthesis Asyl L. Hawa, Research Fellow, Health Economics Jo Picot, Senior Research Fellow, Evidence Synthesis Fay Chinnery, Research Fellow, Health Economics David Scott, Principal Research Fellow, Statistics Jonathan Shepherd, Principal Research Fellow, Evidence Synthesis
Correspondence to	Jonathan Shepherd Southampton Health Technology Assessments Centre (SHTAC) School of Healthcare Enterprise and Innovation Alpha House Enterprise Road, University of Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	25/08/2022

Source of Funding: This report was commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme as project number 135574.

Acknowledgements

We are grateful to the following for providing expert clinical advice for this report:

- Dr Mark Jones, Associate Professor in Respiratory Medicine, University of Southampton and Honorary Consultant Physician & ILD Clinical Lead, University Hospitals Southampton

We also thank

- Dr Geoff Frampton, SHTAC, for providing a quality assurance review of the draft EAG report.
- Lois Woods, Senior Research Assistant, SHTAC for performing literature searching and critically appraising search strategies.

Declared competing interests of the authors and advisors

- The authors declare none.
- Dr Jones is a co-investigator and a principal investigator on preclinical studies of novel treatments for lung diseases, including idiopathic pulmonary fibrosis, funded under a collaborative framework agreement between the University of Southampton and Boehringer Ingelheim.

Rider on responsibility for report

The view expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Cooper, K; Hazell, L; Hawa, AL; Picot, J; Chinnery, F; Scott, D; and Shepherd, J. Nintedanib for treating idiopathic pulmonary fibrosis in people with a forced vital capacity above 80% predicted (part-review of technology appraisal guidance 379): A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2022

Contributions of authors

Keith Cooper critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Lorna Hazell critically appraised the clinical effectiveness systematic review, and drafted the report; Asyl L. Hawa, critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Jo Picot, critically appraised the clinical effectiveness systematic review, and drafted the report; Fay Chinnery critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; David Scott critically appraised the clinical effectiveness systematic review, and drafted the report; Jonathan Shepherd critically appraised the clinical effectiveness systematic review, drafted the report and is the project co-ordinator and guarantor.

Table of Contents

LIST OF TABLES

LIST of FIGURES

LIST of APPENDICES

LIST OF ABBREVIATIONS

6MWD	6-Minute Walk Distance
6MWT	6-Minute Walk Test
ACD	Appraisal Consultation Document
AE	Adverse event
AIC	Akaike Information Criterion
ANCOVA	Analysis of covariance
BD	Twice daily
BSC	Best supportive care
CASA-Q	The Cough and Sputum Assessment Questionnaire
CG	Clinical Guideline
CI	Confidence interval
CrI	Credible interval
CS	Company submission
CSR	Clinical study report
CT	Computerised tomography
DIC	Deviance Information Criterion
DLco	Carbon monoxide Diffusing Capacity
DSA	Deterministic sensitivity analysis
EAG	Evidence Assessment Group
EMA	European Medicines Agency
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 3 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
FDA	Food and Drug Administration
FEV_1	Forced expiratory volume in the first second
FVC	Forced Vital Capacity
GI	Gastro-intestinal
HCRU	Health Care Research Unit
HRCT	High-resolution computerised tomography
HRQoL	Health-related quality of life
ICER	Incremental cost-effectiveness ratio
ICU	Intensive Care Unit

ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
IPD	Individual patient level data
ITC	Indirect treatment comparison
ITT	Intention-to-treat
LOCF	Last observation carried forward
LYG	Life year gained
NAC	N-acetylcysteine
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
NMA	Network meta-analysis
OD	Once daily
OR	Odds ratio
PaO2	Partial pressure of oxygen in arterial blood
PAS	Patient Access Scheme
PFS	Progression-free survival
PGI-C	Patient's Global Impression of Change
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SGRQ	St. George's Hospital Respiratory Questionnaire
SGRQ-I	IPF-specific St. George's Hospital Respiratory Questionnaire
SmPC	Summary of Product Characteristics
SpO2	Oxygen saturation by pulse oximetry
STA	Single Technology Appraisal
TA	Technology Appraisal
UCSD-SOBQ	University of California in San Diego Shortness of Breath Questionnaire
VC	Vital Capacity

1 EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the external assessment group (EAG) as being potentially important for decision making. It also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, health technology, evidence and information on the issues are in the main body of this EAG report.

All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1 Overview of the EAG's key issues

Table 1 Summary of EAG's key issues

Issue number	Headline description	EAG report sections
1	Uncertainty in whether all relevant observational study evidence has been included in the systematic literature review	3.1.2
2	Exclusion of survival data from a published trial (Lancaster 2020 et al) which could inform the company's pooled survival analyses	3.1.2
3	The company's economic model base case uses overall survival estimates for the whole trial population, rather than the FVC > 80% predicted subgroup	4.2.6
4	Nintedanib-treated patients are followed up for much longer than placebo patients, which increases uncertainty in the longer-term comparison of clinical effectiveness	4.2.6

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves life expectancy (overall survival) and quality of life in quality-adjusted life years (QALYs). An ICER is the ratio of the additional costs to the QALYs gained.

The company report their base case cost-effectiveness results in company submission (CS) Table 115 and CS Table 116 using the list price and Patient Access Scheme (PAS) price for nintedanib respectively, reproduced in Table 2 and Table 3 below.

Table 2 Company base case results for nintedanib vs. best supportive care (using the list price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,262	4.08	3.21				
Nintedanib	£89,177	7.40	5.69	£69,915	3.32	2.49	£28,094

Reproduced from CS Table 115.
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.

Table 3 Company base case results for nintedanib vs. best supportive care (using PAS price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,262	4.08	3.21				
Nintedanib	[redacted]	7.40	5.69	[redacted]	3.32	2.49	[redacted]

Reproduced from CS Table 116.

The base case results show that nintedanib offers a mean QALY gain of 2.49 for an additional mean cost of £69,915 (list price) and [redacted] (PAS price) versus best supportive care, producing ICERs of £28,094 and [redacted] per QALY gained respectively.

In reply to clarification question B5, the company provided additional results for the FVC>80% predicted subgroup using fitted OS curves for this subgroup. The FVC > 80% predicted subgroup results had an ICER, using the PAS price for nintedanib, of [redacted] per QALY (Table 4).

Table 4 Company results for nintedanib vs. best supportive care (PAS price for nintedanib) using OS curves for the FVC > 80% predicted subgroup

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£18,724	3.87	3.06				

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
Nintedanib	[redacted]	8.50	6.51	[redacted]	4.63	3.44	[redacted]
Produced by the EAG using OS parameter estimates provided in clarification response document Table 10 and 11							

1.3 The decision problem: summary of the EAG's key issues

No key issues were identified with respect to the decision problem, notwithstanding those issues listed below which stem from the company's use of whole trial population data instead of data from the decision problem population of people with FVC > 80% predicted.

1.4 The clinical effectiveness evidence: summary of the EAG's key issues

Issue 1 Uncertainty in whether all relevant observational study evidence has been included in the systematic literature review

Report section	3.1 Critique of the methods of review(s)
Description of issue and why the EAG has identified it as important	The process for screening clinical effectiveness studies for inclusion in the systematic literature review (SLR) is not fully clear and at times appears unsystematic. The only observational evidence included in the SLR is from the INPULSIS-ON and TOMORROW open-label extension studies – both are follow-on studies from company sponsored nintedanib RCTs. However, it is not plausible that these are the only relevant available observational studies of nintedanib and best supportive care. For example, the CS cites a selection of IPF registries worldwide to validate model assumptions and outcomes or to asses trial generalisability. However, the inclusion/exclusion status of these registry studies is not clear. Any such studies that do meet the inclusion criteria should undergo the same systematic processes and reporting as the open-label extension studies.
What alternative approach has the EAG suggested?	A more explicit description of the inclusion/exclusion status of observational studies identified through the SLR literature searches.
What is the expected effect on the cost-effectiveness estimates?	Uncertain. It is possible that additional observational studies may provide data to inform clinical effectiveness estimates in the economic model.

What additional evidence or analyses might help to resolve this key issue?	The company should consider the potential impact on the model assumptions and results of all eligible observational studies.

Issue 2 Exclusion of survival data from a published trial (Lancaster 2020 et al) which could inform the company's pooled survival analyses.

Report section	3.1 Critique of the methods of review(s)
Description of issue and why the EAG has identified it as important	The CS excludes a published company-sponsored phase IIIb nintedanib RCT by Lancaster et al (2020) from their systematic literature review due to methodological limitations caused by protocol amendments (e.g. trial enrolment difficulties; lack of statistical power). The EAG notes that some of the trial outcomes are relevant to the decision problem and could also inform certain model assumptions (e.g. survival estimates). In our view not all of the methodological limitations cited would necessarily bias the trial's results to a significant degree. In response to an EAG clarification question, the company asserted that the results of the trial are supportive of (i.e. consistent with) the TOMORROW and INPULSIS trials and that inclusion of Lancaster et al (2020) would have a minimal impact on the overall results in their submission. Whilst this is reassuring, the company do not provide evidence to show the impact of this study on the model cost- effectiveness estimates.
What alternative approach has the EAG suggested?	A cost effectiveness scenario analysis including survival data from the Lancaster 2019 trial, in addition to the INPULSIS and TOMORROW trials, would illuminate the effect any apparent bias associated with Lancaster et al (2019).
What is the expected effect on the cost-effectiveness estimates?	This is uncertain at present.
What additional evidence or analyses might help to resolve this key issue?	As above, the company should provide a scenario analysis including survival data from the Lancaster 2019 trial, ideally using data for the subgroup of patients with FVC >80% if available. This would represent a more complete nintedanib evidence base than that of the current submission.

1.5 The cost-effectiveness evidence: summary of the EAG's key issues

Issue 3 The company's economic model base case uses overall survival estimates for the whole trial population, rather than the FVC > 80% predicted subgroup

Report section	4.2.6 Treatment effectiveness and extrapolation
Description of issue and why the EAG has identified it as important	The company's base case economic model uses OS data for the whole trial population, rather than the FVC > 80% predicted subgroup.
What alternative approach has the EAG suggested?	The EAG suggests that the base case economic model should use OS data for the FVC > 80% subgroup as this population is specified in the decision problem.
What is the expected effect on the cost-effectiveness estimates?	Using the EAG's corrected model, the ICER using the whole trial OS data is slightly higher at [redacted] per QALY compared to the ICER based on OS data for the FVC > 80% predicted subgroup, [redacted] per QALY.
What additional evidence or analyses might help to resolve this key issue?	The EAG recommend using OS data for FVC > 80% predicted subgroup as the base case.

Issue 4 Nintedanib-treated patients are followed up for much longer than placebo patients, which increases uncertainty in the longer-term comparison of clinical effectiveness

Report section	4.2.6 Treatment effectiveness and extrapolation
Description of issue and why the EAG has identified it as important	The pivotal RCTs allowed placebo participants to receive nintedanib open-label at the end of the 52 week blinded trial. The open-label extension studies followed-up nintedanib patients for over five years, disproportionately longer than the follow-up period for placebo.
What alternative approach has the EAG suggested?	Based on the Kaplan Meier data submitted by the company in their clarification response (B5), the EAG considers there is no difference in survival between the nintedanib and placebo arms. We therefore assume that mortality is initially the same for both the trial arms for the FVC > 80% predicted subgroup. When the mean FVC % predicted of the FVC > 80% predicted subgroup has declined to that of the whole trial population, the placebo OS curve is assumed to follow the placebo parametric curve for the whole trial population. We estimate this happens after 5.5 years.
What is the expected effect on the cost-effectiveness estimates?	Using the EAG corrected model, the ICER using the OS data for the FVC > 80% predicted subgroup is [redacted] per QALY. Applying the EAG's assumptions for the extrapolation of the placebo arm, the ICER increases to [redacted] per QALY.
What additional evidence or analyses might help to resolve this key issue?	Longer term follow-up of people receiving best supportive care, eg. from real-world data sources, evidence would help to clarify this issue.

1.6 Secondary issues: summary of the EAG's view

The EAG has identified the following secondary issues for consideration. The common theme among them is uncertainties relating to the subgroup of people with IPF and an FVC > 80% predicted.

- **Network meta-analysis (NMA)**. The company use the odds ratios estimated from their NMA to inform clinical effectiveness parameters in their economic model, but these are not stratified by FVC % predicted subgroup. Historically, the NMA includes placebo arms from pirfenidone trials, but these arms do not include patients with FVC >80% predicted. Given that pirfenidone is no longer a comparator it is arguable whether the NMA is required in the current appraisal. Instead, these odds ratios could have been estimated from the INPULSIS and/or TOMORROW RCTs directly or from a pairwise meta-analysis of these trials. This would have allowed the odds ratios to be computed for the FVC >80% predicted subgroup. The EAG notes, however, that the odds ratios from the whole trial population(s) will likely be more precise due to the larger sample size.
- **Subgroup interaction tests.** The company base their assumption of similar treatment effects across FVC % predicted subgroups, at least in part, on the non-significant results of statistical interaction tests in the INPULSIS trials. However, as the company notes, these tests are likely to be underpowered to detect a significant difference between treatment and subgroups. There remains some uncertainty about the validity of assumptions of similarity or difference in treatment effects across patient subgroups. Further expert clinical advice would be beneficial.
- **Open-label extension studies.** The OLE studies from the INPULSIS and TOMORROW RCTs only include patients who have completed the respective parent trials and therefore may comprise a more skewed sample of patients (e.g. healthier, more motivated) than general IPF patient population seen in practice. Also, the results of the extension studies are not stratified by FVC % predicted subgroups and it is therefore uncertain whether the results are fully generalisable to the FVC > 80% predicted subgroup. Further expert clinical advice would be beneficial.

1.7 Summary of EAG's preferred assumptions and resulting ICER

Based on the EAG critique of the company's model (discussed in section 5.3.5), we have identified four key aspects of the company's base case with which we disagree with the assumptions made. Our preferred model assumptions are the following:

- **Population modelled for OS**: FVC >80% predicted, rather than the whole trial population.

- **Extrapolation of OS:** For the first 5.5 years, we use the same survival curve for the BSC arm as for the nintedanib arm as the mortality rate for both arms is considered equal; thereafter we use the BSC survival curve from the whole trial population for the BSC arm.
- **OS hazard ratio for acute exacerbations:** we use a HR of 2.79, rather than 1.4.
- **Time horizon:** we use a time horizon of 35 years, rather than 50 years.

Table 5 below presents the results obtained from the model with the above preferred EAG model assumptions implemented. The results are most sensitive to the extrapolation of OS assumption.

Table 5 EAG deterministic base case results (using PAS price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£23,264	5.71	4.49				
Nintedanib	[redacted]	7.20	5.62	[redacted]	1.49	1.14	[redacted]
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.							

Modelling errors identified and corrected by the EAG are described in section 5.3.4. For further details of the exploratory and sensitivity analyses done by the EAG, see section 6.2.2.

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This EAG report is a critique of the company's submission (CS) from Boehringer Ingelheim which informs NICE's part-review of health technology guidance TA379 'Nintedanib for treating idiopathic pulmonary fibrosis (IPF)' published in 2016.

TA379 was informed by a company submission from Boehringer Ingelheim and critiqued by SHTAC in an EAG report, published in 2015.[1] (NB. At that time NICE referred to the EAG as the Evidence Review Group (ERG). To avoid potential confusion in this report arising from historical citing of the ERG (original 2015 appraisal) and the EAG (this current appraisal), from this point onward we only use the term EAG to describe this group in the past and the present). NICE's guidance recommends nintedanib as an option for adults with IPF but only in patients with a forced vital capacity (FVC) between 50% and 80% predicted. NICE have noted that this threshold for treatment was not supported by UK clinicians and recommended a part-review of TA379 [2].

The scope of this part-review is to assess the clinical and cost effectiveness of nintedanib in the subgroup of IPF patients with a FVC above 80% predicted. TA379 included evidence from two replicate phase III nintedanib randomised controlled trials (RCTs) (INPULSIS trials) and the phase II TOMORROW RCT. Our critique identifies the strengths and weakness of the current CS, focusing on new evidence submitted by the company for this subgroup of patients:

- post-hoc subgroup analyses of the INPULSIS and TOMORROW RCTs in patients with FVC >80% predicted, and
- longer-term clinical effectiveness and safety data from two open-labelled extension (OLE) studies (INPULSIS-ON and TOMORROW OLE).

One clinical expert was consulted to advise the EAG and inform this report. Clarification on some aspects of the CS was requested from the company by the EAG via NICE on 8th July 2022. A response from the company via NICE was received by the EAG on 27th July 2022, and a further response was received on 4th August 2022; these can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on idiopathic pulmonary disease

CS section B.1.3.1 provides an overview of the effects of IPF on patients and their quality of life. IPF is a progressive, irreversible lung disease with no cure. The rate of disease progression is described as heterogenous and unpredictable with a median survival from diagnosis between 2 and 5 years. Forced vital capacity (FVC) is a lung function test that measures the amount of air that can be forcibly exhaled after a deep breath. The FVC % predicted expresses the FVC as a percentage of the predicted value based on population norms adjusted for age, gender and height. In the previous NICE nintedanib appraisal (TA379), the appraisal committee acknowledged that the FVC % predicted has some limitations but concluded this is the most widely used measure in clinical practice for monitoring lung function in IPF. Clinical expert advice to the EAG is that IPF is not described in terms of severity as this is not determined by lung function alone. Some patients with significant fibrosis/symptom burden, e.g., co-existent emphysema with IPF, have an FVC % predicted that is maintained above 80% despite advancing disease (as the emphysema prevents FVC decline). Typically, in clinical practice the preferred terminology is to describe IPF as early or advanced, but this does not directly relate to FVC thresholds.

2.2.2 Background information on nintedanib

Nintedanib (OFEV®) is a tyrosine kinase inhibitor which inhibits several steps in the process of lung fibrosis. It is licensed for use in IPF in adults, regardless of the patient's FVC % predicted value. The recommended dose is 150mg orally twice daily. CS Table 2 provides a comprehensive description of treatment with nintedanib including details of other conditions for which the product has a marketing authorisation.

2.2.3 The position of nintedanib in the current treatment pathway

CS section B.1.3.2 provides an accurate description of the current clinical pathway of care in IPF. Current management of IPF in the UK includes best supportive care and pulmonary rehabilitation. NICE clinical guideline 163 on IPF defines best supportive care as including non-pharmacological approaches aimed at symptom relief, management of co-morbidities, withdrawal of therapies suspected to be ineffective or causing harm and end of life care. The CS states that pulmonary rehabilitation includes educational and exercise components (CS section B.1.3.2) and should be tailored to the individual patient. Clinical expert advice to the EAG is that patients would also undergo regular assessments for oxygen requirements and

be given ambulatory oxygen as appropriate. Lung transplantation may improve survival and quality of life, but many patients are ineligible due to increasing age and comorbidities.

Pharmacological interventions aim to slow the rate of decline in lung function. Two antifibrotic drugs, nintedanib and pirfenidone, are licensed for the treatment of IPF. Pirfenidone is licensed for treatment of mild to moderate IPF only, while nintedanib is licensed for use in IPF regardless of severity[3,4]. Both drugs are currently recommended by NICE as options for adults with IPF but their use is restricted to patients with an FVC between 50% and 80% predicted. Our clinical expert advised that best supportive care and pulmonary rehabilitation usually continues alongside pharmacological treatments as appropriate, but as disease progresses the approach shifts to discontinuation of antifibrotics, symptom relief and palliative care. NICE guidance currently recommends that pirfenidone and nintedanib are stopped if disease progresses by a 10% or more decrease in FVC % predicted in any 12-month period [1,5]. A stopping rule has not been considered in the current CS (see section 4.2.6.5 of this report for our discussion of this).

2.2.4 Management of patients with IPF and a FVC >80% predicted

Patients with an FVC > 80% predicted are estimated to represent around a third of UK IPF patients.[6] These patients currently receive best supportive care but are not eligible to receive pharmacological treatment until their lung function (as measured by FVC) has declined below 80% predicted.

In NICE TA379, the clinical and cost-effectiveness of nintedanib was compared to both pirfenidone and best supportive care, however in this current appraisal the relevant comparator is best supportive care only. Pirfenidone is not an appropriate comparator because it is not recommended by NICE for treating IPF patients with FVC >80% predicted. In TA379, the NICE Committee concluded that the incremental cost effectiveness ratios (ICERs) for nintedanib compared to best supportive care were not in the range considered to be a cost-effective use of NHS resources. This recommendation was based on ICERs estimated by the company from their economic model (in patients with FVC % predicted over 50%) and an exploratory analysis provided by the EAG (including only patients with FVC >80% predicted).[1]

EAG comment on background

The company has provided an appropriate description of the disease burden for IPF, the intervention and the current treatment pathway. They have also presented

background information that is relevant to the patient population for whom nintedanib is not currently recommended.

2.3 Critique of the company's definition of the decision problem

The company's decision problem broadly matches the final scope issued by NICE (Table 6). The CS presents evidence for the majority of the outcomes listed in Table 6 for the subgroup of patients FVC >80% predicted in the INPULSIS RCTs. However, only selected clinical outcomes are presented for this subgroup for the TOMORROW RCT (further detail is given in section 3.2.3 of this report). The effect of nintedanib on overall survival is not presented in the CS for the subgroup of patients with FVC >80% predicted in the individual INPULSIS and TOMORROW RCTs. However, pooled Kaplan Meier survival curves from these RCTs and their OLE studies were provided for this subgroup on request (company's response to clarification question B6).

Table 6 Summary of the decision problem

	Final scope issued by NICE	Company's decision problem	EAG comments
Population	Adults with idiopathic pulmonary fibrosis with FVC >80% predicted	Same as final scope issued by NICE	No concerns
Intervention	Nintedanib	Same as final scope issued by NICE	No concerns
Comparators	Established clinical management without nintedanib	Same as final scope issued by NICE	No concerns
Outcomes	The outcome measures to be considered include: • pulmonary function parameters • physical function • exacerbation rate • mortality • adverse effects of treatment • health-related quality of life.	Same as final scope issued by NICE	The outcomes in the CS are appropriate and match the final scope with the following exception: • Physical function, e.g., 6-minute walk test (6MWT) was presented in the previous appraisal (TA379) but is not included in the current submission. • The 6MWT is not included in the company's economic model and therefore we do not consider this to be a major omission.

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	No concerns
Subgroups	Not applicable	Not applicable	No concerns
Special considerations including issues related to equity or equality	Not applicable	Not applicable	No concerns

3 CLINICAL EFFECTIVENESS

In each of the following sub-sections we provide a brief re-cap on the evidence assessed in the previous NICE appraisal of nintedanib for IPF (NICE TA379) followed by a description and critique of the new evidence submitted by the company for this current part-review of TA379.

3.1 Critique of the methods of review(s)

3.1.1 Evidence submitted in TA379

The company conducted a systematic literature review (SLR) to identify RCTs on nintedanib and relevant comparators covering appropriate efficacy, safety and health related quality of life (HRQoL) outcomes. The EAG's critique of the review methods is described in section 3.1.1. of the 2015 EAG report [7]. No major concerns with respect to the review methods were noted.

3.1.2 New evidence submitted

An updated SLR was conducted to identify RCTs published from September 2014 up to 14th January 2022. The inclusion criteria for the company's updated SLR also include non-randomised trials and observational studies. The CS does not explain *a-priori* how observational studies would be used to inform the current appraisal. Appendix D.1.1 of the current CS provides details of the methods of this review. Appendix 9.1 of this report below presents the EAG's assessment of the methods of the company's updated SLR.

Following screening of titles and abstract records, a total of **150** records were selected for full text eligibility screening. CS Appendix D Table 137 lists the **89** records that were eligible for inclusion in the review but does not indicate how many unique studies these records describe. Of the remaining **61** full texts screened and excluded, reasons for exclusion by PICO criteria are summarised in CS Table 138 (see also company clarification response A1.b).) (NB. The EAG are unclear what is meant by the exclusion criterion 'timeframe out of scope' which was applied to 25 of these 61 studies).

CS Appendix D Table 136 provides an overview of **nine** "identified clinical trials". No details are given about criteria for selecting these nine trials from the 89 included records. Of these nine the CS presents evidence for a sub-set of five trials:

- phase III INPULSIS I and II RCTs,
- phase II TOMORROW RCT,

- INPULSIS-ON study and
- TOMORROW open label extension study.

In response to clarification question A1.a), the company justify why the other four studies were "not considered relevant to the decision problem" and thus excluded from the submission:

- The first excluded study ("**INMARK"; NCT02788474**) [8] had a much shorter duration (12 weeks) than the INPULSIS and TOMORROW trials (52 weeks). An open label extension to the INMARK trial (including nintedanib only; up to 40 weeks) provides relevant data for disease progression but no survival data are reported. The EAG acknowledges that due to the shorter duration of this study it is less informative for economic modelling than the INPULSIS and TOMORROW trials. Nonetheless, the INMARK study and its OLE appear to fulfil the company's PICO selection criteria and, as such, details of the study should have been reported in the CS to allow similarities or differences in characteristics and findings to be fully considered.
- The second excluded study was a company-sponsored phase IIIb trial of nintedanib by **Lancaster et al 2020 (NCT01979952**) [9] which reported relevant outcomes including deaths. The company reports that this was excluded because it is not a pivotal trial and due to substantial protocol changes (e.g. the primary analysis was conducted at six months instead of 52 weeks, thus compromising statistical power; possible bias due to premature treatment discontinuations which were greater in the placebo arm). Notwithstanding these issues, this study also appears to fulfil the company's PICO selection criteria and we would have expected the company to have provided details of this study, including its results, to allow an independent assessment of risk of bias and certainty of the findings.
- Furthermore, the EAG notes that the Lancaster et al 2020 trial was combined with the INPULSIS and TOMORROW trials and their open-label extensions in a published extrapolation of long-term survival in IPF patients (Lancaster et al 2019)[10]. The CS describes a similar method of extrapolation to inform the economic modelling for this appraisal, but without inclusion of the Lancaster trial. In response to EAG clarification questions A1b and B1 the company asserts that inclusion of this study would have minimal impact on the overall results for this submission. However, they do not provide any evidence to support this.
- The third excluded study was a safety and pharmacokinetic study in a Japanese population and not necessarily considered generalisable to the UK IPF population (**NCT01136174**).[11] We consider this a reasonable exclusion.

- The fourth excluded study had no results available (**UMIN0000020682**). [12] The EAG notes this study is likely out of scope as it compares nintedanib with pirfenidone (CS Table 136).

The company's reasons for exclusion of these four studies do not appear to fulfil the SLR exclusion criteria listed in CS Table 135. Rather, it appears that additional *ad-hoc* exclusion criteria have been applied relating to factors such as study generalisability, risk of bias and methodological quality. From the study information available to the EAG, it appears that none of the four excluded studies can be considered to provide findings with the same degree of certainty as those of the INPULSIS and TOMORROW RCTs and their extensions (we discuss study risk of bias in section 3.2.2 of this report). The company does not mention whether these four excluded studies were considered as providing supportive evidence, for example potentially informing cost effectiveness scenario analyses.

The EAG notes that the lack of consistency in the application of the PICO selection criteria to the full text articles raises the question of whether ad hoc exclusion criteria were also applied to records excluded at the title and abstract screening stage of the SLR. If so, then this suggests a bigger risk of bias in the selection of clinical effectiveness studies informing this appraisal.

INPULSIS-ON and TOMORROW OLE are the only non-randomised studies included. The EAG is unable to verify whether any other relevant non-randomised or observational studies may have been excluded from the company's SLR.

Finally, as the company's literature search was six months out of date, the EAG performed an updated search of the same databases used in the company searches. One EAG systematic reviewer screened the titles and abstracts from this search (n=311 records). No new RCTs, relevant to the decision problem, were identified.

ERG comment on the methods of review:

The EAG considers the SLR methods to be appropriate with the exception of:

- A lack of transparency in the process and criteria for study selection,
- Apparent *ad-hoc* reasons for exclusion applied to some studies,
- Lack of detail on the selection of observational studies

We are therefore unclear whether all the relevant evidence has been identified.

3.2 Critique of studies of the technology of interest, the company's analysis and interpretation

3.2.1 Included studies

3.2.1.1 Evidence submitted in TA379

The company included two replicate phase III double blind, placebo-controlled RCTs (INPULSIS I and II) and a phase II dose-escalation RCT (TOMORROW trial) on nintedanib in the original submission. In all studies, the primary endpoint was the rate of decline in FVC (ml/year) from baseline to 12 months of treatment. A summary of the methodology of the INPULSIS and TOMORROW trials is presented in section 4.3 of the original submission. [1] We assume that all patients in these trials continued to receive best supportive care as appropriate in addition to their allocated trial medication (nintedanib or placebo).

The company conducted subgroup analyses in the INPULSIS trials according to patients' baseline FVC:

- FVC predicted ≤70% vs. >70% (prespecified) (previous CS section 4.8)
- FVC predicted ≤90% vs. >90% (post-hoc) (previous CS section 4.8}.
- FVC predicted ≤80% vs. >80% (post-hoc, in response to clarification question A3).[13]

3.2.1.2 New evidence submitted

The CS provides the following new evidence:

- Further details of the post-hoc subgroup analysis from the INPULSIS RCTs for patients with baseline FVC predicted ≤80% vs. >80%,
- Post-hoc subgroup analysis from the TOMORROW RCT for patients with baseline FVC predicted ≤80% vs. >80% (provided in the company's response to clarification question A9),
- Open-label extension (OLE) studies for the INPULSIS (INPULSIS-ON) and TOMORROW (TOMORROW OLE) trials.

3.2.1.3 RCTs: Study characteristics

The methodology of the INPULSIS and TOMORROW RCTs are summarised in the current CS Tables 3 and 5 and CS section B.2.3 and key design features are summarised below in Table 7.

Table 7 Key design features of the INPULSIS and TOMORROW trials

Study	Key features
INPULSIS I and II	• Replicate 52-week, double-blind, randomised (3:2), placebo-controlled trials, evaluating the effect of oral nintedanib, 150 mg twice daily, on annual FVC decline, in patients with IPF • 487 patients with an FVC >80% predicted were randomised into the trial (295 nintedanib; 192 placebo) • Two randomised patients in the placebo arm were not treated
TOMORROW	• A 52-week, double-blind, randomised, placebo-controlled dose-escalation trial evaluating the effect of nintedanib administered at oral doses of 50 mg qd, 50 mg bid, 100 mg bid and 150 mg bid on FVC decline during one year, in patients with IPF (five trial arms in total) • 219 patients with an FVC >80% predicted were randomised into the trial: nintedanib 50mg qd (n=43), 50mg bd (n=45), 100mg bid (n=50), 150 mg bd (n=41); placebo (n=40) • One patient randomised to nintedanib 150mg was not treated

Source: CS Table 7 and responses to clarification question A2

3.2.1.4 RCTs: baseline characteristics of patients with FVC >80% predicted

Baseline characteristics for patients with FVC % predicted >80% in the TOMORROW trial are not provided in the CS as this was not a planned subgroup analysis in this study. Table 8 shows the baseline characteristics of patients in the pooled INPULSIS I and II trials stratified by FVC >80% and ≤80% predicted. CS Tables 9 and 10, respectively, present baseline characteristics for the FVC >90% and ≤90% predicted subgroups and the whole trial population. Baseline characteristics were broadly comparable between trial arms within each subgroup.

The age, sex and smoking history of the patients with FVC >80% predicted from UK sites in the INPULSIS trials were comparable with that of all patients in the British Thoracic Society (BTS) registry in 2021 (CS Table 18). This registry comprises demographic and clinical data for over 4000 patients with interstitial lung disease (including IPF and sarcoidosis) collected from 75 UK centres (largely specialist tertiary care hospitals) over an 8-year period. [14] UK trial participants with FVC >80% predicted had a similar smoking history to patients with FVC >80% predicted in the BTS registry but had a lower diffusing capacity for carbon monoxide (DLco). Clinical expert advice to the EAG is that the BTS Registry is a valuable resource, however a recognised limitation is it does not recruit consecutive patients, and only a limited number of centres contribute data.

The EAG notes that lung function parameters such as mean FVC and diffusing capacity of the lung for carbon monoxide (DLco) are higher in the group with FVC >80% predicted at baseline in the INPULSIS trials (Table 8). However, this group were slightly older on average, had a slightly higher proportion of smokers, a higher proportion of patients with centrilobular emphysema and a lower mean St George's Respiratory Questionnaire (SGRQ) score. Clinical expert advice to the ERG is that:

- FVC is not of use in patients with emphysema in determining the extent of disease in IPF or its progression over time. Radiological assessment of fibrosis and gas transfer testing are more useful. In patients with co-existent emphysema FVC may never decline below 80% despite significant radiological progression of fibrosis. Emphysema prevents FVC decline and is expected to be more frequent in patients with FVC >80% predicted.
- Our expert also commented that the higher prevalence of emphysema also explains the slightly lower FEV1/FVC ratio in these patients (as the emphysema lowers the FEV1 but not the FVC, whilst lung fibrosis alone will lower both FEV1 and FVC proportionally.)
- The lower SGRQ score indicates a better quality of life status in the FVC >80% predicted subgroup which is as expected.
- Our expert did not note any other meaningful differences in characteristics of patients between trial arms or subgroups.

Table 8 Baseline characteristics of participants in the INPULSIS trials stratified by baseline FVC >80% vs. FVC ≤80% predicted

Baseline characteristic	Baseline FVC >80% predicted		Baseline FVC ≤80% predicted	
	Nintedanib (n=295)	Placebo (n=190)	Nintedanib (n=343)	Placebo (n=233)
Male, n (%)	218 (73.9)	148 (77.9)	289 (84.3)	186 (79.8)
Age, yrs mean (SD)	68.0 (7.8)	67.6 (7.6)	65.4 (8.2)	66.5 (8.1)
Race, n (%)				
White	154 (52.5)	109 (57.4)	206 (60.1)	139 (59.7)
Asian	95 (32.2)	59 (31.1)	99 (28.9)	699 (29.6)
Black	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.6)
Missing†	22 (11.6)	46 (15.6)	25 (10.7)	36 (10.5)
Smoking status, n (%)				
Never smoked	77 (26.1)	50 (26.3)	97 (28.9)	72 (30.9)
Ex-smoker	199 (67.5)	126 (66.3)	236 (68.8)	157 (67.4)
Current smoker	19 (6.4)	14 (7.4)	10 (2.9)	4 (1.7)

Time since diagnosis, yrs mean (SD)	1.56 (1.34)	1.52 (1.35)	1.72 (1.37)	1.61 (1.27)
Centrilobular emphysema, n (%)	137 (46.4)	91 (47.9)	117 (34.1)	75 (32.2)
FVC, mean mL (SD)	3102 (783)	3241 (812)	2379 (546)	2309 (515)
FVC, % predicted mean (SD)	95.1 (12.5)	95.4 (13.7)	66.6 (8.0)	66.1 (8.1)
FEV_1/ FVC ratio, % mean (SD)	80.0 (5.8)	79.7 (5.7)	83.1 (5.4)	83.3 (5.7)
DL_{CO}, % predicted mean (SD)	51.4 (13.5)	51.2 (11.9)	44.0 (12.6)	43.5 (13.6)
SGRQ total score, mean (SD)	34.3 (18.5)	34.1 (17.1)	43.9 (18.6)	44.0 (18.5)

Abbreviations: DL_{CO}, diffusing capacity of the lung for carbon monoxide; FVC, forced vital capacity; FEV_1, forced expiratory volume; SD, standard deviation; SGRQ, St George's Respiratory Questionnaire.

†In France, regulation did not permit the collection of data on race.

Source: Table reproduced from CS Table 8

3.2.1.5 Open label extension studies: study characteristics

The methodology of the INPULSIS-ON and TOMORROW OLE studies is summarised in the current CS Tables 4 and 6 and CS section B.2.3. The flow of participants between the parent trials and their respective OLE studies is depicted graphically in CS Figures 2 and 3. Key design features are summarised below in Table 9. The studies were conducted at sites in Europe, Asia, the Americas and Australasia with 173 participants enrolled in INPULSIS-ON study and 58 participants in the TOMORROW OLE.

Table 9 Key design features of the INPULSIS-ON and TOMORROW OLE studies

Study	Key features
INPULSIS-ON	• **Design.** A phase III open-label extension trial of the long-term safety of oral nintedanib in patients with IPF • **Eligibility.** Patients who completed the 52-week treatment period of the INPULSIS RCTs, and the 4-week follow-up visit. • **Treatment.** Patients received nintedanib up to a maximum dose of 150mg bd • **Follow up:** 68 months (CS Table 17)
TOMORROW OLE	• **Design.** A phase II open-label extension study of the long-term tolerability, safety and efficacy of oral nintedanib in patients with IPF • **Eligibility.** Patients who completed 52 weeks' treatment in the TOMORROW RCT (period 1) continued treatment in a blinded phase (period 2), until the last patient had completed 52 weeks' treatment in period 1.

	• **Treatment.** Patients in the placebo arm of the TOMORROW RCT switched to nintedanib 50mg qd during period 2. Patients received nintedanib at a range of doses between 50 mg qd and 150 mg bd in the extension • **Follow up**: Almost 8 years from start of period 1 to database lock (15th October 2015) (CS Table 17)

CS Figure 47 presents the participant flow from the INPULSIS trials to INPULSIS-ON. In summary:

- Of the 1061 patients treated in INPULSIS I and II, 807 (76.1%) completed the trials.
- Of the 807 completers, 734 (90.9%) continued in INPULSIS-ON.
- The proportions of patients continuing were similar between the parent trial arms: 430 (90.5% of the 475 randomised to nintedanib) continued on nintedanib and 304 (91.6% of patients randomised to placebo) switched to nintedanib.
- 457 (62.3%) of the 734 patients had an FVC ≤80% predicted and 277 (37.7%) had an FVC>80% predicted (company clarification response A7) at the start of the extension period.

CS Figures 43 and 44 show the participant flow from the TOMORROW trial to its OLE. Further details are provided in the company's response to clarification questions A2 and A10. In summary:

- 316 (73.1%) of the 432 randomised patients from the parent trial completed the planned observation time.
- 198 patients entered the OLE (45.8% of those originally randomised and 59.8% of those with complete observation time in the parent trial).
- 37 patients switched from placebo to nintedanib and 161 remained on nintedanib.
- The EAG notes that the company's economic model (and network meta-analysis) only uses data from the TOMORROW trial and/or its OLE from patients who were originally randomised to placebo (n=85) or the licensed dose of nintedanib 150mg bd (n=85). In these two groups, 71 patients entered the OLE: 37 patients switched from placebo to nintedanib and 34 patients continued on nintedanib 150mg bd in the OLE
- The proportion of patients with FVC >80% predicted entering the OLE was not reported in the CS.

3.2.1.6 Open label extension studies: Patients' baseline characteristics

Table 10 shows the baseline characteristics of the patients entering INPULSIS-ON. In response to clarification question A5.b), these represent characteristics at the point of

entering the OLE. These characteristics were similar to that reported for all patients at the start of the parent trials (CS Table 10) with the exception of mean baseline FVC % predicted which was, on average, slightly lower than the baseline in the parent trial (76.21% in INPULSIS-ON compared to 78.1 % to 80.5% in the parent trials). The baseline characteristics for the subgroup of patients with FVC >80% predicted (n=277) at the start of the INPULSIS-ON study were not provided in the CS. Table 10 Baseline characteristics of participants in INPULSIS-ON

Baseline characteristic	INPULSIS-ON (n=734)
Male, n (%)	587 (80.0)
Age, yrs mean (SD)	67.2 (7.8)
Race, n (%)	
White	431 (58.7)
Black	2 (0.3)
Asian	215 (29.3)
Missing†	86 (11.7)
Smoking status, n (%)	
Never smoked	204 (27.8)
Former smoker	503 (68.5)
Current smoker	27 (3.7)
BMI, Kg/ m^2 mean (SD)	27.5 (4.4)
Weight, Kg mean (SD)	78.22 (16.17)
FVC, % predicted mean (SD)	76.21 (19.06)
FVC, mL mean (SD)	2622.9 (811.1)

Abbreviations: BMI, body mass index; FVC, forced vital capacity; SD, standard deviation.
†Race was not collected in patients treated at French sites as this is prohibited by French law.
Source: Table reproduced from CS Table 11

The characteristics of the patients in the nintedanib 150mg bd and placebo arms of the parent TOMORROW trial at the start of the extension phase (Table 11) were broadly similar to the characteristics of patients in these two trial arms at the start of the parent study (CS Table 12). An exception was that the FVC and FVC % predicted were lower at the start of the extension study in those who switched from placebo. This is expected in patients who did not receive any active treatment in the parent trial. Baseline characteristics for patients with FVC >80% predicted were not presented for the TOMORROW open label trial.

Table 11 Characteristics of patients in the TOMORROW OLE

Baseline characteristic	Nintedanib 150 mg bid (N=35)	Comparator† (N=37)
Male, no. (%)	28 (80.0)	23 (62.2)
Age in years, mean (SD)	67.2 (7.0)	66.2 (7.3)
Time since IPF diagnosis, years, mean (SD)	2.9 (1.1)	3.5 (1.6)
FVC, L, mean (SD)	2.7 (0.9)	2.4 (0.7)
FVC, % predicted, mean (SD)	77.1 (21.4)	73.0 (17.9)
DL_{CO}, % predicted, mean (SD)	40.1 (14.4)	38.9 (10.5)
Smoking status		
Never smoked	12 (34.3)	14 (37.8)
Ex/ current smoker	23 (65.7)	23 (62.2)

Abbreviations: BMI, body mass index; DL_{CO}, diffusing capacity for carbon monoxide; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; SD, standard deviation
†Patients in the comparator group entered the extension trial on nintedanib 50 mg daily but had the option to increase dose to nintedanib 150 mg twice daily. Dose reduction from 150 mg twice daily to 100 mg twice daily and treatment interruption were permitted in both groups for the management of AEs
Source: Table reproduced from CS Table 13.

The EAG's clinical expert advised that patients in INPULSIS-ON and TOMORROW were generally representative of patients who would be treated with nintedanib in clinical practice. The exceptions are that the trial populations are slightly younger than the average UK population (early 70's), and the distribution of ethnicity is different from the UK in INPULSIS-ON (a lower proportion of white patients).

EAG comment on included studies

The baseline characteristics of patients in the subgroups with FVC >80% predicted and FVC ≤80% predicted in the INPULSIS and TOMORROW RCTs were similar between trial arms. Expert clinical advice to the EAG confirms no apparent unexpected differences in characteristics between subgroups.

A higher proportion of patients from INPULSIS RCTs entered the INPULSIS-ON study than was the case in the TOMORROW RCT and its open-label extension study. The baseline characteristics of patients entering the extension studies were similar to baseline characteristics of their respective parent RCTs. The exception was lung function (FVC) which had declined at entry to the extension studies, though this is to be expected over time. Our clinical expert noted that patients in the INPULSIS-ON and TOMORROW OLE were slightly younger and a lower proportion were white

compared to patients commonly seen in practice; this is in keeping with observations in the parent trials.

3.2.2 Risk of bias assessment

3.2.2.1 Evidence submitted in TA379

The company critically appraised the TOMORROW and INPULSIS RCTs using the NICE recommended criteria. Their judgements are repeated in Table 15 of the current CS. In the original appraisal we agreed with the company's judgments, with the following exceptions:

- Question 5 (Were there any unexpected imbalances in dropouts between groups?): we judged this 'uncertain' (unclear risk of bias) for TOMORROW.
- Question 6 (Is there any evidence to suggest that the authors measured more outcomes than they reported?): we judged 'yes' (increased risk of bias) for TOMORROW.
- Question 7 (Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?) We judged the last observation carried forward analysis in TOMORROW to be inappropriate for addressing missing data; for INPULSIS the lack of information available on analysis methods for missing data led us to judge the risk of bias as 'unclear'.

3.2.2.2 New evidence submitted

3.2.2.2.1 Risk of bias assessment for RCT subgroup analyses

The EAG assumes that the risk of bias judgements provided in the CS Table 16 were made in relation to the whole trial population in the INPULSIS and TOMORROW RCTs. We therefore requested from the company details of drop-out rates and missing data for the subgroup of patients with FVC >80% predicted (response to clarification question A2). These are the domains in which risk of bias may potentially vary between patient subgroups.

The company clarified that planned observation time was considered as complete if all visits until week 52 and the following follow-up visit were performed.

- In the pooled **INPULSIS** RCTs, a slightly higher proportion of patients did not complete the planned observation time in the nintedanib arm (43 patients; 14.6%) compared to the placebo arm (19 patients; 10.0%), however this was largely due to differences in adverse event rates which is not unexpected.

- Drop-out rates were more variable across the trial arms (ranging from 6% to 23.3%) in the **TOMORROW** RCT but were similar between the licensed dose, nintedanib 150mg bd, (7 patients; 17.5%) and placebo trial arms (8 patients, 20.0%).

Analyses of the primary outcome in the INPULSIS and TOMORROW trials required patients to have a minimum number of on-treatment measurements (we assume a baseline measurement was also required but this is not explicitly stated in the CS). This means that patients with partially complete data could still be included in the analysis and did not need to have completed the planned observation time.

- The reported proportions of patients with missing data for the analysis of the primary outcome (annual decline in FVC) in the subgroup of patients with FVC >80% predicted range from 1.1% to 2.4% in INPULSIS and 0% to 5% in TOMORROW.
- These proportions reflect the numbers of patients excluded from the analysis rather than the numbers of missing data points over the trial visits for those who were included.
- Appropriate regression methods were used to account for missing data under the 'missing at random' assumption but no sensitivity analyses were provided to test this assumption in this subgroup analysis due to lack of statistical power.
- The EAG notes, however, that the primary analyses were robust to other missing data assumptions in sensitivity analyses conducted for the whole trial population which is reassuring.

The company provides a narrative description of the potential issues associated with analysing INPULSIS trial patients in subgroups defined by different baseline FVC % predicted values (CS pages 54-54). In particular, such analysis may be subject to chance findings when multiple analyses are performed, and lack of statistically significant interactions may reflect underpowered tests and do not necessarily indicate a lack of true difference in treatment effect between subgroups. The EAG agrees with that these are valid considerations.

3.2.2.2.2 Risk of bias assessment for the open label extension studies

The company critically appraised the TOMORROW and INPULSIS-ON open-label extension studies using the STA User Guide 2022 criteria [15] and a checklist proposed by Bowers et al.[16] which assesses reporting quality, internal validity and external validity in OLE studies. The EAG is not aware of any other standardised tools for assessing OLE studies specifically, so this approach seems reasonable. Our own assessments of the studies using these

criteria differed in some places from those of the company and we summarise these in Appendix 9.2. In brief, we note that four countries which contributed data to the TOMORROW RCT are not represented in the TOMORROW OLE. The rate of sample slippage is a potential concern because less than 50% of randomised patients from the parent trial entered the OLE which appears lower than the average (74%) reported in a review of OLE studies by Bowers et al.[16] We also observe that potential confounders and effect modifiers are not clearly identified as such in either the CS or the published paper for the TOMORROW OLE or INPULSIS-ON.

EAG comment on risk of bias in included studies

We did not note any major risks of bias in the conduct of the subgroup analyses. However, as noted by the company, these analyses may be subject to limitations commonly associated with subgroup analyses in clinical trials such as multiplicity (type I error) and lack of statistical power for interaction tests (type II error). Results of the subgroup analyses should therefore be interpreted with caution.

Similarly, we have no significant concerns with the conduct of the OLE studies, with the caveat that less than half of the patients in the TOMORROW RCT entered its OLE study. The patients who entered the INPULSIS-ON and TOMORROW OLE studies, however, did appear to be similar to their respective parent trial populations, though only a limited set of baseline characteristics were available for the OLE studies.

3.2.3 Outcomes assessment

3.2.3.1 Evidence submitted in TA379

Table 2 of the previous CS details the clinical efficacy, safety and HRQoL outcomes measured in the INPULSIS and TOMRROW trials. The primary outcome for both trials was the annual rate of decline in FVC (ml/year) at 12 months for nintedanib compared to placebo. Following expert clinical advice, the EAG concluded that the company had included the most clinically meaningful outcomes with the exception of activities of daily living which were not measured in the trials or specified in the NICE final scope. The company's economic model derived the baseline risk of mortality, disease progression (defined by a 10-point drop in FVC% predicted) and time to first acute exacerbation using outcome data from the placebo arms of the INPULSIS and TOMORROW trials. The corresponding risks for nintedanib were derived by applying an odds ratio from the company's NMA to these respective baseline risks (see section 3.5 of this report).

3.2.3.2 New evidence submitted

No new outcome measures are presented in the current CS. Selected outcomes from the parent trials are included in the post-hoc analysis of the INPULSIS and TOMORROW RCTs for the subgroup of patients with FVC>80% predicted (Table 12). We describe the outcomes measured in the open-label extension studies in Table 13.

Table 12 Outcomes reported in the INPULSIS and TOMORROW RCTs for patients with FVC >80% predicted

Outcome	INPULSIS	TOMORROW
Efficacy	Annual rate of decline in FVC (mL/ year); Change from baseline in FVC (mL/ year); Time to first acute exacerbation	Annual rate of decline in FVC
Safety	Number of adverse events (overall, severe, serious, fatal, leading to discontinuation).	Adverse events reported for *whole trial population only*
HRQoL	Change from baseline in St George's Respiratory Questionnaire (SGRQ) total score at week 52	Change from baseline in St George's Respiratory Questionnaire (SGRQ) total score at week 52 reported for *whole trial population only*

Source: CS section B.2.6, company response to clarification question A9, CS Table 37

Although these data are mostly reported for the subgroup with FVC>80% predicted, data from the whole trial population are used to inform the baseline risks in the company's model, as done in TA379. The CS does not present overall survival for the subgroup of patients in the RCTs with FVC>80% predicted, either in the form of survival curves or as a hazard ratio. However, pooled Kaplan Meier survival curves from these RCTs and their open-label extension studies are provided for the subgroup with FVC >80% predicted in the company's response to clarification question B6.

A summary of the most frequently reported adverse events is shown in CS Table 38 stratified by baseline FVC >90% vs. FVC ≤90% predicted. The most frequently reported adverse events is shown in CS Table 38 stratified by baseline FVC >90% vs. FVC ≤90% predicted.

Additional HRQoL measures were recorded during the INPULSIS trials but are not presented in the CS for patients with FVC>80% predicted (e.g. UCSD-SOBQ, PGI-C and

CASA-Q cough score). NICE's preferred HRQoL tool, the EQ-5D, was measured in the INPULSIS trials but the CS does not present the change in EQ-5D over time for patients with FVC>80% predicted. Section 4.2.7.2 of this EAG report provides further details of how the trial-based EQ-5D data are used in the economic model.

A summary of the clinical outcomes in the INPULSIS-ON and TOMORROW open-label extension study is provided in Table 13.

Table 13 Outcomes reported in open label extension studies

Outcome	INPULSIS-ON	TOMORROW OLE
Efficacy	Annual rate of decline in FVC calculated over 192 weeks; absolute change in FVC (mL and % predicted) from baseline to week 192; number and rate of acute exacerbations; mortality over 5 years	Annual rate of decline in FVC from first drug administration until 15th October 2015;
		Overall survival; progression-free survival; incidence (and %) of patients with at least one acute IPF exacerbation; Annual rate of decline in DL_{CO}
Safety outcomes	Incidence of AEs (primary outcome)	Percentage of patients with at least one AE
HrQoL	Not reported	Not reported

Source: CS Tables 4, 6, 7, 19 & 20

EAG comment on outcomes assessment

Consistent with TA379, the company include efficacy, safety and HRQoL outcomes, appropriate to IPF. Presentation of survival data for the subgroup of patients with FVC >80% in the INPULSIS and TOMORROW RCTs would have been informative to assess the consistency between these trials.

3.2.4 Statistical methods of the included studies

3.2.4.1 Evidence submitted in TA379

The statistical approach used in INPULSIS and TOMORROW RCT was reported in the previous company submission (TA379).[1] The EAG considered the approach to be appropriate with the exception of the last observation carried forward imputation method for missing data for secondary outcomes in the TOMORROW trial. [7] We considered this method increased the risk of bias.

3.2.4.2 New evidence submitted

As described in section 3.2.3 of this report, the company has provided results of post-hoc subgroup analysis for selected outcomes in patients with FVC >80% predicted in the INPULSIS and TOMORROW RCTs. We assume that the same statistical approach has been applied to these analyses as was applied to analyses conducted in the whole trial population(s). However, we note that sensitivity analyses to account for missing data (e.g. using multiple imputation techniques) do not appear to have been provided for the post-hoc subgroup analyses. As often the case with subgroup analyses, results should be interpreted with caution due to smaller sample sizes. Similarly, tests of interaction between treatments and subgroups are likely to be underpowered to detect a difference in treatment effect between subgroups.

The statistical approach for the open-label extension studies is reported in CS Table 14 and in CS Appendix M. Sample size calculations and methods to account for multiplicity were not required for these studies due to their “descriptive” efficacy and safety analyses (CS page 306). No analysis for the subgroup of patients with FVC >80% predicted were conducted for the TOMORROW OLE study.

In **INPULSIS-ON** the outcomes were analysed as follows:

- The primary outcome was the incidence of adverse events during treatment period (up to 56.3 months in total). The CS reports that event rates per 100 patient exposure-years were calculated, however CS Table 39 appears to report simple percentages. Missing adverse event dates were imputed according to company conventions (not otherwise described).
- The annual rate of decline in FVC over the full 192 weeks of the extension was calculated using a similar approach to the analysis in the parent trial (random coefficient regression). This was compared numerically with the rate of decline during the parent trial. All patients with at least one post-baseline FVC measurement were included in the analysis. Missing data were not imputed for this outcome.
- Missing data on time to death and time to acute exacerbations were accounted for through censoring, however censoring rules are not presented in the CS.
- Analyses were based on patients who received at least one dose of nintedanib in INPULSIS-ON.

- Analyses were reportedly run separately for those patients who had received nintedanib in the parent trials and those who had received placebo, however CS Table 19 presents outcomes for the whole study population.
- Post-hoc subgroup analyses were conducted for patients with FVC ≤50% vs >50% predicted and for patients with an increase/no decline in FVC % predicted vs those with declines in FVC <10% and ≥10% predicted from baseline to week 24 (CS Appendix E).

In the **TOMORROW OLE** study:

- In keeping with the parent trial, a mixed model for repeated measures was used to estimate the annual rate of decline in FVC (primary outcome) using all available assessments from first drug administration in the extension study to trial database lock (15th October 2015), up to 61.8 months.
- Handling of missing data is not described in detail in the CS.
- Analyses were based on patients who received at least one dose of nintedanib in the blinded phase of the parent trial (period 1 of TOMORROW).
- Results are presented (CS Table 20) stratified by the parent trial treatment allocation and are given separately for the whole period from the start of parent trial to end of the OLE and for the OLE phase only.

EAG comment on study statistical methods:

The statistical methods used for the subgroup analyses mirrored that of the analysis of the whole trial population(s) in the parent RCTs and were generally appropriate. However, no sensitivity analyses were performed by the company to test the assumption that missing data on FVC was 'missing-at-random' (due to lack of power).

For the OLE studies, the analyses were largely descriptive and the statistical approach appeared to be appropriate to the outcomes measured.

3.3 Efficacy results of the intervention studies

3.3.1 Evidence submitted in TA379

The results from the INPULSIS and TOMORROW RCTs were discussed by the NICE appraisal committee and a summary of the evidence can be found in the ACD committee papers.[1] In the company's submission for TA379 three subgroup analyses from the INPULSIS trials were described:

- **FVC ≤70% versus >70% of predicted value** at baseline conducted for the primary and key secondary endpoints (prespecified; no numerical data were presented)
- **FVC >90% vs. ≤90% predicted value** at baseline (post hoc; numerical data presented for the primary outcome)
- **Emphysema vs no emphysema** at baseline (post hoc; no numerical data presented).

The overall conclusion from these analyses was no statistically significant differences in outcomes by subgroup.

The open label-extension studies were ongoing at the time of TA379 in 2016 and no evidence was available to inform decision making.

3.3.2 New evidence submitted

Subgroup analyses from the INPULSIS RCTs are reported in three places within the CS: section B.2.6, section B.2.7 and Appendix E. The company provided results from a post-hoc subgroup analysis in patients with FVC >80% predicted for the TOMORROW RCT in response to clarification question A9. The EAG's summary and critique of these subgroup analyses is presented in the next section (3.3.3) and additionally in Appendix 3 of this EAG report. We also summarise the results from the INPULSIS-ON and TOMORROW OLEs (section 3.3.4).

3.3.3 Post-hoc subgroup analyses from the RCTs: FVC ≤80% vs. >80%

Post-hoc subgroup analyses of the INPULSIS trials (reported in a conference abstract[17] and/or drawn from company unpublished data on file[13]) are presented for three outcomes: adjusted annual rate of decline in FVC, time to first acute exacerbation and adjusted mean change from baseline in SGRQ total score. These data are shown in Table 14. The company also shows the change from baseline in FVC over 52 weeks for these subgroups in CS Figure 5.

Table 14 Subgroup analyses by FVC% predicted ≤80% versus >80%

Outcome	baseline FVC >80% predicted			baseline FVC ≤80% predicted		
Adjusted annual rate of decline in FVC, mL/year	Nintedanib n=295	Placebo n=190	difference	Nintedanib n=343	Placebo n=233	difference
	-99.6	-228.0	128.4 mL (95% CI: 78.0, 178.8)	-125.7	-220.5	94.8 mL (95% CI: 48.3, 141.4)
	Treatment-by-time-by-subgroup interaction p=0.4959					
Time to first acute exacerbation	Hazard ratio: 0.49 (95% CI: 0.17, 1.35) in favour of nintedanib			Hazard ratio:0.72 (95% CI: 0.41, 1.27) in favour of nintedanib		
	Treatment-by-subgroup interaction p=0.6505					
Adjusted mean change from baseline in SGRQ total score at week 52	Nintedanib n=278	Placebo n=185	difference	Nintedanib n=331	Placebo n=228	difference
	2.99	4.05	−1.07 (95% CI: −3.45, 1.32)	4.04	5.71	−1.66 (95% CI: −3.97, 0.64)
	Treatment-by-subgroup interaction p=0.5814					
Source: CS text pages 62-64, CS Figure 4, CS Figure 6						

Post-hoc subgroup analyses of the TOMORROW trial were provided in response to clarification question A9 for the primary outcome only. The EAG notes that p-values here are nominal as this was not a prespecified analysis. Numerically, the greatest observed difference is for the nintedanib 150mg bd arm (-9mL decline in FVC/year) relative to placebo (-185mL decline in FVC/year) in patients with FVC >80% predicted (Table 15).

Table 15 Rate of decline in FVC (L/year) at 12 months* by FVC % predicted at baseline, observed cases (TOMORROW trial)

	Treatment	N patients in RS	N analysed patients	Adjusted rates (SE)**	Adjusted rates of difference (SE)**	95% CI	p-value***
FVC >80% predicted							
No	Placebo	47	45	−0.188 (0.049)			

	Treatment	N patients in RS	N analysed patients	Adjusted rates (SE)**	Adjusted rates of difference (SE)**	95% CI	p-value***
	Nintedanib 50mg qd	44	42	−0.219 (0.052)	−0.030 (0.071)	−0.170, 0.109	0.6718
	Nintedanib 50mg bid	41	41	−0.274 (0.050)	−0.086 (0.070)	−0.223, 0.051	0.2194
	Nintedanib 100mg bid	36	35	−0.221 (0.055)	-0.032 (0.074)	−0.177, 0.112	0.6607
	Nintedanib 150mg bid	45	44	−0.118 (0.055)	0.071 (0.074)	−0.074, 0.216	0.3384
Yes	Placebo	40	38	−0.185 (0.053)			
	Nintedanib 50mg qd	43	43	−0.133 (0.052)	0.053 (0.074)	−0.093, 0.199	0.4777
	Nintedanib 50mg bid	45	45	−0.154 (0.048)	0.031 (0.072)	−0.110, 0.172	0.6631
	Nintedanib 100mg bid	50	50	−0.124 (0.045)	0.062 (0.069)	−0.074, 0.198)	0.3733
	Nintedanib 150mg bid	41	40	−0.009 (0.053)	0.177 (0.075)	0.030, 0.323	0.0182

RS, randomised set (all randomised patients whether treated or not)
The p−value for the interaction FVC %pred > 80% * treatment for the model including the subgroup and the interaction term FVC %pred > 80% * treatment is: 0.1408.
* Based on visits up to visit 9
** Based on a Mixed linear regression Model repeated measures with terms for treatment*time, gender*age, subject effect, subject*time, treatment, (subject effect and subject*time random, all other effects fixed) and a variance component variance−covariance matrix
*** Nominal p−value
Source: Response to clarification question A9 received on 4th August 2022

3.3.4 Results from the open-label extension studies

3.3.4.1 Clinical outcomes from INPULSIS-ON

The company presents the clinical outcomes from the INPULSIS-ON study in CS Table 19:

- Participants treated with nintedanib for 52 weeks in the parent INPULSIS trials (not stratified by FVC % predicted) had an adjusted annual rate of decline in FVC of -113.6 mL.
- In comparison, over the 192 weeks of INPULSIS-ON, the adjusted rate of decline in FVC for all patients treated with nintedanib (i.e. also including placebo patients newly treated with nintedanib when they entered the open-label extension) was -135.1 mL.
- The company suggests that the 22 mL difference in the adjusted rate of decline at 192 weeks vs 52 weeks is not clinically meaningful because the minimum clinically important

difference in FVC% predicted of 2-6% would equate to 75-80 mL for patients in INPULSIS-ON. Our clinical expert agreed that this difference is not clinically meaningful.

In response to clarification question A8, the company provides further details of the rate of decline in FVC in INPULSIS-ON stratified by baseline FVC % predicted:

- This analysis showed a slightly higher rate of decline in the subgroup with FVC >80% predicted in INPULSIS-ON (-133.60mL) in the nintedanib group than observed for the same subgroup in the pooled INPULSIS trials (-99.57 mL) i.e., a difference of 34 mL.
- The company states that this is still a clinically insignificant difference (i.e. suggesting that the effect of nintedanib on slowing IPF progression persists over the longer-term), Again, our expert agreed this was not a clinically significant difference.

Additional outcomes are reported for INPULSIS-ON including post-hoc subgroup analyses of patients with FVC >50% predicted vs ≤50% predicted and patients with/without a decline in FVC ≥10% at the end of the INPULSIS parent trials. (CS text pages 73-74 and Appendix E).

3.3.4.2 Clinical outcomes from the TOMORROW open-label extension

The company presents the clinical outcomes from the TOMORROW OLE in CS Table 20.

- For participants who received nintedanib 150mg twice daily (licensed dose) in the 52-week TOMORROW RCT (period 1), continued to receive nintedanib during the blinded phase (period 2) and who then entered the open-label extension, the adjusted annual rate of decline in FVC was −125.4 mL/year (95% CI: −168.1 to −82.7).
- For participants who received placebo in the TOMORROW RCT (period 1), who were switched to nintedanib during the blinding phase (period 2) and who continued to receive nintedanib in the open-label extension, the adjusted annual rate of decline in FVC was −189.7 mL/year (95% CI: −229.8 to −149.6).

3.4 Safety results of the intervention studies

3.4.1.1 Safety outcomes from the RCTs

The safety results from the TOMORROW RCT and INPULSIS RCTs were provided for TA379 and can be found in the company's current submission in Appendix F. Diarrhoea was the most frequently reported adverse event in patients allocated to the nintedanib 150mg bd arm in the INPULSIS trials (398 patients; 62.4%) and the TOMORROW trial (47 patients; 55.3%).

The company presents two subgroup analyses of safety data (from the INPULSIS trials only) patients stratified by baseline FVC >80% vs. ≤80% predicted (Table) and by baseline FVC >90% vs. ≤90% predicted (see Appendix 9.3.4 of this report). A greater proportion of people receiving nintedanib in the baseline FVC >80% predicted subgroup experienced a severe or serious adverse event or an adverse event leading to treatment discontinuation. Adverse events rates were more comparable for the subgroup of patients with FVC ≤80% predicted. The EAG notes that a higher proportion of patients had one or more serious (or severe) adverse event in the subgroup with baseline FVC ≤80% predicted (nintedanib and placebo arms) compared to the FVC >80% predicted (nintedanib and placebo arms).

Table 16 Adverse events in INPULSIS trials by baseline FVC >80% vs. FVC ≤80% predicted

Event n (%)	Baseline FVC>80% predicted		Baseline FVC ≤80% predicted	
	Nintedanib (n=295)	Placebo (n=190)	Nintedanib (n=343)	Placebo (n=233)
AE(s)	277 (93.9)	167 (87.9)	332* (96.8)	211 (90.6)
Severe AE(s) [a]	76 (25.8)	30 (15.8)	98 (28.6)	69 (29.6)
Serious AE(s) [b]	80 (27.1)	44 (23.2)	114 (33.2)	83 (35.6)
Fatal AE(s)	11 (3.7)	6 (3.2)	26 (7.6)	25 (10.7)
AE(s) leading to treatment discontinuation [c]	66 (22.4)	14 (7.4)	57 (16.6)	40 (17.2)

Source: CS Table 37 edited by the EAG
Abbreviations: AE, adverse event.
[a] An event that was incapacitating or that caused an inability to work or to perform usual activities.
[b] An event that resulted in death, was immediately life threatening, resulted in persistent or clinically significant disability or incapacity, required or prolonger hospitalisation, was related to a congenital anomaly or birth defect, or was deemed serious for any other reason.
[c] AEs leading to treatment discontinuation in >2% of patients in any treatment group.
* The EAG have corrected this value as per company response to clarification question A11.

3.4.1.2 Safety outcomes from the open-label extensions

The frequencies of adverse events in the INPULSIS-ON and TOMORROW OLE are summarised in Table 17. The proportions of patients experiencing severe or serious adverse events and events leading to discontinuation were higher in the OLE studies when compared to the parent trials. In keeping with the observations of the parent trials, diarrhoea was the most frequently reported adverse event in the INPULSIS-ON trial (519 patients; 70.7%) and

the TOMORROW-OLE (63 patients in nintedanib 150mg bd dose group; 74.1%). Again, event rates for diarrhoea were higher in the OLE studies compared to the parent trials. This could potentially be explained by patients switching from placebo to nintedanib begin to experience these adverse events in the extension.

Table 17 Adverse events in INPULSIS-ON and TOMORROW OLE

Event n (%)	INPUSIS-ON (n=734)	TOMORROW OLE	
		Nintedanib 150 mg Twice daily (n=85)	Comparator† (n=85)
≥1 AE(s)	723 (98.5)	84 (98.8)	83 (97.6)
≥1 Severe AE(s) [a]	412 (56.1)	41 (48.2)	50 (58.8)
≥1 Serious AE(s) [b]	506 (68.9)	47 (55.3)	55 (64.7)
Fatal AE(s)	Not reported	12 (14.1)	31 (36.5)
≥1 AE(s) leading to treatment discontinuation [c]	313 (42.6)	48 (56.5)	49 (57.6)

Source: CS Tables 39 & 40 edited by the EAG
Abbreviations: AE, adverse event.
[a] An event that was incapacitating or that caused an inability to work or to perform usual activities.
[b] An event that resulted in death, was immediately life threatening, resulted in persistent or clinically significant disability or incapacity, required or prolonger hospitalisation, was related to a congenital anomaly or birth defect, or was deemed serious for any other reason.
[c] AEs leading to treatment discontinuation in >2% of patients in any treatment group.

3.5 Critique of the network meta-analysis (NMA)

3.5.1 Evidence submitted in TA379

In the absence of any head-to-head trials comparing nintedanib with pirfenidone, an NMA was constructed to allow an indirect comparison of these two treatments. The nintedanib outcome data used in the NMA were from the placebo-controlled INPULSIS I and II and TOMORROW trials. The pirfenidone comparator trials included in the NMA were also all placebo-controlled RCTs; therefore all comparisons were made via placebo. A total of nine outcomes were included in the NMA, of which six informed the economic model (mortality, acute exacerbations, loss of lung function, serious cardiac events, serious gastrointestinal events, overall discontinuations). For each outcome measure a series of scenario analyses examined the effect of removing specific studies from the analysis due to differences in potential effect modifiers (e.g. duration of disease).

3.5.2 Current NMA approach

The CS states that the NMA has not been updated from TA379 as no new relevant nintedanib RCTs were identified. Instead, “only results relevant to the scope of the decision problem are presented” (CS page 74). The EAG interprets this to mean that only results for the comparison of nintedanib vs placebo are given; results of the indirect comparison between nintedanib versus pirfenidone are not included as the latter is outside the current decision problem.

3.5.2.1 Outcome measures included

The economic model in the current submission uses the original (i.e. 2015) NMA effect estimates for the following outcomes: acute exacerbations, loss of lung function, serious cardiac events, serious gastrointestinal events and overall treatment discontinuation. Survival estimates in the model are no longer informed by the NMA – as we discuss below. The EAG cross-checked the NMA results presented in the current CS for the above outcomes with those reported in TA379 and found that they were consistent, as would be expected. (NB. We checked against the committee papers for TA379, noting that NMA results in the company submission for some outcomes were later superseded by corrected NMA results provided by the company in response to EAG clarification questions). The EAG assumes that given the absence of data from new trials, the company have retained the NMA estimates in order to maintain consistency with TA379.

In the current economic model a different approach is used to that of TA379 for extrapolating overall survival (OS). Individual parametric survival curves were fitted to both the nintedanib and placebo arms given some (inconsistent) evidence of an early proportional hazards violation (CS section B.3.3). Thus, the original NMA ORs for OS no longer inform the economic model (see section 4.2.6 of this report for further detail).

3.5.2.2 NMA patient population

The NMA patient population is people with IPF regardless of their baseline FVC % predicted value. We asked the company to rerun the NMA restricting the patient population to those with FVC >80% predicted, where feasible. The company declined, stating that “no significant treatment by subgroup interactions for the primary or secondary endpoints were observed hence the cost-effectiveness model is based on the treatment effect obtained from the NMA results for the overall population for nintedanib versus placebo” (clarification question response B11). However, in the CS the company also acknowledges that the INPULSIS trials were not designed to investigate the effects of nintedanib in subgroups and therefore

“the interaction tests were likely underpowered, and as such, lack of significance does not necessarily imply the absence of a true, underlying difference” (CS page 54). In the EAG’s opinion it is plausible that the non-significant results of the interaction tests are due to lack of statistical power, a consequence of reduced numbers of patients in the subgroups. Therefore, we consider it equally justifiable to restrict the NMA to the FVC >80% predicted subgroup as it is not to restrict the NMA to this subgroup. In other words, both the EAG’s and the company’s preferred approaches to the NMA population should be considered.

3.5.2.3 Purpose of the NMA in the current appraisal

The above issues, however, are eclipsed by the conclusion we have reached which is that, given an indirect comparison between nintedanib against pirfenidone is no longer required, the NMA is effectively redundant. Instead, the EAG suggests that a pairwise meta-analysis of nintedanib versus placebo from the INPULSIS I and II and TOMORROW trials would be sufficient The company do not comment on the purpose of the NMA in the current appraisal, nor whether there are advantages or disadvantages from its inclusion. The EAG notes a potential benefit of the NMA is greater precision of effects from the increased number of placebo participants in the network (i.e. placebo participants from the INPULSIS and TOMORROW trials as well as the placebo participants from the pirfenidone trials). However, a potential disadvantage of the NMA is increased heterogeneity and consequent confounding of effects caused by differences between the nintedanib and pirfenidone trials in study characteristics. Moreover, the pirfenidone placebo trial arms do not include patients with FVC >80% predicted and thus could not be included in any NMA restricted to this subgroup. Hence, this is another reason why a pairwise nintedanib vs placebo comparison would be more appropriate to inform this appraisal.

EAG comment on the NMA

With the exception of survival, the company use the same NMA effect estimates from TA379 for the clinical effectiveness and safety outcomes in the base case economic model. The estimates are, therefore, based on the whole trial population rather than the FVC >80% predicted subgroup. Given that pirfenidone is no longer a relevant comparator treatment in the decision problem, the EAG suggests a more appropriate approach would be a pairwise meta-analysis of nintedanib versus placebo from the INPULSIS I and II and TOMORROW trials, stratified by FVC% predicted subgroups. The CS reports the results of the pooled analysis of the INPULSIS trials alongside the results of the NMA. The EAG notes that the results of these two sets of analyses (based on the whole trial population) are similar.

3.6 Additional work on clinical effectiveness undertaken by the EAG

None

4 `COST EFFECTIVENESS

4.1 EAG comment on the company's review of cost-effectiveness evidence

The company conducted a systematic literature review to identify cost-effectiveness studies and economic evaluations published since September 2014, which evaluated nintedanib and its comparators in adults with IPF. The company completed searches in relevant electronic databases, conference proceedings and Health Technology Assessment (HTA) databases (CS Appendix G 1.1). The electronic searches were supplemented by hand searching to identify other published or unpublished material (grey literature). The search strategy was not limited by country, language, study design or date, but the company limited their full text review of studies to those published in English (CS Appendix G1.1). Databases were searched on 14 January 2022. Eligibility criteria are described in CS Appendix G Table 145.

Six publications were included after full text screening; two were considered by the company as relevant to UK clinical practice: Rinciog et al. (2017)[18] and Loveman et al. (2014).[19]

Rinciog et al. conducted an NMA and developed a cost-effectiveness model assessing the cost-effectiveness of nintedanib vs. pirfenidone, N-acetylcysteine and placebo (best supportive care) for the treatment of IPF.[18] The evaluation used pooled patient-level data from three randomised RCTs of nintedanib: the phase II TOMORROW trial[20] and two phase III INPULSIS trials (INPULSIS-1 and INPULSIS-2[21]). In keeping with the decision problem, the CS discusses the results for the comparison of nintedanib versus best supportive care, but it includes patients with a starting FVC ≥50% predicted. Rinciog et al.[18] is the published version of the model submitted for the company's original submission for nintedanib (TA379).[1]

Loveman et al.[19] reports a systematic review and an economic evaluation of the clinical and cost effectiveness of IPF treatments, and this was discussed in the original CS in TA379. The current CS points out that the NMA and cost-effectiveness model did not include the INPULSIS[21] trials, and also that the estimated cost of nintedanib did not match the list price.

Consequently, the current economic evaluation follows the same approach used in TA379 as detailed in Rinciog et al.[18] with addition of evidence from the nintedanib OLE studies. The Rinciog et al.[18] publication does not include the stopping rule for patients treated with nintedanib whose predicted FVC falls by more than 10% in a year (as specified by the NICE recommendations in TA379 for nintedanib). The CS presents details of the study, and base-case results (Appendix G Table 148).

EAG conclusion

The company's review of the economic evaluation evidence was thorough and appropriate and the EAG is not aware of any additional relevant economic evaluations.

4.2 Summary and critique of the company's submitted economic evaluation by the EAG

4.2.1 NICE reference case checklist

Table 18 shows the EAG's assessment of the concordance between the company's economic evaluation and the NICE reference case. We consider that the company's model is consistent with the reference case.

Table 18 NICE reference case checklist

Element of health technology assessment	Reference case	EAG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and PSS	Yes
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes
Synthesis of evidence on health effects	Based on systematic review	Yes
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes
Equity considerations	An additional QALY has the same weight regardless of the other	Yes

	characteristics of the individuals receiving the health benefit	
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes
PSS: Personal Social Services		

4.2.2 Model structure

4.2.2.1 Overview of the model structure

The company proposes using the same Markov model in the current appraisal as previously developed for NICE TA379[1], in which a three-month cycle length is employed in line with observation periods in the clinical trials. Half-cycle correction was applied in the model. The company maintained that the original structure used in the TA379 economic model was appropriate for the current submission with the justification that survival evidence from long-term follow-up studies can be included without the need to alter the original model structure. The model, implemented using Microsoft Excel, represents IPF lung function decline using an established clinical measure, FVC% predicted, for the health states. FVC% predicted was selected to represent health states due to its consistent use in clinical trials in IPF patients and the ability to reflect the absolute state of patient condition in the model. Figure 1 depicts the company's model structure (CS Figure 10).

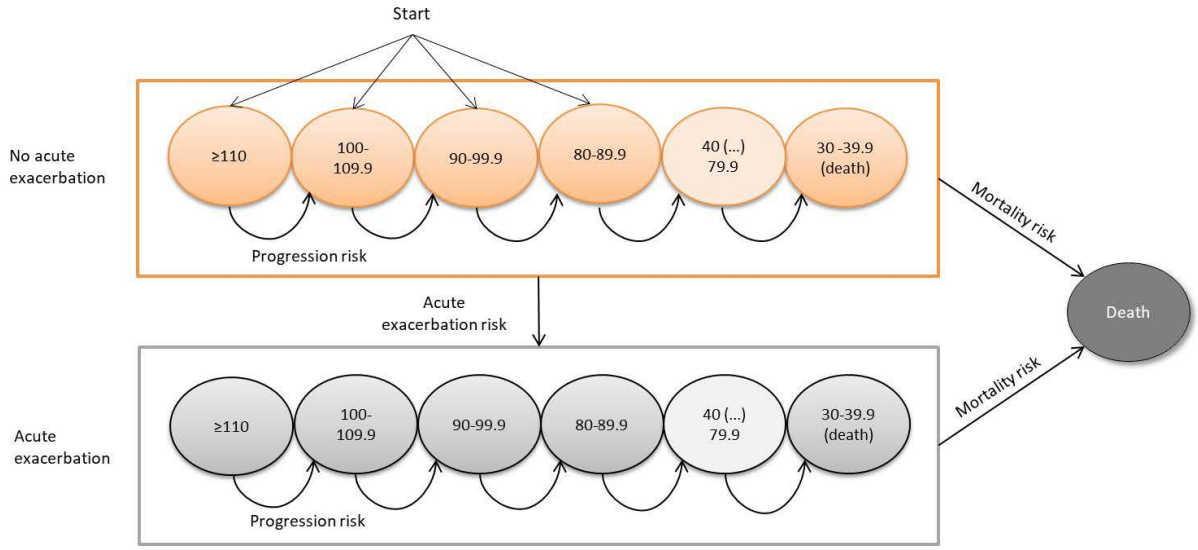

Figure 1 Model structure
Reproduced from CS Figure 10.

Health states are defined by 10-point percentage intervals in FVC% predicted from 30-39.9 to ≥110, with the lower FVC% predicted category representing death due to insufficient lung function. There are health states for patients who have not yet experienced an exacerbation event, and for patients who have experienced at least once exacerbation event. Patients who have exacerbations possess different health outcomes and costs compared with those who have not had an exacerbation. The final health states, death, is an absorbing state. Patients can move from their current health state to the death state at any point during the model.

At the start of the model, all patients begin in one of the non-exacerbation health states with FVC >80%. The distribution of patients among the initial four health states at the start of the model is based on the distribution of patients in the INPULSIS-1 and INPULSIS-2 trials,[21] detailed in Section 4.2.3. The starting age of all patients is stated as 66.75 years.

Patients can progress to different health states in the following ways: (1) loss of lung function; (2) exacerbation; (3) loss of lung function and exacerbation; and (4) death. Loss of lung function is a 10% decrease in FVC% predicted within 3 months (constant risk). Once a patient has progressed to a lower health state, i.e., a health state corresponding to a lower FVC% predicted category, the patient is unable to move back to a higher health state. Furthermore, once a patient experiences an exacerbation event and moves from a non-exacerbation health state to an exacerbation health state, the patient is unable to move back to a non-exacerbation health state. There is also an additional mortality hazard rate associated with patients in exacerbation health states; this parameter was not included in TA379. The model also allows for further adverse events including serious cardiac and gastrointestinal (GI) events, GI perforations, and mild-moderate diarrhoea.

The primary outcome measure of the economic model is incremental cost per QALY (ICER), although cost per life years (LYs) gained and exacerbation events avoided are also considered. In accordance with NICE IPF guidelines[22], the company did not explicitly model patients who transitioned from the FVC 40-49.9% predicted to the FVC 30-39.9% predicted health states, as the latter health state is assumed to be an unsustainable level of lung function; thus, the 30-39.9 health state is considered as representing death. The company assumes independence between mortality and loss of lung function in order to avoid double counting as the overall survival data includes all deaths. The EAG considers this to be a reasonable approach.

EAG comment on model structure

The economic model structure in the current appraisal is based on the model previously accepted by the NICE appraisal committee in TA379.[1] The EAG has no concerns regarding the model structure currently presented.

4.2.3 Population

The modelled population is adults with IPF. The analysis uses pooled data from the phase III INPULSIS 1 and 2 RCTs, the INPULSIS-ON OLE study, the phase II TOMORROW RCT and the TOMORROW OLE study. The baseline characteristics of the nintedanib and placebo patients are shown in CS Table 45. The baseline age in the model was 66.76 years. In accordance with the NICE scope for this appraisal, patients entering the model had an FVC >80% predicted. The distribution of FVC% predicted thresholds at baseline is shown in Table 19 (CS Table 60).

Table 19 Distribution of FVC % predicted in patients at the start of the model

Health state (FVC% predicted)	Distribution (%)
≥110	13.14%
100-109.9	16.43%
90-99.9	27.10%
80-89.9	43.33%
40-79.9	0.00%

Reproduced from CS Table 60.

EAG comment on model population

The EAG agrees that the economic model uses a population consistent with the NICE scope for this appraisal.

4.2.4 Interventions and comparators

The economic model compares the incremental cost effectiveness of nintedanib 150 mg twice daily to best supportive care. The intervention and comparator are consistent with the NICE scope.

4.2.5 Perspective, time horizon and discounting

The company analyses take the perspective of the NHS and Personal Social Services (PSS) in England, which aligns with the NICE manual for health technology assessments.[23] Costs and outcomes (life years and QALYs) are discounted at 3.5%. The company uses a lifetime horizon to reflect the chronic nature of IPF, where lifetime is assumed to be 50 years from the start of the model. Given that the starting age of the patient population is approximately

67 years, a shorter time horizon of 35 years is deemed more appropriate and used in the EAG base case analyses in section 6.1.

4.2.6 Treatment effectiveness and extrapolation

The clinical effectiveness parameters used in the model consist of OS, acute exacerbation, loss of lung function, treatment discontinuation and adverse events. Data from these studies have been taken from the TOMORROW trial and extension study, INPULSIS 1 and 2 trials, and the INPULSIS-ON extension study. More details on the extension studies are given in section 3.2 of this report. All the data used for the clinical effectiveness parameters were from the full trial populations, rather than for the FVC >80% predicted subgroup. The EAG considers that OS estimates for this subgroup should be included in the analysis to reflect the lower mortality rate for these patients.

4.2.6.1 Mortality (overall survival)

The company checked whether the proportional hazards (PH) assumption is supported by visual inspection of the log-cumulative hazard plot (CS Figure 13) and assessment of the Schoenfeld residuals. They concluded that the PH assumption does not hold as the lines in the figure are non-parallel and therefore the ratio of the hazard rates between arms does not remain constant over the follow-up period. As the PH assumption does not hold, independent parametric models were fitted for each treatment arm for OS.

The pooled Kaplan-Meier (KM) survival curves for nintedanib and placebo are shown in Figure 2 (CS Figure 12). The duration of follow-up for nintedanib is approximately 5.5 years which is longer than it was in the original appraisal in 2016 (NICE TA379). Further, nintedanib has markedly better survival probability than was predicted in the previous appraisal at 5 years (60% vs 40%).

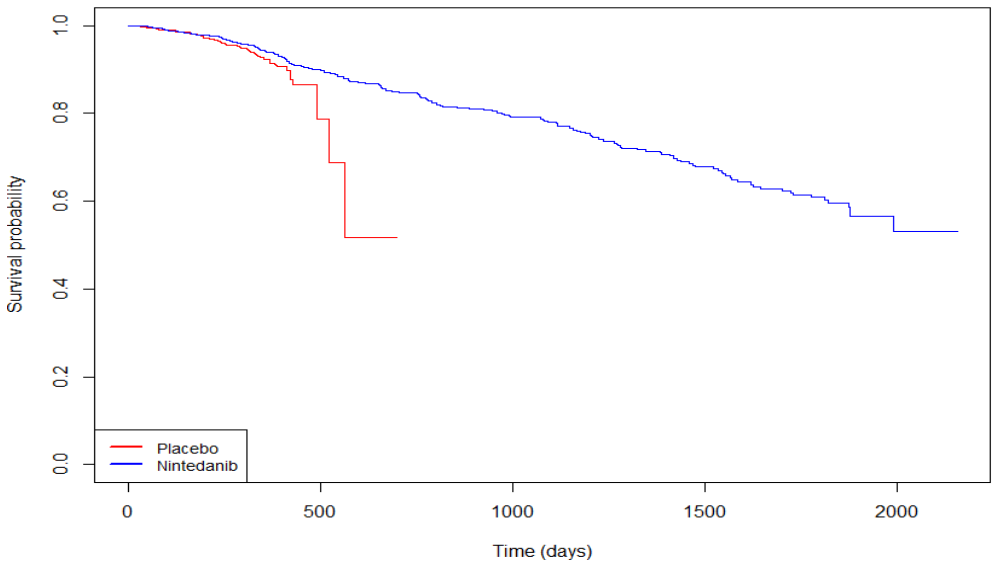

Figure 2 Kaplan-Meier curves of overall survival for nintedanib vs placebo

Reproduced from CS Figure 12

The parametric models, fitted using R software, were: exponential, Weibull, log-logistic, Gompertz, lognormal and generalised gamma. Goodness of fit was assessed using visual inspection and statistical fit using Akake/Bayesian Information Criterion (AIC/BIC). The AIC/BIC values are shown for both treatment arms in CS Tables 47 and 48.

In the nintedanib arm, the models with the best fit by AIC/BIC are the log-logistic, Weibull and generalised gamma. For the placebo arm, the models with the best fit are the Gompertz, Weibull and log-logistic. As the trial data available for the placebo is only 52 weeks, the company also compares the model fit with external data from a study of risk factors for acute exacerbations in IPF by Kondoh et al[24] (CS Figures 15 and 16) and an Australian IPF registry[25] (CS Figure 22). The CS states that in TA747, clinical experts and the NICE committee agreed that the Australian registry was most representative of UK clinical practice. The company concludes that the log-logistic model is the most suitable and used this in their base case. The parameter values for the parametric models are shown in CS Table 49 and 51. The CS also compares data from other international registries and these are discussed in more detail in section 5.3.3.

The EAG agrees with the log-logistic parametric model chosen for OS based on the statistical fit and visual fit to the Australian IPF registry and the Kondoh et al.[24] study. The extension studies report follow-up data for nintedanib in excess of five years, which provides more certainty in this treatment arm. The EAG also notes that the trial results are consistent with those from the Australian IPF registry for the nintedanib arm. The duration of follow-up

data for the placebo arm, however, remains relatively short at under two years. The EAG agrees that survival for placebo patients is likely to be similar to that reported in the Australian IPF registry.

In addition, the company has included a hazard ratio (HR) for the risk of mortality for patients who have had an exacerbation (not previously included in TA379). A HR of 1.395 was applied to these patients, based on the study by Kondoh et al.[24] which reported a HR of 2.79 for a six-month period; the company halved this value to account for the 3-month cycle length. The EAG considers it is appropriate to include a HR for these patients, however it is inappropriate to divide the hazard ratio by two as the HR is independent of time. The EAG notes that the HR from Kondoh et al[24] is consistent with a study by Kakugawa et al.,[26] which investigated risk factors for acute exacerbations in patients with IPF.

4.2.6.1.1 Mortality for the FVC >80% predicted subgroup

The EAG requested further information on the OS estimation of patients with a baseline FVC >80% predicted (clarification questions B5 and B6). The company provided a Kaplan-Meier plot for the overall population and patients with baseline FVC >80% predicted, reproduced in Figure 3 (clarification response document Figure 4). As with the full dataset, there are only 52 weeks follow-up for the placebo arm and more than five years follow-up for the nintedanib arm.

The company fitted parametric survival curves independently to the nintedanib and placebo (best supportive care) treatment arms using the methods described above. The log-logistic model was selected based on AIC/BIC statistical criteria and visual inspection. The parameters for the parametric models are shown in Tables 10 and 11 in the clarification response document. The company conducted a scenario analysis using the log-logistic, Weibull and lognormal parametric models. The resulting ICERs ranged from [redacted] to [redacted] per QALY.

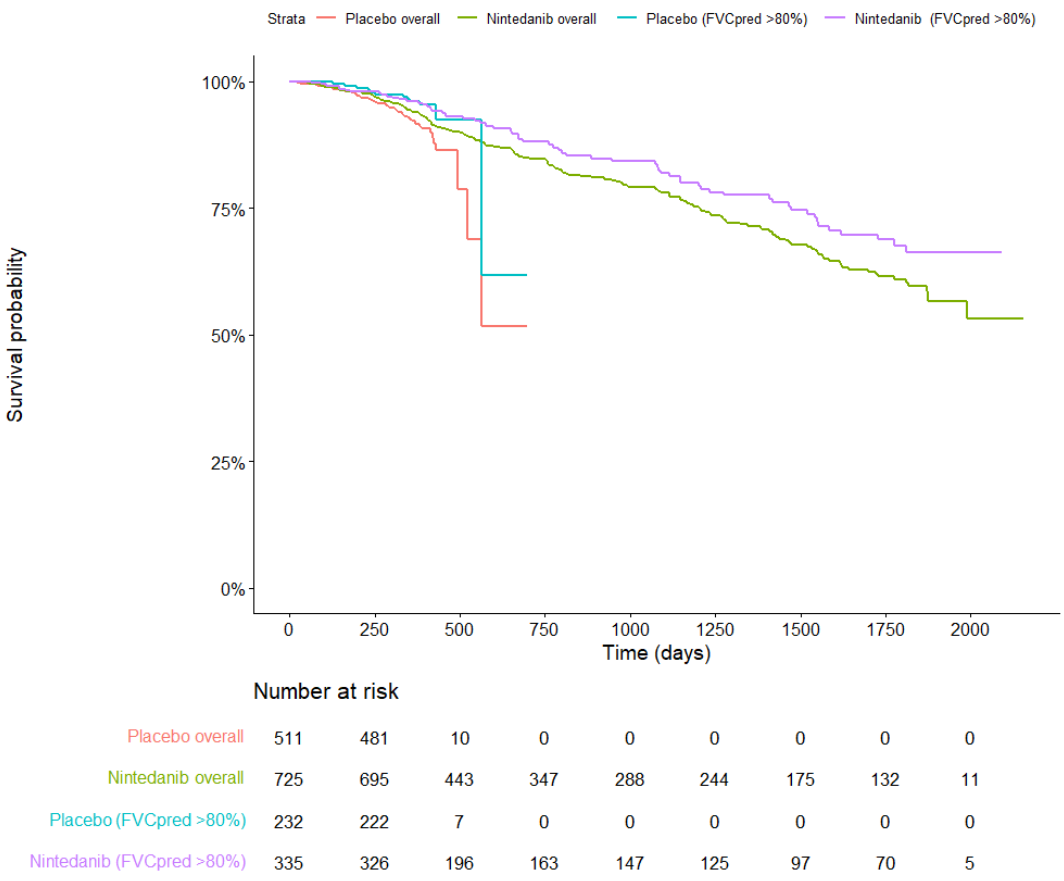

Figure 3 Kaplan-Meier curves for the overall population and patients with baseline FVC >80% predicted

Reproduced from Figure 4 of company clarification response document.

The clarification response document specifies how the OS curve for the placebo arm was fitted. However, the long-term extrapolation is uncertain due to the fact there is only one year follow-up data. The EAG notes that there is no difference between the OS curves for the nintedanib and placebo arms in the FVC >80% predicted subgroup for the first 52 weeks.

For the nintedanib arm, the EAG agrees with the company's fitted curve. Based on no difference between the OS curves in the KM data (Figure 3), the EAG assumes that mortality is initially the same for the best supportive care arm and the nintedanib arm in the FVC >80% predicted cohort. The initial mean FVC % predicted of the cohort is 95% and declines over time. When the modelled cohort reaches the same FVC % predicted as the whole trial population (FVC = 79%), we assume the best supportive care OS curve has the same survival as the whole trial population. Therefore, from this point the best supportive care arm uses the whole trial parametric curve. We estimate it to be 5.5 years before the mean FVC for the cohort is the same as the whole trial FVC.

A study by Jo et al.[25] reports the OS of a cohort with mild impairment (FVC>80%) in the Australian IPF registry (median follow-up 2.1 years). Of this cohort, 25% were receiving anti-fibrotic therapy. We compare the modelled survival for best supportive care using the company and EAG assumptions against this study (Table 20). The company's estimate for survival over four years is notably lower than the survival data from the Australian registry.[25]

Table 20 Company and EAG OS estimates for FVC>80% subgroup vs Australian IPF registry (mild[a] patients)

Year	Australian IPF registry (mild FVC)[25] OS	Company base case for BSC OS	EAG base case for BSC OS
0	100	100	100
1	99	96	96
2	89	80	88
3	73	58	79
4	71	39	70

[a] Patients were classified as 'mild' if FVC ≥80%

BSC, best supportive care; OS, overall survival

4.2.6.2 Acute exacerbations

The risk of first acute exacerbation was estimated from the INPULSIS trials. The company considered the risk was best represented as a constant risk. The exacerbation risk was 1.47% per 3-month cycle for the placebo arm for the adjudication committee estimate and 1.97% for the investigator-reported estimate. In the base case analysis, the adjudication committee-reported exacerbation risk was used, while the investigator-reported value was used in sensitivity analyses. The EAG notes that the investigator-reported value was used in the base case in the previous appraisal TA379.

The risk of exacerbation for nintedanib was informed by the NMA ORs applied to the baseline placebo risk. The OR value for nintedanib vs placebo is 0.56 (95% CI 0.35 - 0.89). (CS Table 59).

4.2.6.2.1 Acute exacerbations for the FVC >80% predicted subgroup

In response to clarification question B9, the proportion of patients with an acute exacerbation and the hazard ratio for time to first exacerbation event in patients with FVC above and below 80% predicted is shown in Table 15 of the company clarification response document.

The company comments that acute exacerbations were a rare event in the overall population in the INPULSIS trial and even more so in the subgroup with the FVC >80% predicted. For this reason, they did not consider it possible to run a scenario based on acute exacerbations in the subgroup with FVC >80% predicted. The subgroup analyses for FVC% predicted ≤80% versus >80% are shown in section 3.3 of this report.

The EAG agrees with the company that there is uncertainty in the results of the subgroup analyses due to the low number of events in the trials. Further, we note some unexplained inconsistencies in HRs between the FVC% predicted subgroups. For example, the HR for time to first acute exacerbation for the comparison of nintedanib versus placebo varies from 0.46 in the FVC **>**90% predicted subgroup, to 0.49 in the FVC >80% predicted subgroup and to 1.00 in the FVC **>**70% predicted subgroup. Clinical advice to the EAG is that although lower FVC (more severe IPF) is a recognised risk factor for acute exacerbation at any disease stage the actual HR benefit is most likely comparable. Hence, the HRs for the >90% and <80% FVC subgroups seem, in his opinion, more realistic. We therefore base the risk of exacerbation for nintedanib versus placebo on the whole trial population and the risks estimated for the subgroups in the EAG base case scenarios. The acute exacerbation probability per 3-month cycle for the best supportive care arm is 1.05% for the FVC >80% predicted subgroup and 2.58% for the FVC ≤80% predicted subgroup.

4.2.6.3 Loss of lung function

The company defines loss of lung function as a 10-point drop in FVC% predicted. Patients entered the model at different FVC% predicted health states to reflect the INPULSIS clinical trial as shown in CS Table 60. Lung function decline, with and without exacerbation, was incorporated using a logistic model derived from a logistic regression of the phase III clinical trial data. In both cases (i.e., with and without exacerbation), there was a diminishing effect in progression with loss of lung function; that is, the probability of progression was lower for patients with lower FVC% predicted. However, the absolute risk of progression was significantly higher when there was an exacerbation. This is graphically presented in CS Figure 29.

The risks associated with loss of lung function for nintedanib were obtained by applying ORs from a NMA to the baseline risk from the INPULSIS trials, assuming a constant hazard over time. The OR estimate for nintedanib vs placebo was 0.54 (95% CI: 0.42 to 0.69). The company investigated whether the rate of decline of lung function would be similar for the >80% predicted group. The CS states that the probabilities of progression were similar,

though slightly lower, in the >80% predicted group (CS Figure 30). The progression probabilities for the >80% predicted group were examined in a sensitivity analysis. The EAG agrees that the probabilities of progression are similar for the FVC >80% predicted group and it is reasonable to use the lung function decline from the whole population.

4.2.6.4 Treatment discontinuation

The company estimates overall discontinuation risk for the baseline best supportive care arm to be 5.5% per cycle. The associated risk for nintedanib (OR 1.42; 95% CI: 1.08 - 1.87) was calculated by applying ORs obtained from all trial evidence from the NMA to the baseline risk. The company assumes that patients would not discontinue from best supportive care, but they used this discontinuation risk to estimate the relative discontinuation risk in patients receiving nintedanib.

The company estimated the discontinuation rate for the FVC >80% predicted subgroup in response to clarification question B9. The discontinuation rate for this population is 3.8% per cycle for best supportive care. The EAG have included this discontinuation rate in our scenario analyses in section 6.2.2.

4.2.6.5 Treatment stopping rule

In TA379 the NICE committee recommended that nintedanib should be subject to a stopping rule for those patients whose FVC % predicted declines by 10% in a year. In the current CS the company dispenses with this stopping rule on the basis that:

- It was implemented to be consistent with the stopping rule for pirfenidone. However, pirfenidone is not a comparator in the current appraisal.
- Expert clinical advice to the company was that a stopping rule according to the above criteria would be difficult to impose.
- In NICE TA747 the appraisal committee noted that clinicians would stop treatment in patients with rapid disease progression, hence a stopping rule was not required.

Clinical advice to the EAG agrees that this stopping rule, based on a decline of FVC alone, is not used in routine clinical practice. The EAG notes that in TA379 the base case ICERs were not considered cost effective without the stopping rule. As discussed above in section 4.2.6.1, the OS data for nintedanib shows better survival than predicted in the previous appraisal at 5 years (60% vs 40%). Given the improvement in OS since that appraisal, the stopping rule may not be necessary, [redacted], due

to the improvement in cost effectiveness. The EAG has included scenarios using the company corrected model and the EAG base case in sections 5.3.4.1 and 6.2.2 respectively.

4.2.6.6 Adverse events

The CS model included AEs which had a substantial impact on costs and QALYs, had an incidence of more than 5%, or an incidence 1.5 times greater than the comparator arm. Serious cardiac events and serious GI events were included in the analysis. Gastrointestinal perforations were also included, based on their clinical importance.

The incidence of each of the serious AEs were estimated from the best supportive care arm and their associated risks for nintedanib were measured using OR values from the NMA presented in CS Table 71. Following recommendation from the NICE committee in TA379, mild-moderate diarrhoea was also included (CS Table 72).

EAG comment on treatment effectiveness and extrapolation

The company uses clinical effectiveness data for the whole trial population in their base case analysis, rather than using data for the FVC >80% predicted subgroup. The EAG considers that overall survival for this subgroup should be included in the analysis to reflect the lower mortality rate for these patients. The extrapolation of the best supportive care curve is uncertain, because patients receiving placebo were only followed up for 52 weeks in the trial. The EAG suggests an alternative assumption for modelling best supportive care whereby the initial mortality is equal for both treatment arms. Clinical effectiveness data for acute exacerbations and discontinuation are more uncertain and the EAG suggests the whole trial population effectiveness data may be appropriate for these parameters.

4.2.7 Health related quality of life (HRQoL)

4.2.7.1 Systematic literature review for utilities

The company conducted a systematic literature review in January 2022 to identify studies published since August 2014 that evaluated HRQoL in patients with IPF (CS Appendix H). The search strategy and database searches were in line with those for the cost-effectiveness review (CS Appendix G1.1). In addition, clinical trial databases were searched to identify ongoing and recently completed studies that met the inclusion criteria for the review. Eligibility criteria are given in CS Appendix H Tables 153 and 154.

Nine publications were identified after full text screening. Three studies reported HRQoL outcomes and health state utility values for relevant health states (including FVC% predicted status and adverse events of interest).[18,27,28] All three studies derived utility values from analysis of patient-level EQ-5D-3L data from the INPULSIS trials.[21] Two studies used UK preference weights to derive utilities: one was a cost-effectiveness analysis developed from the Belgian healthcare payer perspective,[28] and Rinciog et al. was a cost-effectiveness analysis from the UK NHS and Personal Social Services perspective.[18] The utility values reported were the same in both studies, so the company focusses on the analysis conducted from the UK perspective. Rinciog et al. presented utilities by FVC % predicted status, acute exacerbation-related disutility and adverse event-related disutility.[18] These values are used in the model in the current submission.

4.2.7.2 Study-based health related quality of life

The utility values applied in the model are utilities by FVC % predicted status, acute exacerbation-related disutility and adverse event-related disutility. FVC % predicted health state utility values were taken from EQ-5D 3L data from the INPULSIS trials[21] (CS B.3.4 table). These utility values were previously used in TA379.[1] They are described in CS section B.3.4 and were discussed in section 4.2.5 of the EAG report on the company's original nintedanib submission in 2015.

The economic model includes adverse events that had a substantial impact on costs and QALYs, had an incidence of more than 5% or an incidence 1.5 times greater than the comparator arm. These are: serious gastrointestinal events; serious cardiac events, gastrointestinal perforation and, at the request of the NICE Committee in TA379, mild-moderate diarrhoea.

Acute exacerbations were associated with disutility, estimated from the INPULSIS trials[29] (CS section B.3.4, CS Table 86). The model uses the investigator reported exacerbation rate in the base case and explores the effect of the adjudicated committee exacerbation disutilities in a sensitivity analysis. These disutility values were previously used in TA379 and are shown in CS Table 84.

Utility decrements for serious cardiac events, and gastrointestinal perforation were obtained from a retrospective analysis of a UK database (CS section B.3.4, CS Table 87).[30] The company's search strategy did not identify any utility values for skin disorders, dizziness or anorexia. The EAG repeated the search and concludes that there are no missing sources.

Disutility values for the adverse events are given in Table 21 (CS section B.3.4, Table 87). Disutilities associated with adverse events were based on TA379 (shown in column 2), but the duration was reduced from one year to one month (column 3) following EAG and NICE Committee feedback during the TA379 appraisal. Disutility due to diarrhoea used the same assumptions as those in TA747 (nintedanib for progressive fibrosing interstitial lung diseases[22]) and was applied for one month.

TA379 reported disutility values for skin disorders, based on the study by Ara and Brazier.[30] These utility decrements are not reported in the current CS as rashes were an adverse event associated only with pirfenidone, not nintedanib. The EAG agrees this approach is appropriate.

Table 21 Adverse events-related disutility

Event	Mean value (2015)	Mean value (2022)	Source
Serious cardiac events	-0.198	-0.0165	Ara and Brazier[30]
Serious GI	-0.068	-0.0057	INPULSIS 1 and 2[29]
GI perforation	-0.118	-0.0098	Ara and Brazier[30]
Mild-moderate diarrhoea	N/A	-0.0028	Assumption: 50% of serious GI events[22]

Reproduced from CS Table 158 and adapted by the EAG
GI: gastrointestinal; N/A: not applicable.

EAG comment on HRQoL

The company's utility values used for FVC% predicted health states and disutilities for acute exacerbations have not been changed from the previous nintedanib submission (TA379) and were previously accepted by the NICE Committee.

The utility decrements for acute exacerbations presented in the CS were taken from the INPULSIS trials. The EAG were unable to find any alternative sources of disutility for acute exacerbations.

The disutilities calculated for adverse events are appropriate, following the changes to the duration for which they are applied, reflecting committee recommendations in TA379. Disutilities for mild-moderate diarrhoea have also been included following NICE committee comments in TA379 and use the same values as were used for TA747.

4.2.8 Resources and costs

4.2.8.1 Resource use review

The company completed a systematic literature review in January 2022 to identify costs and healthcare resource use (published since September 2014) evaluating nintedanib and its comparators in adults with IPF. The search strategy and database searches are described in CS Appendix G1.1; these were supplemented by hand-searching published and unpublished material and searching appropriate registries and clinical trial databases. Eligibility criteria are given in CS Appendix I Table 164.

Following full text screening, 16 articles were included in the review, of which three were relevant to UK clinical practice.[31] Two studies reported costs associated with the treatment of IPF as part of an economic analysis, [18,31] previously discussed in section 4.1. Cost inputs (values, sources, and assumptions) used by Rinciog et al.[18] are presented in CS Appendix I Tables 168. Diamantopoulos et al.[32] conducted a retrospective analysis of 1,014 patients from the INPULSIS trials to evaluate how many hospitalisations and physician visits patients experienced over three months, and the results are presented in CS Appendix I Table 169.

4.2.8.2 Drug acquisition costs

The list price of nintedanib is £2151.10 for a 30-day supply of 60 capsules (150mg each). Dosage is two capsules a day (150mg bd), giving a cost of £71.70 per day. Nintedanib is available with a patient access scheme (PAS) price discount of [redacted], lowering the cost to [redacted] per day. The company does not associate a cost with best supportive care, because this was the placebo (control) arm of the trial. Nintedanib is taken orally and there are no associated administration costs.

4.2.8.3 Health state unit costs and resource use

The company's economic model includes the following components:

- Drug acquisition costs
- Liver function test costs
- Patient monitoring (background follow-up) costs (hospitalisation, emergency department visits, GP visits, specialist visits, physiotherapist visits, chest HRCT [high-resolution computerised tomography], chest X-ray, oxygen requirement assessment, bronchoalveolar lavage, CT [computerised tomography] pulmonary angiogram, right heart catheterization procedure, and general diagnostic procedures (e.g. bronchoscopy)
- Oxygen use costs

- Treatment-related adverse event costs
- Acute exacerbation costs (hospitalisations, emergency department visits, GP visits and specialist visits)
- End-of-life palliative care cost

Costs were calculated using UK unit cost data from the National Schedule of Reference Costs (2019-20)[33] and the PSSRU Unit Costs of Health and Social Care[34] inflated to 2020/21 values using appropriate inflation indices.[34]

The economic model uses resource data obtained from a post-hoc analysis of patient-level data from the INPULSIS trials.[29] The company analysed and adjusted health care resource use data for the model health states (FVC % predicted groups) and calculated the probability of the resource usage within a 3-month cycle. The number of resource use observations for each FVC % predicted group is shown in CS section B.3.5, Table 89.

The costs for patient monitoring for each health state were calculated as a 3-month probability of using each resource (hospitalisation, emergency department visits, GP visits, etc), weighted by the number of patients in each FVC % predicted group. Total per-cycle and annual monitoring costs for each FVC % predicted group are given in CS section B.3.5, Table 104.

The NICE draft clinical guideline for the diagnosis and management of suspected IPF states patients with IPF should receive long-term oxygen therapy to prevent resting hypoxemia.[35] The CS highlights those patients with FVC >80% predicted would not require oxygen supplementation.

The model uses the safety data set from the INPULSIS trials[29] to determine the probability of patients visiting the hospital, the emergency department, a GP, and a specialist following an acute exacerbation within a 3-month cycle. The total exacerbation cost and breakdown by health care resource are shown in CS section B.3.5, Table 108. The model uses a total exacerbation cost (£4,628) for patients in both trial arms (placebo and nintedanib) who experience a new exacerbation.

The model assumes that all patients receive palliative care (in addition to ongoing monitoring) for the last year of their lives. The cost for end-of-life care consists of hospice and home care (excluding hospital) and was estimated to be £3,037.50 per 3-month cycle (the average cost of hospital and social care for the final year of life is £12,150).[34]

Evidence from the TOMORROW[20] and INPULSIS trials[21] shows that patients taking nintedanib can experience elevated liver enzyme levels. Consequently, the company model assumes all patients on active treatment would have routine liver function tests every three months.

EAG comment on resources and costs

Costs for each FVC % predicted group were calculated in the same manner as in TA379 and TA474, which had been accepted by the NICE appraisal committee. Costs have been inflated to 2020/21 values appropriately. Resource use data given in the CS were obtained from individual patient level data from the INPULSIS trials (this is the same approach used for TA379) and are relevant to the clinical pathway of patients with IPF. The EAG are not aware of any other source of resource use data for this patient group.

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company reports their base case cost-effectiveness results in CS Tables 115 and 116 using the list price and Patient Access Scheme (PAS) price respectively. Table 22 and Table 23 below present the base case results using the list price and PAS price for nintedanib, respectively.

Table 22 Base case results for nintedanib vs. best supportive care (using list price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,262	4.08	3.21				
Nintedanib	£89,177	7.40	5.69	£69,915	3.32	2.49	£28,094

Reproduced from CS Table 115.
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.

Table 23 Base case results for nintedanib vs. best supportive care (using PAS price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,262	4.08	3.21				
Nintedanib	[redacted]	7.40	5.69	[redacted]	3.32	2.49	[redacted]

Reproduced from CS Table 116.
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.

The base case results show that nintedanib offers a mean QALY gain of 2.49 for an additional mean cost of £69,915 (list price) and [redacted] (PAS price) versus best supportive care, producing ICERs of £28,094 and [redacted] per QALY gained respectively.

5.2 Company's sensitivity analyses

5.2.1 Deterministic sensitivity analyses

The company reports results from their deterministic sensitivity analyses with the PAS discount applied, in CS Table 129, CS Figure 40, and CS Figure 41. Fourteen scenarios were considered for the one-way sensitivity analysis across four main parameters: probabilities, costs, utilities and adverse events, listed in CS Table 121. The variations in input parameters were based on the 95% confidence intervals. The company's results indicate that the discontinuation probabilities and mortality probabilities due to exacerbation are the main drivers of the model results, increasing the ICER to [redacted] and [redacted] per QALY respectively. The maximum range of the ICER in the one-way sensitivity analysis results ranges from [redacted] to [redacted] per QALY (using the PAS price for nintedanib).

[redacted]

5.2.2 Scenario analyses

The company considers 19 distinct scenarios for their scenario analyses, described in CS Tables 122 to 128, numbered from 15 to 33. The scenarios cover seven parameter groups: overall survival, exacerbations, loss of lung function, adverse events, costs, discontinuation, and FVC% predicted categories. Many of the scenarios explored involve implementing alternative odds ratios obtained from published literature.

Changing the choice of parametric distribution from the log-logistic model to the generalised gamma distribution for both nintedanib and best supportive care (scenario 16) had the largest effect on the ICER, reducing the ICER to [redacted] per QALY. All other scenarios did not have a substantial impact on the ICER. In these scenarios, the ICERs ranged from [redacted] per QALY when transition probabilities for FVC >80% predicted and an alternative odds ratio for nintedanib were used (scenario 24) to [redacted] per QALY when the exacerbation coefficient was included (scenario 21).

5.2.3 Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) with input parameter distributions as presented in CS Table 117. The results from 1,000 iterations are reported in CS Table 120, whilst CS Figures 38 and 39 depict the scatterplot and cost-effectiveness acceptability curve (CEAC), respectively. The company assigns a multivariate normal distribution to overall survival and loss of lung function baseline transition, a beta distribution to exacerbation and discontinuation baseline transitions, adverse event risks, health state utilities and disutilities, and resource use proportions, and a lognormal distribution to costs and resource use. The EAG confirms that the probabilistic results are similar to the deterministic results.

5.2.4 Company base case results for FVC >80% predicted subgroup

In reply to clarification question B5, the company provided results for the FVC >80% predicted subgroup. The analysis used the log-logistic model for OS using the parameter values in Tables 10 and 11 of the clarification response document.

Table 24 Company results for nintedanib vs. best supportive care with OS for FVC >80% predicted subgroup (using PAS price for nintedanib)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£18,724	3.87	3.06				
Nintedanib	[redacted]	8.50	6.51	[redacted]	4.63	3.44	[redacted]

BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.

The FVC >80% predicted subgroup results have an ICER (with PAS) of [redacted] per QALY (Table 24Table 24). The company provides scenario analyses using the Weibull and

lognormal model. The ICERs using the Weibull and lognormal models were [redacted] and [redacted] per QALY respectively.

The company were asked to consider scenario analyses using baseline transition probabilities for acute exacerbation, treatment discontinuation, and loss of lung function outcomes for the FVC >80% predicted subgroup. The company did not run any scenarios based on acute exacerbations in the subgroup with FVC >80% predicted, because acute exacerbations were a rare event in the trials, especially in the FVC >80% predicted population (seven (2.4%) patients in nintedanib group and eight (4.2%) patients in the best supportive care group).

However, the company conducted a combined scenario concerning the probabilities of treatment discontinuation and loss of lung function for this subgroup. As in their base case analysis they assumed a constant risk of discontinuation. The company estimates the coefficient for the risk of discontinuation in the best supportive care group to be 7.777, corresponding to a discontinuation rate of 3.75% per 3-month cycle. The probability of discontinuation for patients taking nintedanib was informed by the odds ratio (OR) from the NMA (1.42), which was applied to the baseline best supportive care risk.

For loss of lung function, the company uses the overall trial OR for nintedanib vs placebo (OR = 0.54) and an OR of 0.50 which was derived from the subgroup with FVC >80% predicted (combined scenario 2 in Table 25). Both combined scenarios produce an ICER below [redacted] per QALY.

Table 25 Combined scenario analyses for nintedanib vs best supportive care: treatment discontinuation and loss of lung function derived from the FVC >80% predicted subgroup

Combined scenarios	Coefficient for discontinuation hazard rate	ICER (with PAS)
1) Treatment discontinuation and loss of lung function with base-case OR=0.54 for loss of lung function applied to nintedanib	7.777	[redacted]
2) Treatment discontinuation and loss of lung function with scenario 24 OR=0.50 for loss of lung function applied to nintedanib	7.777	[redacted]
Reproduced from Table 16 in the company clarification response document		

5.3 Model validation and face validity check

5.3.1 Company's model validation

The company states their approach to model validation in CS Section B.3.14. They report that the model structure, approaches, inputs and assumptions were validated as follows:

- Clinical expert advisory board[36] (April 2014), which included two clinical experts to validate the assumptions within the model and the model structure, to ensure that the model adheres to the clinical course of the disease and reflects current clinical practice.
- Validation by model developers: a senior modeller within the model developer's organisation (with no involvement in the development of the model for nintedanib) performed a detailed QA check on the model.
- Validation by the company: 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.

5.3.2 EAG model validation

The EAG conducted a series of quality checks on the company model, assessing its transparency and validity. A range of tests were performed to verify model inputs, calculations, and outputs:

- Cross-checking all parameter inputs against values reported in the CS, model, and cite sources.
- Checking all model outputs against results stated in the CS, including the base case, PSA, DSA, and company's scenarios.
- Checking the individual formulae within the model.
- Manually running scenarios and checking model outputs against results reported in the CS for the DSA and scenario analyses.
- Applying a range of extreme value and logic tests to check the plausibility of changes in results when parameters are changed ('black box' checks).
- Checking Visual Basic (VBA) code for errors and re-running the code to ensure expected outputs were produced.

No errors were identified in the model.

5.3.3 External validation

The company compares their fitted survival curves with data from external clinical studies and international IPF registries. IN the CS, the extrapolated survival curves for nintedanib and best supportive care, using log-logistic, Weibull, and generalised gamma distributions,

were compared with the Australian IPF registry data,[25] and the Greek IPF registry[37,38] and the EMPIRE registry.[38]. The cost-effectiveness model also enables the comparison against the Finnish and the European IPF registries. The company has compared their model results to the European IPF registry[39] and the Finnish IPF registry[40] in the model. Table 26 shows the patient characteristics for these sources.

Table 26 Characteristics of patients in the IPF clinical trials and registries

Data source	Mean age (years)	FVC % pred	Male	Smoking history[a]
INPULSIS I and II trials[41]	66.8	79.51%	79.3%	72.2%
European IPF registry[39]	68.1	68.40%	73.3%	64.7%
EMPIRE registry[38]	67.3	77.08%	68.0%	ND
Long-term NDB IPF data[29]	66.8	79%	78.0%	67.5%
Australian registry[25]	70.9	81.00%	67.7%	71.1%
Greek IPF registry[37]	71.8	73.30%	79.1%	78.2%
Finnish IPF registry[40]	73.0	80.20%	65.1%	55.0%
Kondoh et al.[24]	64.1	77.0%	61.0%	54.0%

Reproduced from the company's model and adapted by the EAG
[a] Ex-smokers and current smokers
ND, No data; Pred, predicted

The company opts to use the Australian registry as the primary source of validation based on NICE TA747,[22] where clinical experts and the NICE committee considered the registry to be a close representation of UK clinical practice. The company further notes that the baseline characteristics of the Australian IPF registry are comparable to those of patients in the TOMORROW and INPULSIS clinical trials, which are reported in CS Table 54. CS Figure 21 and Figure 22 depict the three parametric survival models versus Australian IPF registry data for nintedanib and best supportive care, respectively. For the first three years, all the three best fitting parametric curves (log-logistic, Weibull and generalized gamma) for nintedanib closely match the Australian IPF registry survival data. After year three, the closest fit is provided by the log-logistic curve. The pattern is similar for best supportive care, except the parametric curves start to deviate from the registry data after two years.

A Greek IPF registry,[37] reporting 5-year survival for patients on nintedanib, was compared with extrapolated survival for nintedanib. The company's models consistently predict higher overall survival than that seen in the registry data, as shown in CS Figure 23. The mean age of patients in the Greek IPF registry is 71.8 years, which is higher than that of patients in the TOMORROW and INPULSIS trials (66.5 years). Furthermore, the Greek registry comprised

more patients who were current or former smokers (78.2%) in comparison with clinical trials (72.6%).

The company also compares the extrapolated survival for nintedanib against the EMPIRE study,[38] a long-term real world study reporting 10-year survival rates. For the first two years the model predictions match the registry data, after which the survival rates with nintedanib in the model are higher than the Kaplan-Meier data, as can be seen in CS Figure 24. The extrapolated survival for best supportive care was also assessed against the EMPIRE study data for best supportive care; the modelled survival rates were higher than in EMPIRE (CS Figure 25). Although the mean age of patients in the EMPIRE study is the same as in the clinical trials (66.5 years), this is taken at the point of diagnosis rather than the start of treatment.

The company's OS extrapolation for best supportive care using the Weibull and log-logistic models are also compared to a retrospective study of 110 patients with IPF in Japan by Kondoh et al.[24] (CS Figure 15 and 16). The KM data from Kondoh et al are presented for patients with / without an acute exacerbation.

5.3.4 EAG corrections to the company model

The company model does not include general population mortality. Including general population mortality, where it is higher than IPF mortality (when patients are about 85 years old), increases the ICER to [redacted] per QALY (Table 27).

Table 27 Corrected company base case results using general population mortality for lifetime horizon (PAS price)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,247	4.05	3.18				
Nintedanib	[redacted]	7.00	5.44	[redacted]	2.96	2.26	[redacted]

BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.

The company uses a lifetime (50-year) horizon in their model results. The EAG considers this too long as the starting age of the patient population is approximately 67 years old and so includes patients until age 117 years. We believe a 35-year time horizon is more appropriate. This change does not affect the ICER (Table 28).

Table 28 Scenario analysis using general population mortality for time horizon of 35 years (PAS price)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,246	4.05	3.18				
Nintedanib	[redacted]	7.00	5.44	[redacted]	2.95	2.25	[redacted]
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.							

5.3.4.1 Stopping rule for nintedanib

A stopping rule was recommended for nintedanib in TA379,[1] whereby patients experiencing an absolute decline of 10% or more in predicted FVC within any 12-month period discontinue treatment. This rule is not modelled in the current CS as clinicians consider the stopping rule difficult to impose. However, we have included the stopping rule in an EAG scenario analysis (Table 29; section 6.2.2). Using the stopping rule decreases the ICER to [redacted] per QALY.

Table 29 Scenario analysis using the EAG corrections model with nintedanib treatment discontinuation for patients experiencing a decline of ≥FVC 10% predicted

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£19,246	4.05	3.18				
Nintedanib	[redacted]	6.99	5.41	[redacted]	2.94	2.23	[redacted]
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.							

5.3.4.2 Analysis using OS parameters for the FVC >80% predicted subgroup

The company's model uses clinical effectiveness data for the whole trial population, rather than being restricted to the FVC >80% predicted subgroup. The company provides OS parameter values for this population as part of their clarification response (question B5). Table 30 shows the cumulative effect of including general population mortality, and using OS for the FVC >80% predicted subgroup with a time horizon of 35 years. The ICER for this analysis is [redacted] per QALY.

Table 30 Subgroup analysis with OS from for FVC >80% predicted subgroup using general population mortality with time horizon of 35 years (PAS price)

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£18,712	3.85	3.05				
Nintedanib	[redacted]	7.95	6.15	[redacted]	4.09	3.10	[redacted]
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.							

5.3.5 EAG summary of key issues and additional analyses

A full summary of EAG observations on key aspects of the company's economic model is presented in Table 31.

Table 31 EAG observations of the key aspects of the company's economic model

Parameter	Company base case	EAG comment	EAG base case
Population	Uses FVC >80% predicted subgroup.	We agree	No change
Lung disease progression	CS Table 66;	Similar pattern of decline in lung function observed in patients with baseline FVC >80% predicted to whole trial population (CS Figure 30).	No change
Overall survival (OS)	Uses mortality for whole trial population.	Mortality for FVC >80% predicted population should be used.	No difference in mortality for placebo vs nintedanib for FVC >80% predicted population. Mortality from whole trial population used after 5.5 years.
Risk of mortality after exacerbation	HR of 1.4 for those with exacerbation, based on a study by Kondoh et al.[24] who reported HR of 2.79. The company divided this by two, to account for the cycle length.	It is inappropriate to divide the HR by two, as it is independent of time.	HR of 2.79.
Acute exacerbation	Uses OR for acute exacerbation for	Results are contradictory for FVC	No change but tested in scenario analyses.

	whole trial population.	>70% predicted and FVC >80% predicted analyses.	
Acute exacerbation rate	Uses acute exacerbation rate for whole trial population	Use acute exacerbation rate for FVC >80% predicted population in scenario analysis (1.05% per 3 month cycle for FVC >80% predicted and 2.58% for FVC ≤80% predicted)	No change but tested in scenario analyses.
Treatment discontinuation	Uses discontinuation rate for whole trial population.	We agree	No change but tested in scenario analyses.
Time horizon	50 years	Patient age is 117 years at end of time horizon.	35 years.
Utilities			
Health state utilities	CS Table 88	We agree. Uses values from TA379.	No change
AE disutility	CS Table 88	We agree.	No change
Resource use and costs			
Unit costs	CS Table 112	We agree. Uses updated values from TA379.	No change
Resource use	CS Table 112	We agree. Uses values from TA379.	No change

6 EAG'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

6.1.1 Deterministic sensitivity analyses

[redacted] below shows the results of deterministic sensitivity analyses for the FVC >80% predicted subgroup using the EAG's corrected model and applying the PAS discount for nintedanib. We explored the same 14 scenarios as provided in the CS (CS Table 121) with one-way sensitivity analysis across four parameters: probabilities (of mortality, exacerbation, progression and discontinuation), costs, utilities and adverse events. Input parameter modifications were based on 95% confidence intervals.

The ICERs from the one-way sensitivity analysis range from [redacted] to [redacted] per QALY. The cost-effectiveness of nintedanib is most influenced by mortality probabilities due to exacerbation (scenario 1) and discontinuation probabilities (scenario 4), increasing the ICER to [redacted] and [redacted] per QALY, respectively.

[redacted]

Figure 4 Tornado diagram of nintedanib vs best supportive care (FVC >80% predicted subgroup)

As noted in section 5.2.2, the CS scenario analyses describe 19 separate assumption-testing scenarios in the model (CS Tables 122-128, numbered 15-33), covering seven parameters: overall survival, exacerbations, loss of lung function, adverse events, costs, discontinuation, and FVC% predicted categories. The company uses clinical effectiveness data from the whole trial population in their base case analyses. Table 32 shows the cost-effectiveness results for the 19 scenarios, using the EAG's corrected model and the PAS price for nintedanib for the FVC >80% predicted subgroup.

Table 32 Scenario analyses results using EAG corrected model for FVC >80% subgroup (using PAS price for nintedanib)

Scenario	Parameter	Description of parameter varied	ICER (per QALY)
EAG corrected model			[redacted]
15	Overall survival	Parametric distribution: Weibull model (NDB and BSC)	[redacted]
16		Parametric distribution: Generalised gamma model (NDB and BSC)	[redacted]

17		Baseline risk: allow progression from FVC40-49.9% pred to FVC30-39.9% pred (death)	[redacted]
18	Exacerbation	Baseline risk: use investigator estimates	[redacted]
19		Baseline risk: exclude recurrent exacerbation risk	[redacted]
20		Relative risk: NMA results, scenario 4 excluding Richeldi 2011 (OR=0.62)	[redacted]
21	Loss of lung function	Baseline risk: include exacerbation coefficient	[redacted]
22		Relative risk: NMA results, scenario 3 excluding Richeldi 2011 (OR=0.53)	[redacted]
23		Transition probabilities for FVC >80% predicted	[redacted]
24		Transition probabilities for FVC >80% predicted and OR for NDB patients with FVC >80% predicted (OR=0.50)	[redacted]
25	Safety	Relative risk: serious cardiac events, NMA results, scenario 2 excluding Richeldi 2011 (OR=0.92)	[redacted]
26		Relative risk: serious GI events, NMA results, scenario 2 excluding Richeldi 2011 (OR=1.88)	[redacted]
27		Serious AE disutility value: use alternative value for serious cardiac events (-0.00825)	[redacted]
28		Serious AE disutility value: use alternative value for GI perforation (-0.0021)	[redacted]
29		Serious AE disutility value: use extreme value for all serious AEs: maximum disutility – serious cardiac events value	[redacted]
30	Costs	Cost of right heart catheterisation. Cost for respiratory physiology used (£96.68)	[redacted]
31	Discontinuation	Relative risk: NMA results, scenario 3 excluding Richeldi 2011 (OR=1.39)	[redacted]
32	FVC% predicted values	Use the lowest value of each FVC% pred category (e.g. 80 for the 80-89.9 FVC% pred category) as starting point	[redacted]
33		Use the highest value of each FVC% pred category (e.g. 89.9 for the 80-89.9 FVC% pred category) as starting point	[redacted]

BSC: best supportive care; NDB: nintedanib; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year; NMA: network meta-analyses; GI: gastrointestinal; AE: adverse event; OR: odds ratio.

Using the generalised gamma distribution for both nintedanib and best supportive care (scenario 16), rather than the log-logistic model, had the most significant effect on the cost-effectiveness results, reducing the ICER to [redacted] per QALY. The remaining scenarios caused no significant changes to the ICER, which ranges from [redacted] (scenario 24) to [redacted] (scenario 21) per QALY.

6.1.2 Probabilistic analyses

The company conducted a probabilistic sensitivity analysis (PSA) to explore uncertainty in the model. The parameters are described in CS Table 117 and the results from 1,000 iterations are reported in CS Table 18. We repeated this PSA using the EAG corrected model and restricted the analysis to data from the FVC >80% predicted subgroup. The EAG

confirms that the deterministic and probabilistic results for nintedanib versus best supportive care are comparable (Table 33).

Table 33 Deterministic results vs probabilistic results using EAG corrected model for the FVC >80% predicted subgroup (using PAS price for nintedanib)

Intervention/comparator	Total costs	LYs	QALYs	ICER (per QALY)
Deterministic analysis				
Nintedanib	[redacted]	7.95	6.15	[redacted]
BSC	[redacted]	3.85	3.05	-
Probabilistic analysis				
Nintedanib	[redacted]	7.95	6.14	[redacted]
BSC	[redacted]	3.92	3.10	-
LYs: life years; QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; BSC: best supportive care.				

6.2 EAG's preferred assumptions

Based on the EAG critique of the company's model, we have identified the following aspects of the company base case with which we disagree. Our preferred model assumptions are the following:

- **Population used for overall survival**: FVC >80% predicted, rather than whole trial population.
- **Extrapolation of OS:** For the first 5.5 years, we use the same survival curve for the best supportive care arm as for the nintedanib arm as the mortality rate for both arms is considered equal; thereafter we use the best supportive care survival curve from the whole trial population for the best supportive care arm.
- **OS Hazard ratio for acute exacerbations:** we implement a HR of 2.79, rather than 1.4.
- **Time horizon:** we opted for a time horizon of 35 years, rather than 50 years.

6.2.1 Results from the EAG preferred model assumptions

Table 34 below presents the results obtained from the model with the above preferred model assumptions implemented.

Table 34 EAG base case model results (using PAS price for nintedanib) for the FVC >80% predicted subgroup

Technology	Total			Incremental			
	Costs	LYG	QALY	Costs	LYG	QALY	ICER (£/QALY)
BSC	£23,264	5.71	4.49				
Nintedanib	[redacted]	7.20	5.62	[redacted]	1.49	1.14	[redacted]
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALY: quality-adjusted life year.							

Table 35 shows the cumulative cost-effectiveness results of applying the EAG preferred model assumptions to the corrected company's base case. Incorporating the EAG assumptions leads to an increase in the ICER from [redacted] to [redacted] per QALY. The change that has the most significant impact on the cost-effectiveness results is the OS extrapolation. The other suggested changes have a small impact on the ICER.

Table 35 Cumulative change from the EAG corrected model with the EAG preferred model assumptions (using PAS price for nintedanib)

Assumption	Treatment	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
EAG corrected model	BSC	£19,247	3.18			[redacted]
	NDB	[redacted]	5.44	[redacted]	2.26	
+ Time horizon of 35 years	BSC	£19,246	3.18			[redacted]
	NDB	[redacted]	5.44	[redacted]	2.25	
+ FVC >80% pred population for OS	BSC	£18,712	3.05			[redacted]
	NDB	[redacted]	6.15	[redacted]	3.10	
+ HR = 2.79 for OS for acute exacerbations	BSC	£18,252	2.90			[redacted]
	NDB	[redacted]	5.62	[redacted]	2.72	
+ Equal OS for both arms for 5.5 years (EAG base case)	BSC	£23,264	4.49			[redacted]
	NDB	[redacted]	5.62	[redacted]	1.14	
BSC: best supportive care; NDB: nintedanib; QALY: quality-adjusted life year; ICER: incremental cost-effectiveness ratio; OS: overall survival; HR: hazard ratio.						

6.2.2 Scenario analyses conducted on the EAG base case model

We performed scenario analyses using the EAG base case model to analyse the impact of changing some model assumptions on the overall cost-effectiveness results. In addition to replicating some of the company's scenarios, we also conducted further scenarios regarding acute exacerbation rates as follows:

- Acute exacerbation rate per 3-month cycle of 1.05% for FVC >80% predicted and 2.58% for FVC ≤80% predicted (rather than 1.47% - adjudicator-committee reported).
- Acute exacerbation (time to first acute exacerbation) HRs: for subgroups split by FVC 90% predicted, FVC 80% predicted, and FVC 70% predicted, as shown in Table 36.

Table 36 Scenario analysis: hazard ratios for time to first acute exacerbation for varying subgroups of patients

Outcome	**Baseline FVC >90% predicted**	**Baseline FVC ≤90% predicted**
Time to first acute exacerbation	0.46 (95% CI 0.09–2.48) in favour of nintedanib	0.66 (95% CI 0.39–1.11) in favour of nintedanib
	Baseline FVC >80% predicted	**Baseline FVC ≤80% predicted**
	HR: 0.49; 95% CI 0.17, 1.35	HR; 0.72; 95% CI 0.41, 1.27
	Baseline FVC >70% predicted	**Baseline FVC ≤70% predicted**
	HR: 1.00; 95% CI 0.44, 2.30	HR; 0.52; 95% CI 0.28, 0.99
Source: CS Figure 49 and CS text p.63, 65		

Table 37 presents the results from the scenarios conducted on the EAG base case model. Using the Weibull and lognormal distributions in the model results in the highest ICERs, [redacted] and [redacted] per QALY, respectively. Using the hazard ratio for the time to first acute exacerbation in the FVC 70% predicted subgroup also increases the ICER to just over the willingness-to-pay threshold of [redacted] per QALY.

Assuming overall survival was the same in both treatment groups for one year reduced the ICER to [redacted] per QALY, and to [redacted] per QALY if overall survival was assumed to be the same for three years. Including the stopping rule, whereby patients who experience a decline of ≥FVC 10% predicted within a year discontinue and the treatment effect is lost, reduced the ICER to [redacted] per QALY. Using a generalised gamma distribution in the model also notably affected the ICER, reducing it to [redacted] per QALY. The remaining scenarios did not change the ICER significantly, which ranged from [redacted] to [redacted] per QALY.

Table 37 Scenario analyses results using the EAG base case model (using PAS price for nintedanib) for the FVC >80% predicted subgroup

Scenario	Treatment	Total costs	Total QALYs	ICER (£/QALY)
EAG base case	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.62	
From Company Submission				
OS: Parametric distribution - Weibull (NDB and BSC) (CS scenario 15)	BSC	£22,161	4.20	[redacted]
	NDB	[redacted]	5.07	
OS: Parametric distribution – Generalised Gamma (NDB and BSC) (scenario 16)	BSC	£21,642	4.06	[redacted]
	NDB	[redacted]	5.64	
OS: Allow progression from FVC 40-49.9% pred to FVC 30-39.9% pred (scenario 17)	BSC	£23,111	4.46	[redacted]
	NDB	[redacted]	5.54	
Loss of lung function: Transition probabilities for FVC >80% pred (scenario 23)	BSC	£22,737	4.50	[redacted]
	NDB	[redacted]	5.65	
Loss of lung function: Transition probabilities for FVC >80% pred and OR for NDB in patients with FVC >80% pred (OR=0.5) (scenario 24)	BSC	£22,737	4.50	[redacted]
	NDB	[redacted]	5.65	
EAG scenarios				
OS: Parametric distribution - Lognormal (NDB and BSC)	BSC	£25,833	5.09	[redacted]
	NDB	[redacted]	6.03	
Acute exacerbation rate: 1.05% for FVC >80% pred and 2.58% for FVC ≤80% pred	BSC	£22,650	4.50	[redacted]
	NDB	[redacted]	5.61	
Acute exacerbation HR for FVC >90% pred and FVC ≤90% pred	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.63	
Acute exacerbation HR for FVC >80% pred and FVC ≤80% pred	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.64	
Acute exacerbation HR for FVC >70% pred and FVC ≤70% pred	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.47	
Equal OS for both arms for 1 year	BSC	£19,590	3.34	[redacted]
	NDB	[redacted]	5.62	
Equal OS for both arms for 3 years	BSC	£21,557	3.99	[redacted]
	NDB	[redacted]	5.62	
20-year time horizon	BSC	£23,099	4.47	[redacted]
	NDB	[redacted]	5.49	

50-year time horizon (lifetime)	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.62	
NDB: Discontinue treatment and lose treatment effect for patients that experience a decline of ≥FVC 10% predicted	BSC	£23,264	4.49	[redacted]
	NDB	[redacted]	5.57	
BSC: best supportive care; NDB: nintedanib; QALY: quality-adjusted life year; ICER: incremental cost-effectiveness ratio; OS: overall survival; HR: hazard ratio; OR: odds ratio.				

6.3 Conclusions on the cost effectiveness evidence

The company's model generated a base case ICER of [redacted] per QALY for nintedanib vs best supportive care (using the PAS price for nintedanib). The model used clinical effectiveness estimates from the whole trial populations. In response to clarification question B5, the company produced a scenario analysis using the OS for the FVC >80% predicted subgroup, which had an ICER of [redacted] per QALY.

Our preferred model assumptions are the following:

- **Population used for the overall survival**: FVC >80% predicted, rather than the whole trial population.
- **Extrapolation of OS:** For the first 5.5 years, we use the same survival curve for the best supportive care arm as for the nintedanib arm as the mortality rate for both arms is considered equal; thereafter we use the best supportive care survival curve from the whole trial population for the best supportive care arm.
- **OS Hazard ratio for OS for acute exacerbations:** we use a HR of 2.79, rather than 1.4.
- **Time horizon:** we use a time horizon of 35 years, rather than 50 years.

The EAG's corrections and preferred assumptions increase the ICER for nintedanib vs best supportive care to [redacted] per QALY. These estimates are most sensitive to changes in the assumptions related to the OS extrapolation.

7 SEVERITY

The company calculates the QALY shortfall using the SCHARR QALY shortfall calculator,[42] and:

- General population QALYs calculated from EQ-5D health state profiles[43]
- HRQoL, measured using the EQ-5D-5L and mapped to the EQ-5D-3L[44]

- National life table data for age and sex-specific survival times[45]

The sex distribution (78% male) and starting age (68 years) were based on the baseline characteristics of people with FVC >80% predicted (CS Section B.2.3, Table 8). The company does not consider nintedanib suitable for a QALY weighting, because the absolute QALY shortfall compared with best supportive care in IPF is lower than 12 years; and the proportional shortfall is less than 85%.

EAG comment on severity

The EAG checked the company's calculations and we agree with the company's evaluation. We do not believe that there is a high degree of severity, as the absolute QALY shortfall is less than 12 years and the proportional shortfall is below 85%.

8 References

1. NICE. *Nintedanib for treating idiopathic pulmonary fibrosis [TA379]*. https://www.nice.org.uk/guidance/ta379 (accessed 27/01/2022).
2. NICE. *Review of TA379; Nintedanib for treating idiopathic pulmonary fibrosis and TA504; Pirfenidone for treating idiopathic pulmonary fibrosis*. https://www.nice.org.uk/guidance/ta504/evidence/review-decision-paper-pdf-9135716510 (accessed 26/01/2022).
3. EMC. *Esbriet 267 mg Hard Capsules SPC*. https://www.medicines.org.uk/emc/product/3705/smpc#gref (accessed 21/01/2022).
4. EMC. *Ofev 150 mg soft capsules SmPC*. https://www.medicines.org.uk/emc/product/7705/smpc#gref (accessed 13/05/2022).
5. NICE. *Pirfenidone for treating idiopathic pulmonary fibrosis [TA504]*. https://www.nice.org.uk/guidance/ta504 (accessed 24/01/2022).
6. Spencer LG, Loughenbury M, Chaudhuri N, et al. Idiopathic pulmonary fibrosis in the UK: analysis of the British Thoracic Society electronic registry between 2013 and 2019. *ERJ Open Res* 2021;7(1).
7. Cooper K, Kalita N, Rose M, et al. Nintedanib for treating idiopathic pulmonary fibrosis: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC). 2015.
8. Maher TM, Stowasser S, Nishioka Y, et al. Biomarkers of extracellular matrix turnover in patients with idiopathic pulmonary fibrosis given nintedanib (INMARK study): a randomised, placebo-controlled study. *Lancet Respir Med* 2019;7(9):771-79.

9. Lancaster L, Goldin J, Trampisch M, et al. Effects of Nintedanib on Quantitative Lung Fibrosis Score in Idiopathic Pulmonary Fibrosis. *The open respiratory medicine journal* 2020;14:22-31.
10. Lancaster L, Crestani B, Hernandez P, et al. Safety and survival data in patients with idiopathic pulmonary fibrosis treated with nintedanib: pooled data from six clinical trials. *BMJ Open Respir Res* 2019;6(1):e000397.
11. ClinicalTrials.gov. *Safety and PK Study of BIBF 1120 in Japanese Patients With IPF*. https://clinicaltrials.gov/ct2/show/NCT01136174 (accessed 25th August 2022).
12. UMIN. *The comparison of the efficacy and safety of pirfenidone and nintedanib in patients with idiopathic pulmonary fibrosis*. https://center6.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000022819 (accessed 25th August 2022).
13. Maher TM, Flaherty KR, Noble PW, et al. Data on file: Effect of baseline FVC on lung function decline with nintedanib in patients with IPF. 25th Ann Cong of the European Respiratory Society (ERS), Amsterdam, 26 - 30 Sep 2015 (Oral Presentation), 2015.
14. British Thoracic Society. *BTS ILD Registry Annual Report 2021*. https://www.brit-thoracic.org.uk/document-library/quality-improvement/ild-registry/bts-ild-registry-annual-report-2021/ (accessed 17/02/2022).
15. NICE. *Single technology appraisal and highly specialised technologies evaluation: User guide for company evidence submission template* https://www.nice.org.uk/process/pmg24/chapter/clinical-effectiveness (accessed 25th August 2022).
16. Bowers M, Pickering R, Weatherall M. Design, objectives, execution and reporting of published open-label extension studies. *Journal of Evaluation in Clinical Practice* 2012;18(2):209-15.
17. Maher TM, Flaherty KR, Noble PW, et al. Effect of baseline FVC on lung function decline with nintedanib in patients with IPF: Eur Respiratory Soc, 2015.
18. Rinciog C, Watkins M, Chang S, et al. A Cost-Effectiveness Analysis of Nintedanib in Idiopathic Pulmonary Fibrosis in the UK. *Pharmacoeconomics* 2017;35(4):479-91.
19. Loveman E, Copley VR, Colquitt JL, et al. The effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: systematic review, network meta-analysis and health economic evaluation. *BMC Pharmacol Toxicol* 2014;15:63.
20. Richeldi L, Costabel U, Selman M, et al. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. *New England Journal of Medicine* 2011;365(12):1079-87.
21. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *New England Journal of Medicine* 2014;370(22):2071-82.

22. NICE. *Nintedanib for treating progressive fibrosing interstitial lung diseases [TA747]*. https://www.nice.org.uk/guidance/ta747 (accessed 18/01/2022).
23. NICE. *NICE health technology evaluations: the manual*. https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation (accessed 01 July 2022).
24. Kondoh Y, Taniguchi H Fau - Katsuta T, Katsuta T Fau - Kataoka K, et al. Risk factors of acute exacerbation of idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2010(1124-0490 (Print)).
25. Jo HE, Glaspole I, Grainge C, et al. Baseline characteristics of idiopathic pulmonary fibrosis: analysis from the Australian Idiopathic Pulmonary Fibrosis Registry. *Eur Respir J* 2017;49(2).
26. Kakugawa T, Sakamoto N, Sato S, et al. Risk factors for an acute exacerbation of idiopathic pulmonary fibrosis. *Respir Res* 2016;17(1):79.
27. Porte F, Cottin V, Catella L, et al. Health economic evaluation in idiopathic pulmonary fibrosis in France. *Curr Med Res Opin* 2018;34(10):1731-40.
28. Rinciog C, Diamantopoulos A, Gentilini A, et al. Cost-Effectiveness Analysis of Nintedanib Versus Pirfenidone in Idiopathic Pulmonary Fibrosis in Belgium. *Pharmacoecon Open* 2020;4(3):449-58.
29. Boehringer Ingelheim. INPULSIS patient data, 2014.
30. Ara R, Brazier JE. Using Health State Utility Values from the General Population to Approximate Baselines in Decision Analytic Models when Condition-Specific Data are Not Available. *Value in Health* 2011;14(4):539-45.
31. Loveman E, Copley VR, Colquitt J, et al. The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation. *Health Technol Assess* 2015;19(20):i-xxiv, 1-336.
32. Diamantopoulos A, Maher TM, Schoof N, et al. Influence of Idiopathic Pulmonary Fibrosis Progression on Healthcare Resource Use. *Pharmacoecon Open* 2019;3(1):81-91.
33. NHS England. *National Schedule of NHS costs Version 2 - Year 2019-20*; 2021.
34. Jones K, Burns A. *Unit Costs of Health and Social Care 2021*; 2021.
35. NICE. *Idiopathic pulmonary fibrosis in adults: diagnosis and management [CG163]*. https://www.nice.org.uk/guidance/CG163/chapter/1-Recommendations#management (accessed 24/01/2022).
36. Boehringer Ingelheim. Advisory board minutes, 2014.
37. Antoniou K, Markopoulou K, Tzouvelekis A, et al. Efficacy and safety of nintedanib in a Greek multicentre idiopathic pulmonary fibrosis registry: a retrospective, observational, cohort study. *ERJ Open Res* 2020;6(1).

38. Vasakova M, Sterclova M, Mogulkoc N, et al. Long-term overall survival and progression-free survival in idiopathic pulmonary fibrosis treated by pirfenidone or nintedanib or their switch. Real world data from the EMPIRE registry: Eur Respiratory Soc; 2019.
39. Guenther A, Krauss E, Tello S, et al. The European IPF registry (eurIPFreg): baseline characteristics and survival of patients with idiopathic pulmonary fibrosis. *Respiratory Research* 2018;19(1):141.
40. Kaunisto J, Salomaa ER, Hodgson U, et al. Demographics and survival of patients with idiopathic pulmonary fibrosis in the FinnishIPF registry. *ERJ Open Res* 2019;5(3).
41. Richeldi L, Du Bois R, Raghu G, et al. Efficacy and safety of nintedanib in patients with idiopathic pulmonary fibrosis: Results of two 52-week, phase III, randomized, placebo-controlled trials (INPULSISTM). *American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS* 2014;189.
42. Schneider P, McNamara S, Love-Koh J, et al. *QALY Shortfall Calculator*. https://r4scharr.shinyapps.io/shortfall/.
43. UK Data Service. *National Centre for Social Research (NatCen), University College London, Department of Epidemiology and Public Health*.
44. Hernandez Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study. *Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. Report 063*, 2020.
45. Office for National Statistics. *National life tables: UK*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables (accessed 01/04/2022).

9 Appendices

9.1 Appendix 1 EAG appraisal of systematic review methods

Table 38 EAG appraisal of systematic review methods

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
Was the review question clearly defined using the PICOD framework or an alternative?	Yes	The PICOD components are detailed in CS Appendix D Table 135.
Were appropriate sources of literature searched?	Yes	Included Medline, Embase, Cochrane Central Register of Controlled Trials, relevant clinical trial registries and conference abstracts, reference lists of key papers and systematic reviews. The CDSR was not searched however the EAG notes

		this would have yielded 11 additional references of which none were relevant.
What time period did the searches span and was this appropriate?	Yes	September 2014 to January 2022. The EAG performed an updated search. No new RCTs were identified.
Were appropriate search terms used and combined correctly?	Yes	CS Appendix D 1.1 Tables 131-133
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes	The company do specify their selection criteria (CS Appendix D Table 135). We note that these are wider than the decision problem as they are based on those used for TA379 which has a wider scope in terms of population and comparators. However, it appears the company have applied additional ad-hoc exclusions to studies identified at full text review.
Were study selection criteria applied by two or more reviewers independently?	Yes	CS Appendix D.1.1
Was data extraction performed by two or more reviewers independently?	No.	One reviewer extracted the study data and a second reviewer validated the extracted data (CS Appendix D.1.1.) The EAG considers this approach adequate.
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes	The company assessed the risk of bias using the following tools/guides: • CRD criteria recommended by NICE for the nintedanib RCTs (CS Table 15) • STA User Guide (2022) and a publication by Bowers et al. for the open label extension studies (CS Tables 16 and 17)[15,16] • Cochrane Risk of Bias tool for other trials in the company's NMA.[1]
Was risk of bias assessment (or other study quality assessment) conducted by two or more reviewers independently?	Unclear	The CS does not state who performed the risk of bias assessments.

Is sufficient detail on the individual studies presented?	Yes	Further details of the trial characteristics are presented in CS sections B.2.2 and B.2.3.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes	Pairwise and network meta-analysis were used to combine study results from the INPULSIS and TOMORROW RCTs. A full critique is provided in section 3.5 of this EAG report.

CDSR: Cochrane Database of systematic reviews; CRD: Centre for Reviews and Dissemination; STA: Single technology appraisal

9.2 Appendix 2 Comparison of company and EAG critical appraisal of open label extension studies

Table 39 Comparison of company and EAG quality assessment (STA User Guide criteria) for the INPULSIS-ON open-label extension study

Trial name	Company assessment	EAG assessment
Was the cohort recruited in an acceptable way?	Yes, patients who completed INPULSIS trials were eligible.	Yes, but with the caveat that only participants who completed INPULSIS RCT could enter the OLE.
Was the exposure accurately measured to minimise bias?	Yes, median (range) exposure for patients continuing and initiating treatment recorded.	Yes, exposure to actual treatment received was recorded during the RCT and the OLE.
Was the outcome accurately measured to minimise bias?	Yes.	Yes
Have the authors identified all important confounding factors?	Yes, decreasing patient numbers over time, potential for selection bias in patients who continued the extension.	Probably yes, with the caveat that there may be unknown confounding factors.
Have the authors taken account of the confounding factors in the design and/ or analysis?	Yes, subgroup analysis conducted by nintedanib dose, dose adjustment, and dose intensity.	Probably yes
Was the follow-up of patients complete?	Yes, data are based on the database lock for the final analysis.	Yes, but with the caveats that i) not all participants entered the open-label extension and ii) some participants dropped out of the OLE.
How precise (for example, in terms of confidence interval	Not applicable, due to small sample size.	Most results are presented with SE, SD or 95% CI. Sample size is larger than for the TOMORROW study which

and p values) are the results?		should mean results from this study are more precise than those of TOMORROW.
Source: CS Table 16 with addition of EAG quality assessment		

Table 40 Comparison of company and EAG quality assessment (STA User Guide criteria) for the TOMORROW open-label extension

Trial name	Company assessment	EAG assessment
Was the cohort recruited in an acceptable way?	Yes, patients who completed TOMORROW were eligible.	Yes, but with the caveat that only participants who completed both the RCT and subsequent blinded phase 2 section of TOMORROW could enter the OLE. CS Table 7 suggests four countries that contributed data to the TOMORROW RCT were not represented in the TOMORROW OLE.
Was the exposure accurately measured to minimise bias?	Yes, exposure by trial and treatment recorded.	Yes, exposure to actual treatment received was recorded across all the periods of study including the OLE.
Was the outcome accurately measured to minimise bias?	Yes.	Yes
Have the authors identified all important confounding factors?	Yes, decreasing patient numbers over time, switch in treatment and dose, potential for selection bias in patients who continued the extension.	Probably yes, with the caveat that there may be unknown confounding factors.
Have the authors taken account of the confounding factors in the design and/ or analysis?	Yes, analysis conducted separately for comparator arm which comprised patients who received placebo in period 1, nintedanib 50 mg once daily in period 2, and a range of nintedanib doses in the extension.	Yes, for changes in treatment and dose. The impact of missing patients (those who did not enter the OLE) on outcomes analysed is uncertain.
Was the follow-up of patients complete?	Yes, data are based on the database lock for the final analysis.	Yes, but with the caveats that i) not all participants entered the OLE and ii) not all patients completed the OLE
How precise (for example, in terms of confidence interval and p values) are the results?	Not applicable, due to small sample size.	Most results are presented with SE or 95% CI, but small sample sizes does mean the results are less certain than if the sample size had been larger.

Source: CS Table 16 with addition of EAG quality assessment

Table 41 Comparison of company and EAG quality assessment (Bowers et al. criteria) for the INPULSIS-ON open-label extension

Features indicative of high quality OLE studies (Bowers et al., 2012)[16]	Company assessment	EAG assessment
"Explicitly stated aims, to minimize the possibility of Type I Error"	☑ The objective was to assess the long-term efficacy and safety of nintedanib. The primary outcome was incidence of adverse events. The database was locked for final analysis on Sept 12, 2017 so all endpoints were recorded up to 192 weeks from baseline. Only descriptive statistics were used. No formal statistical inferences were used, but to aid the interpretation of the data, patients were divided into groups.	Yes. Aim stated as "to assess the long-term efficacy and safety of nintedanib" with the primary outcome to "characterise the long-term safety and tolerability of nintedanib in patients with idiopathic pulmonary fibrosis, and this was analysed in patients who received at least one dose of nintedanib in INPULSIS-ON"
"A well-characterized sample representative of the target population in whom the medication will be used"	☑ Patients who entered INPULSIS-ON were divided into two groups: those who had already received nintedanib (masked) in INPULSIS and continued nintedanib (open-label) in INPULSIS-ON, and those who had received placebo in INPULSIS and initiated nintedanib in INPULSIS-ON. Patients receiving nintedanib 150 mg twice daily or placebo at the end of an INPULSIS trial received nintedanib 150 mg twice daily in INPULSIS-ON. Patients receiving nintedanib 100 mg twice daily or	Partially. The sample is well characterised with baseline characteristics provided for the INPULSIS RCTs with a more limited range of characteristics reported for the participants who entered the INPULSIS-ON extension (CS Table 11).

Features indicative of high quality OLE studies (Bowers et al., 2012)[16]	Company assessment	EAG assessment
	placebo at the end of an INPULSIS trial could receive nintedanib 100 mg twice daily or 150 mg twice daily in INPULSIS-ON. Permanent or temporary dose reductions to 100 mg twice daily and treatment interruptions were allowed, to manage adverse events.	
"Outcome assessment is masked to treatment received where possible"	? Outcomes were assessed with clinical and laboratory evaluation and the recording of adverse events reported during and until 28 days after discontinuation of treatment. The published report does not state whether the outcomes assessors were blind to treatment allocation.	Unclear. As the OLE was not blinded and participants knew they were receiving nintedanib it is likely that outcome assessors were not blind to OLE treatment allocation, but they may have been blind to OLE participants' earlier RCT allocation.
"A low rate of sample slippage in relation to the numbers randomized in the preceding RCT, but the length of follow-up should be considered in making this assessment"	☑ The sample size decreased over time, but this is justified by the long 68-months follow-up duration (NOTE: long-term assessment per se' is an important objective in OLE studies; Bowers et al., 2012), and by the fact that this reduction was partly due to patients switching to prescribed nintedanib in clinical practice once it became available.[16] Sample size calculation was not required and the number of patients eligible depended on the number of patients completing the parent trials INPULSIS-1 and	The EAG considers that rate of sample slippage in relation to the numbers randomised in the preceding RCT is similar to what might be expected for studies of this type. After the 52-week RCT and 4-12 week treatment gap, 71.9% of those who had received placebo in the RCT and 67.4% of those who had received nintedanib entered the OLE (for the total RCT population 69.2% of participants entered the OLE). The proportion of RCT participants entering the OLE is not far below the mean of 74% (min-max 6-100%) calculated for a random sample of 40 OLEs.[16]

Features indicative of high quality OLE studies (Bowers et al., 2012)[16]	Company assessment	EAG assessment
	INPULSIS-2 and willing to participate in this extension trial. Of 807 patients who completed the INPULSIS trials, 734 were treated in INPULSIS-ON, of whom 430 were continuing nintedanib and 304 were initiating nintedanib. 295 of 430 patients continuing nintedanib and 219 of 304 patients who initiated nintedanib in INPULSIS-ON discontinued nintedanib during the trial. All analyses were evaluated using observed case analysis, i.e. using only the available data, without imputation for missing data. Missing or incomplete AE dates were imputed. Missing data for time-to-event endpoints were managed by censored data analyses.	
"The quality of a study can only be judged if objectives, design, conduct, analysis and results are adequately described and the STROBE guidelines for reporting observational studies in epidemiology should be followed"	☑ The published version of the report comply with STROBE guidelines.*	The published version of the INPULSIS-ON open-label extension does not explicitly identify potential confounders or effect modifiers and the statistical method of adjustment for the primary outcome is not described. In most other respects the published paper complies with STROBE guidelines.
"The limitations of the specific study design used and its execution should be discussed"	☑ The limitations are discussed in the published study and include: absence of a comparator group; decreasing patient numbers over time. There was also potential for selection bias due to patients in the	Yes, the published study includes a discussion of study limitations.

Features indicative of high quality OLE studies (Bowers et al., 2012)[16]	Company assessment	EAG assessment
	INPULSIS trials who had a more favourable course of disease or were better able to tolerate nintedanib. These patients would have been more likely to complete the trial and so be eligible for INPULSIS-ON. They might also have been more likely to remain on treatment in INPULSIS-ON, potentially reducing the observed decline in FVC and mortality in INPULSIS-ON.	
Source: CS Table 17 with addition of EAG quality assessment		

Table 42 Comparison of company and EAG study quality assessment (Bowers et al. criteria) for the TOMORROW open-label extension

Features indicative of high quality OLE studies Bowers et al., 2012)[16]	Company assessment	EAG assessment
"Explicitly stated aims, to minimize the possibility of Type I Error"	☑ The main objective was to present long-term efficacy and safety data. The primary efficacy endpoint was the annual rate of decline in FVC and was calculated using all FVC assessments from first drug administration in the extension study until database lock on 15th October 2015, up to 61.8 months.	Yes. The clinicaltrials.gov entry (where nintedanib is called BIBF 1120) for the TOMORROW OLE states "The aim of this trial is to offer continuation of BIBF 1120 treatment for patients with Idiopathic Pulmonary Fibrosis (IPF) who have completed a prior clinical trial with that drug. The primary objective will be to establish the long term tolerability and safety profile of BIBF 1120 in Idiopathic Pulmonary Fibrosis (IPF). As a secondary objective the

Features indicative of high quality OLE studies Bowers et al., 2012)[16]	Company assessment	EAG assessment
	All endpoints were exploratory and only descriptive statistics were performed.	effects of long-term treatment with BIBF 1120 on survival as well as safety and efficacy parameters will be investigated in an open-label, not randomized, un-controlled design".
"A well-characterized sample representative of the target population in whom the medication will be used"	☑ Patients who completed 52 week's treatment in TOMORROW period 1 continued treatment in a blinded phase (period 2), until the last patient had completed 52 weeks' treatment in period 1. In period 2, patients treated with nintedanib in period 1 continued their dose, and placebo-treated patients were switched to nintedanib 50 mg qd in a blinded manner. Patients who completed period 2 could continue/start nintedanib in the open-label extension trial. Patients entered the extension trial on the dose that they were receiving at the end of period 2, but had the option to increase dose to nintedanib 150 mg bid. Dose reduction from 150 mg bid to 100 mg bid and treatment interruption were permitted for the management of adverse events. In the extended period study, the comparator group received placebo in period 1 and nintedanib 50mg qd in period 2.	Partially. The sample is well characterised with baseline characteristics provided for the TOMORROW RCT and a more limited range of characteristics reported for those participants who entered the TOMORROW OLE (CS Table 13).

Features indicative of high quality OLE studies Bowers et al., 2012)[16]	Company assessment	EAG assessment
"Outcome assessment is masked to treatment received where possible"	? The published report does not state whether the outcomes assessors were blind to treatment allocation.	Unclear. As the OLE was not blinded and participants knew they were receiving nintedanib it is likely that outcome assessors were not blind to OLE treatment allocation, but they may have been blind to OLE participants' earlier RCT allocation.
"A low rate of sample slippage in relation to the numbers randomized in the preceding RCT, but the length of follow-up should be considered in making this assessment"	☑ The sample size decreased over time, but this is justified by the nearly 8-years follow-up duration from the start of period 1 (NOTE: long-term assessment per se' is an important objective in OLE studies; Bowers et al., 2012)[16] The number of patients eligible for the extension study depended on the number of patients completing the TOMORROW trial and willing to participate in this extension trial. Of 428 patients treated in period 1, a total of 286 entered period 2, and 198 entered the extension, including 35 in the nintedanib 150 mg twice daily group and 37 in the comparator group (35 of whom increased dose to nintedanib 150 mg twice daily). The full analysis set included all patients in the treated set who provided baseline data (for the first	The EAG considers that rate of sample slippage in relation to the numbers randomised in the preceding RCT is a potential concern. After the 52-week RCT and period 2 (length unclear, CS Figure 1 suggests a maximum of about 30 weeks) 46% of the RCT participants entered the OLE. A 2012 review of OLE studies found across a random sample of 40 OLEs a mean of 74% (min-max 6-100%) of the participants randomized in the preceding RCT(s) were enrolled in the OLE.[16] The rate of sample slippage in relation to the numbers randomized in the preceding RCT would therefore appear to be higher than average. This rate of sample slippage is not unexpected given the long duration of follow up, however, we are uncertain how this compares, on average, with that in studies of a similarly long duration.

Features indicative of high quality OLE studies Bowers et al., 2012)[16]	Company assessment	EAG assessment
	trial visit) for at least 1 endpoint in the open-label extension trial.	
"The quality of a study can only be judged if objectives, design, conduct, analysis and results are adequately described and the STROBE guidelines for reporting observational studies in epidemiology should be followed"	☑ The published version of the report comply with STROBE guidelines.	The published version of the TOMORROW trial extension lacks a clearly reported rationale for the study and does not state specific objectives. {Richeldi, 2018 #4} Confounder & effect modifier terminology are not used so the reader would need to identify potential confounders and effect modifiers themselves by interpreting/inferring from the text. In most other respects the published paper complies with STROBE guidelines.
"The limitations of the specific study design used and its execution should be discussed"	☑ The limitations are discussed in the published study and include: switches in treatments and doses; lack of a true placebo group; potential for selection bias in patients who continued into the extension. Patients who died or were unable to enter the extension due to disease progression were excluded from the analyses. The small patient numbers available for analyses beyond period 1 means these results may underestimate the rate of FVC decline, particularly in the comparator group, in which most patients received nintedanib 150 mg twice daily in the extension.	Yes, the published study includes a discussion of study limitations.
Source: CS Table 17 with addition of EAG quality assessment		

9.3 Appendix 3 Additional clinical effectiveness results

9.3.1 Post-hoc subgroup analyses from INPULSIS trials: FVC ≤90% vs. >90%

Evidence for the primary endpoint (adjusted annual rate of decline in FVC) was presented for TA379. In the current submission the company additionally provides the data as a figure (CS Figure 7). New for this submission are data presented for: time to first acute exacerbation, adjusted mean change from baseline in SGRQ total score and time to an absolute decline in FVC ≥10% predicted or death as shown in Table 43.

Table 43 Subgroup analyses by FVC% predicted ≤90% versus >90% (INPULSIS trials)

Outcome	baseline FVC >90% predicted			baseline FVC ≤90% predicted		
Time to first acute exacerbation	Hazard ratio: 0.46 (95% CI: 0.09, 2.48) in favour of nintedanib			Hazard ratio: 0.66 (95% CI: 0.39–1.11) in favour of nintedanib		
	Treatment-by- subgroup interaction p=0.956					
Adjusted mean change from baseline in SGRQ total score at week 52	Nintedanib n=NR	Placebo n=NR	difference	Nintedanib n=NR	Placebo n=NR	difference
	2.16	3.02	−0.87 (95% CI: −3.97, 2.24)	4.00	5.64	−1.65 (95% CI: −3.60, 0.31)
	Treatment-by-subgroup interaction p=0.3382					
Time to an absolute decline in FVC ≥10% predicted or death	Nintedanib n=166	Placebo n=108	difference	Nintedanib n=472	Placebo n=315	difference
	Hazard ratio: 0.59 (95% CI: 0.38, 0.89) in favour of nintedanib			Hazard ratio: 0.61 (95% CI: 0.48, 0.78) in favour of nintedanib		
	Treatment-by-subgroup interaction p=0.830					
Source: CS text pages.64-65, CS Figure 8						

9.3.2 Prespecified subgroup analysis from INPULSIS trials: FVC ≤70% vs. >70% predicted value

Evidence from the pooled INPULSIS studies for the primary endpoint (adjusted annual rate of decline in FVC) was described in the company submission for TA379 [1] stating that no statistically significant differences in outcomes by subgroup were found, but no numerical data were presented. New for this submission are some numerical data as shown in Table 44

Table 44 Subgroup analyses by FVC% predicted ≤70% versus >70% (INPULSIS trials)

Outcome	baseline FVC >70% predicted			baseline FVC ≤70% predicted		
Annual rate of decline in FVC	Nintedanib n=431	Placebo n=269	difference	Nintedanib n=207	Placebo n=154	difference
	NR	NR	109.0 (95% CI: 68.2, 149.9)	NR	NR	113.5 (95% CI: 51.3, 175.7)
	Treatment-by-time-by subgroup interaction p=0.9505					
Acute exacerbations	Nintedanib n=431	Placebo n=269		Nintedanib n=207	Placebo n=154	
	15 (3.5%)	9 (3.3%)		16 (7.7%)	23 (14.9%)	
Time to first acute exacerbation	Hazard ratio: 1.00 (95% CI: 0.44, 2.30)			Hazard ratio; 0.52 (95% CI: 0.28, 0.99)		
	Treatment-by-subgroup interaction p=0.1747					
Change from baseline in SGRQ total score over 52 weeks	Nintedanib n=410	Placebo n=263	difference	Nintedanib n=199	Placebo n=150	difference
			-0.34 (95% CI: -2.34, 1.65)			-3.34 (95% CI: -6.29, -0.38)
	Treatment-by-subgroup interaction p=0.0631					
Source: CS text page 70, CS Figures 48-50 NR : Not reported						

9.3.3 Subgroup analyses by baseline characteristics other than FVC % predicted

A narrative summary of post-hoc subgroup analyses conducted in the INPULSIS trials for patients with and without emphysema at baseline was presented in the company submission for TA379 and this is expanded on in the current submission with additional subgroup analyses reported for the first time in the current submission in CS section 2.7 and Appendix

E. No statistically significant differences between subgroups were observed for any of the subgroup analyses reported.

9.3.4 Adverse events in INPULSIS trials by baseline FVC >90% vs. FVC ≤90% predicted

In the baseline FVC >90% predicted subgroup receipt of nintedanib also led to a higher proportion of severe adverse events and adverse events that led to permanent drug discontinuation. Severe or serious adverse events occurred more frequently in the subgroup of patients with baseline FVC ≤90% predicted (nintedanib and placebo arms) than FVC >90% predicted (nintedanib and placebo arms).

Table 45 Adverse events in INPULSIS trials by baseline FVC >90% vs. FVC ≤90% predicted

Event n (%)	Baseline FVC >90% predicted		Baseline FVC ≤90% predicted	
	Nintedanib (n=166)	Placebo (n=108)	Nintedanib (n=472)	Placebo (n=315)
Any AE(s)	156 (94.0)	100 (92.6)	453 (96.0)	278 (88.3)
Severe AE(s) [a]	37 (22.3)	18 (16.7)	137 (29.0)	81 (25.7)
Serious AE(s) [b]	38 (22.9)	28 (25.9)	156 (33.1)	99 (31.4)
Fatal AE(s)	4 (2.4)	2 (1.9)	33 (7.0)	29 (9.2)
AE(s) leading to Permanent drug discontinuation [c]	36 (21.7)	8 (7.4)	87 (18.4)	46 (14.6)

Source: CS Table 38 edited by the EAG
Abbreviations: AE, adverse event.
[a] An event that was incapacitating or that caused an inability to work or to perform usual activities.
[b] An event that resulted in death, was immediately life threatening, resulted in persistent or clinically significant disability or incapacity, required or prolonger hospitalisation, was related to a congenital anomaly or birth defect, or was deemed serious for any other reason.
[c] AEs leading to treatment discontinuation in >2% of patients in any treatment group.