4.16

in collaboration with:

Erasmus School of Health Policy & Management

Maastricht University

Nintedanib for progressive fibrosing interstitial lung disease (PF-ILD) excluding idiopathic pulmonary fibrosis (IPF) [ID1599]

Produced by Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University

Authors
Rob Riemsma, Reviews Manager, Kleijnen Systematic Reviews Ltd, UK
Hannah Penton, Health Economist, Erasmus School of Health Policy & Management (ESHPM), EUR, the Netherlands
Pim Wetzelaer, Health Economist, ESHPM, EUR
Sean Harrison, Systematic Reviewer, KSR Ltd
Kevin McDermott, Systematic Reviewer, KSR Ltd
Nigel Armstrong, Health Economist, KSR Ltd
Charlotte Ahmadu, Health Economist, KSR Ltd
Gill Worthy, Statistician, KSR Ltd
Janine Ross, Information Specialist, KSR Ltd
Lisa Stirk, Information Specialist, KSR Ltd
Maiwenn Al, Health Economics Researcher, ESHPM, EUR
Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University

Correspondence to
Rob Riemsma, Kleijnen Systematic Reviews
Unit 6, Escrick Business Park
Riccall Road, Escrick
York, United Kingdom
YO19 6FD

Date completed 18/03/2021

Source of funding: This report was commissioned by the NIHR Evidence Synthesis Programme as project number STA 13/32/44.

Declared competing interests of the authors

None.

Acknowledgements

None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Confidential comparator prices are highlighted in green throughout the report.

Any de-personalised data are highlighted in pink throughout the report.

Rider on responsibility for report

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

This report should be referenced as follows:

Riemsma R, Penton H, Wetzelaer P, Harrison S, McDermott K, Armstrong N, Ahmadu C, Worthy G, Ross J, Stirk L, Al M, Kleijnen J. Nintedanib for progressive fibrosing interstitial lung disease (PF-ILD) excluding idiopathic pulmonary fibrosis (IPF) [ID1599]: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2021.

Contributions of authors

Rob Riemsma acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Hannah Penton acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Pim Wetzelaer, Charlotte Ahmadu and Nigel Armstrong acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Sean Harrison and Kevin McDermott acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Janine Ross and Lisa Stirk critiqued the search methods in the submission and contributed to the writing of the report. Maiwenn Al critiqued the company's economic evaluation, contributed to the writing of the report, and provided general health economic guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

6MWD	Six-minute walking distance
AE	Adverse events
ALAT	Latin American Thoracic Association
ALT	Alanine aminotransferase
aPTT	Activate partial thromboplastin time
AST	Aspartate aminotransferase
ATP	Adenosine triphosphate
ATS	American Thoracic Society
AZA	Azathioprine
Bid	Twice daily
BI	Budget impact
BIC	Bayesian information criterion
BMI	Body mass index
CE	Cost effectiveness
CEA	Cost effectiveness analysis
CEAC	Cost effectiveness acceptability curve
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CNS	Central nervous system
CrCl	Creatinine clearance
CRD	Centre for Reviews and Dissemination
CS	Company's submission
CSR	Clinical study report
CTD	Connective tissue disease
CVD	Cardiovascular diseases
DBL	Database lock
DLco	Diffusing capacity of the lung for carbon monoxide
DSU	Decision Support Unit
EMA	European Medicines Agency
EOT	End of trial
EPAR	European Public Assessment Report
EQ-5D	European Quality of Life-5 Dimensions
ERG	Evidence Review Group
ERS	European Respiratory Society
EUR	Erasmus University Rotterdam
FAD	Final appraisal document
FDA	Food and Drug Administration
FEV1	Forced expiratory volume in one second
FVC	Forced vital capacity
FVC %pred	Forced vital capacity percentage predicted
GI	Gastrointestinal
HP	Hypersensitivity pneumonitis
HR	Hazard ratio
HRCT	High-resolution computed tomography
HRQoL	Health-related quality of life
HSUV	Health state utility value
HTA	Health technology assessment
IC	Indirect comparison
ICER	Incremental cost effectiveness ratio
ICH-GCP	International Conference on Harmonisation Harmonised Tripartite Guideline for Good Clinical Practice
IIP	Idiopathic interstitial pneumonia
ILD	Interstitial lung disease

INR	International normalised ratio
iNSIP	Idiopathic nonspecific interstitial pneumonia
IPF	Idiopathic pulmonary fibrosis
ITC	Indirect treatment comparison
ITT	Intention to treat
IVIG	Intravenous immunoglobulin
JRS	Japanese Respiratory Society
K-BILD	King's Brief Interstitial Lung Disease Questionnaire
KSR	Kleijnen Systematic Reviews
L-PF	Living with pulmonary fibrosis
LYs	Life years
LYG	Life years gained
MeSH	Medical subject headings
MHRA	Medicines and Healthcare Products Regulatory Agency
MI	Myocardial infarction
MTA	Multiple technology appraisal
MTC	Mixed treatment comparison
NA	Not applicable
NAC	N-acetylcysteine
NR	Not reported
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NSCLC	Non-small cell lung cancer
NTD	Nintedanib
OCS	Oral corticosteroids
OS	Overall survival
PAH	Pulmonary arterial hypertension
PAS	Patient access scheme
PFD	Pirfenidone
PFS	Progression-free survival
PF-ILD	Progressive fibrosing interstitial lung disease
PICO	Patients, interventions, comparators, and outcomes
PLB	Placebo
PRESS	Peer Review of Electronic Search Strategies
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
PT	Prothrombin time
QALY	Quality adjusted life year
QoL	Quality of life
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
RR	Relative risk; Risk ratio
SAE	Serious adverse events
ScHARR	School of Health and Related Research
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of product characteristics
SSc	Systemic sclerosis
SSC-ILD	Systemic sclerosis associated interstitial lung disease

STA	Single technology appraisal
TA	Technology assessment
TEAE	Treatment emergent adverse events
TNF	Tumour necrosis factor
TTFAE	Time to first acute exacerbation
TTO	Time trade-off
UIP	Usual interstitial pneumonia
ULN	Upper limit of normal
UK	United Kingdom
UMC	University Medical Centre
USA	United States of America
WHO	World Health Organization
WTP	Willingness-to-pay

Table of Contents

Table of Tables

Table of Figures

1. EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the evidence review group (ERG) as being potentially important for decision making. If possible, it also includes the ERG's preferred assumptions and the resulting incremental cost effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 presents the key model outcomes. Section 1.3 discusses the decision problem, Section 1.4 issues relate to the clinical effectiveness, and Section 1.5 issues related to the cost effectiveness. Other key issues are discussed in Section 1.6 while a summary in presented in Section 1.7.

Information on key as well as non-key issues are in the main ERG report, see Sections 2 (decision problem), 3 (clinical effectiveness) and 4 to 6 (cost effectiveness) for more details.

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

1.1 Overview of the ERG's key issues

Table 1.1: Summary of key issues

ID1457	Summary of issue	Report sections
1	Relevant comparators are not included in the company submission (CS).	Sections 2.3 and 3.6
2	The comparator included in the CS does not reflect best supportive care (BSC) in the UK.	Sections 2.3 and 3.6
3	The ERG and company differed on their preferred extrapolation for overall survival (OS)	Section 4.2.6.1

BSC = best supportive care; CS = company submission; ERG = Evidence Review Group; OS = overall survival.

The key differences between the company's preferred assumptions and the ERG's preferred assumptions are that the company preferred to extrapolate OS using a Bayesian Weibull curve. However, although clinical experts consulted by the company could not choose between the two curves, the ERG preferred to use the frequentist Weibull curve. This was because the frequentist curve provided a better fit to long term survival data in idiopathic pulmonary fibrosis (IPF) patients taking nintedanib, used by the company to validate the long-term extrapolation. The ERG also made a minor adjustment to the health state utility value (HSUV) for the 80-89 predicted FVC percentage health state in order to maintain a consistent decline in utility with the decline in lung function.

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Increasing survival
- Reducing the number of acute exacerbations
- Slowing the decline in lung function

Overall, the technology is modelled to affect costs by:

- its higher unit price than current treatments
- decreasing costs associated with the deterioration of health due to progressive fibrosing interstitial lung disease (PF-ILD)

The modelling assumption that has the greatest effect on the ICER is:

- The extrapolation of overall survival

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

The decision problem addressed in the company submission (CS) is broadly in line with the final scope issued by NICE. However, not all relevant comparators as described in the NICE scope are included in the CS (Table 1.2) and the comparator included in the CS (placebo in the INBUILD trial) may not reflect current best practice or best supportive care (BSC) in the UK (Table 1.3).

Table 1.2: Key issue 1: Relevant comparators are not included in the CS

Report section	Sections 2.3 and 3.6
Description of issue and why the ERG has identified it as important	The description of the comparators in the NICE scope is: "Established clinical management without nintedanib (may depend on underlying cause of ILD) including, but not limited to: • immunosuppressants, such as azathioprine, cyclophosphamide, mycophenolate (do not currently have a marketing authorisation in the UK for this indication) • corticosteroids (do not have currently have a marketing authorisation in the UK for this indication) • infliximab (does not have currently have a marketing authorisation in the UK for this indication) • rituximab (does not have currently have a marketing authorisation in the UK for this indication) • best supportive care." The company only included one comparator, which they referred to as placebo. This was effectively all treatments received in the placebo arm of the INBUILD trial and which excluded immunomodulatory treatments that would have been current clinical practice.
What alternative approach has the ERG suggested?	The company should have included other relevant comparators as described in the NICE scope. However, given the lack of evidence for most comparators it is not clear how that could have been achieved. Therefore, the ERG has no suggestions for an alternative approach.
What is the expected effect on the cost effectiveness estimates?	The expected change to the ICER is unclear. However, if comparator treatments are more effective than those treatments received in the placebo arm (i.e. excluding immunomodulatory treatments for six months), the ICER will be less favourable for nintedanib.
What additional evidence or analyses might help to resolve this key issue?	The ERG is not aware of any additional evidence that would resolve this issue.

Table 1.3: Key issue 2: The comparator included in the CS may not reflect BSC in the UK

Report section	Sections 2.3 and 3.6
Description of issue and why the ERG has identified it as important	The comparator (placebo) in the company submission (CS) is defined as the treatment patients received in the control arm of the INBUILD trial. As stated by the company, "Due to the lack of availability of specific targeted therapies, immunomodulatory treatments (including azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids) have routinely been used in clinical practice for the treatment of ILD. However, their benefit-risk profiles in PF-ILD have not been established and they are not licensed for the treatment of PF-ILD. In order to avoid the potential impact of these drugs on the assessment of nintedanib in PF-ILD, their use was not allowed at randomisation and during the first 6 months of the treatment period. Patients who had taken these drugs could only participate in the trial if a wash-out period was observed before randomisation" (CS, pages 25-26). Therefore, it is clear that the treatments received in the placebo arm of the INBUILD trial do not represent current best practice or best supportive care (BSC) in the UK.
What alternative approach has the ERG suggested?	Given the evidence presented in the CS, the ERG has no suggestions for an alternative approach.
What is the expected effect on the cost effectiveness estimates?	The expected change to the ICER is unclear. However, if current best practice in the UK, which includes immunomodulatory treatments, is more effective than those treatments received in the placebo arm excluding immunomodulatory treatments, the ICER will be less favourable for nintedanib.
What additional evidence or analyses might help to resolve this key issue?	The ERG is not aware of any additional evidence that would resolve this issue.

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

The ERG did not identify any other key issues relating to clinical effectiveness.

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

A full summary of the cost effectiveness evidence review conclusions can be found in Section 6.4 of this report. The company's cost effectiveness results are presented in Section 5, the ERG's summary and detailed critique in Section 4, and the ERG's amendments to the company's model and results are presented in Section 6. The key issue in the cost effectiveness evidence is discussed in Table 1.4.

Table 1.4: Key issue 3: The selection of the parametric curve for overall survival (OS)

Report section	Section 4.2.6.1
Description of issue and why the ERG has identified it as important	The company preferred to extrapolate OS using a Bayesian Weibull curve given that: the Bayesian analysis was guided by external long-term IPF data, which could increase the accuracy of long-term predictions; clinicians considered the two Weibull options (frequentist or Bayesian) the most plausible in the long-term; the Weibull Bayesian provided a reasonably good fit to external IPF data. The choice of extrapolation of OS is a driver of model results.

Report section	Section 4.2.6.1
What alternative approach has the ERG suggested?	The ERG prefers to extrapolate OS using the frequentist Weibull, given that clinicians could not choose between the frequentist and Bayesian Weibull and the frequentist better fits the long-term nintedanib IPF external validation data presented.
What is the expected effect on the cost effectiveness estimates?	Extrapolating OS using the frequentist instead of the Bayesian Weibull adds approximately £8,000 to the company's post-clarification base-case ICER.
What additional evidence or analyses might help to resolve this key issue?	This issue would be resolved with longer term follow-up data in PF-ILD patients taking nintedanib, but this is not currently available.

1.6 Other key issues: summary of the ERG's view

The ERG did not identify any other key issues relating to cost effectiveness.

1.7 Summary of the ERG's view

The ERG's preferred assumptions are described in detail in Section 6.1.2 of this report and summarised in Table 1.5, with the impact of each assumption (applied independently to the company's post-clarification base-case) on results also shown. The results of the ERG preferred base-case, combining all the above assumptions, are displayed in the final row of the table.

An issue in the model submitted in response to clarification created an imbalance in the results of the probabilistic sensitivity analysis (PSA) compared to the determinist results, which should be fixed by the company in future stages of the appraisal in order to allow for the presentation of reliable PSA results to accompany the ERG base-case.

Scenario analyses conducted by the ERG are displayed in Section 6.2.2. The scenario which had the largest impact on results was extrapolating OS with the frequentist Weibull rather than the Bayesian Weibull.

Table 1.5: Summary of ERG's preferred assumptions and ICER

Scenario	Incremental cost	Incremental QALYs	ICER
Company's original CS base-case	[redacted]	[redacted]	[redacted]
Company's post-clarification base-case (including updated/corrected costs from clarification letter, including recurrent exacerbations in the model and including the age-adjustment of utilities)	[redacted]	[redacted]	[redacted]
Extrapolation of OS using the frequentist Weibull instead of the Bayesian Weibull (Key issue 3)	[redacted]	[redacted]	[redacted]
Adjustment of the health state utility value (HSUV) for 80-89 FVC % predicted state to maintain consistent trend in decline.	[redacted]	[redacted]	[redacted]
ERG's preferred base-case	[redacted]	[redacted]	[redacted]

BSC = best supportive care; CS = company submission; ERG = Evidence Review Group; FVC = forced vital capacity; HSUV = health state utility value; ICER = incremental cost effectiveness ratio.

2. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 2.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Population	Adults with progressive-fibrosing interstitial lung disease (excluding idiopathic pulmonary fibrosis)	Adults with progressive-fibrosing interstitial lung disease (excluding idiopathic pulmonary fibrosis)	N/A	The population is not completely in line with the NICE scope.
Intervention	Nintedanib	Nintedanib	N/A	The intervention is in line with the NICE scope
Comparator(s)	Established clinical management without nintedanib including, but not limited to: • immunosuppressants (such as azathioprine, cyclophosphamide, mycophenolate; do not currently have a marketing authorisation in the UK for this indication) • corticosteroids (do not have currently have a marketing authorisation in the UK for this indication) • infliximab (does not have currently have a marketing authorisation in the UK for this indication) • rituximab (does not have currently have a marketing	Placebo	At the trial design stage, there were no approved therapies for the treatment of PF-ILD, other than IPF. Currently, the only approved therapy is nintedanib. When diagnosis of ILD is confirmed, patients receive conventional treatment (such as corticosteroids and immunomodulatory agents) based on the specific type of ILD (see the proposed algorithm in Figure 3, page 19 [of the CS]). If the disease continues to progress despite use of these conventional treatments, a diagnosis of PF-ILD is then confirmed through pulmonary function tests, as well as radiological and clinical assessments. It is at this stage, once PF-ILD has been confirmed, that nintedanib should be considered as a treatment, as it is the only licensed treatment available for PF-ILD.	The comparators are not in line with the NICE scope. Also, placebo cannot be regarded as a comparator because it is not standard care i.e. no-one in actual clinical practice would receive a placebo. The comparator might be regarded instead as all other treatments administered to the patients (See Section 2.3 for further details).

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
	authorisation in the UK for this indication) • best supportive care		A consensus of clinical experts have advised that, whilst immunomodulatory agents may still be used to treat the inflammatory component of the disease, there are no randomised controlled trials to suggest that these unlicensed treatments have a positive impact on the chronic fibrotic progression of PF-ILD (i.e. delaying disease progression). Patients were eligible to participate in the trial if their ILD had worsened despite treatment with unapproved medications used in clinical practice to treat ILD. To minimise a potential impact on the efficacy and safety assessments, treatment for ILD with unapproved anti-inflammatory or immunomodulatory medications was required to be discontinued and a wash-out period was to be observed before randomisation of the patient. As there is currently no other targeted anti-fibrotic therapy licensed for the treatment of chronic fibrosing ILD with a progressive phenotype, the use of placebo as a control group was considered justified. However, initiation of concomitant immunomodulatory treatment as medically indicated was allowed for the management of worsening of the	

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
			disease after the first six months of the trial. Some patients received the treatments specified as comparators within the NICE scope, either for treatment of PF-ILD or the underlying condition (see full description on page 51-52 of the CS). Baseline and concomitant medication use are described in Section B.2.2 of the CS.	
Outcomes	The outcome measures to be considered include: • lung function • physical function • exacerbation rate • progression-free survival • mortality • adverse effects of treatment • health-related quality of life	• Rate of decline in FVC at 52 weeks (primary endpoint) • Absolute change from baseline in total score on K-BILD questionnaire at 52 weeks • Time until acute exacerbation of ILD or death at 52 weeks • Death at 52 weeks • Acute exacerbation of ILD or death up to DBL2 • Death up to DBL2 • AEs, serious AEs and severe AEs	N/A	The outcomes are generally in line with the NICE scope.
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	Not reported.	Not reported.	The economic analysis was conducted in line with the NICE reference case.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
	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 commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account.			
Subgroups to be considered	If the evidence allows subgroup analyses by ILD type will be considered.	Not reported	Not reported.	No subgroup analyses were performed.

Based on Table 1 and pages 11 to 12 of the CS.[1]

AE = adverse event; CS = company submission; DBL1 = database lock 1; DBL2 = database lock 2; FVC = forced vital capacity; ILD = interstitial lung disease; K-BILD = King's Brief Interstitial Lung Disease Questionnaire; N/A = not applicable

2.1 Population

The population defined in the scope is: "People with fibrosing interstitial lung disease that has progressed despite treatment (excluding idiopathic progressive fibrosis)".[2] The population in the CS is "Adults with progressive-fibrosing interstitial lung disease (excluding idiopathic pulmonary fibrosis)".[1] The population is not completely in line with the NICE scope, but is in line with the main trial (the INBUILD trial) described in the company submission, which included patients aged ≥18 years if they had a physician-diagnosed fibrosing ILD present with features of diffuse fibrosing lung disease of ≥10% extent on high-resolution computed tomography (HRCT), and met the protocol criteria for progression within 24 months of screening as assessed by the investigator.

Nintedanib has four approved marketing authorisations:

- As VARGATEF®, it is indicated in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or locally recurrent non-small cell lung cancer of adenocarcinoma tumour histology after first-line chemotherapy
- As OFEV®, it is indicated in adults for the treatment of:
  - Idiopathic pulmonary fibrosis (IPF)
  - Systemic sclerosis associated interstitial lung disease (SSc-ILD)
  - Other chronic fibrosing interstitial lung diseases with a progressive phenotype (PF-ILD)

Nintedanib was granted EMA marketing approval as VARGATEF®, for the treatment of non-small cell lung cancer in November 2014; and as OFEV®, for the treatment of IPF in January 2015, SSc-ILD in May 2020 and PF-ILD in July 2020. There are no restrictions in place under the current marketing authorisations.

The company claims that "patients with SSc-ILD with the progressing fibrosing phenotype are included in the INBUILD trial and are therefore included in the population considered in this submission, in line with the marketing authorisation for nintedanib" (CS, page 10).[1] However, it is unclear how many patients with SSc-ILD with the progressing fibrosing phenotype are included in the INBUILD trial and what their results were.

2.2 Intervention

The intervention (nintedanib) is in line with the scope.

The recommended dose is 150 mg nintedanib orally twice daily, administered approximately 12 hours apart. The 100 mg twice daily dose is only recommended to be used in patients who do not tolerate the 150 mg twice daily dose. In patients with mild hepatic impairment (Child Pugh A), the recommended dose of nintedanib is 100 mg twice daily approximately 12 hours apart.[1]

According to the company, no additional tests or investigations are required prior to the administration of nintedanib (CS, page 14).[1]

2.3 Comparators

The description of the comparators in the NICE scope is as follows: "Established clinical management without nintedanib (may depend on underlying cause of ILD) including, but not limited to:

- immunosuppressants, such as azathioprine, cyclophosphamide, mycophenolate (do not currently have a marketing authorisation in the UK for this indication)

- corticosteroids (do not have currently have a marketing authorisation in the UK for this indication)
- infliximab (does not have currently have a marketing authorisation in the UK for this indication)
- rituximab (does not have currently have a marketing authorisation in the UK for this indication)
- best supportive care".[2]

The company only included one comparator, which they referred to as placebo.[1]

ERG comment: The comparator (placebo) in the CS is defined as the treatment patients received in the control arm of the INBUILD trial. This should not be referred to as placebo because no one receives placebo in actual clinical practice. As stated by the company, "Due to the lack of availability of specific targeted therapies, immunomodulatory treatments (including azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids) have routinely been used in clinical practice for the treatment of ILD. However, their benefit-risk profiles in PF-ILD have not been established and they are not licensed for the treatment of PF-ILD. In order to avoid the potential impact of these drugs on the assessment of nintedanib in PF-ILD, their use was not allowed at randomisation and during the first 6 months of the treatment period. Patients who had taken these drugs could only participate in the trial if a wash-out period was observed before randomisation" (CS, pages 25-26).[1] Lack of license should not be a reason for excluding a treatment as a comparator: the test for inclusion is whether treatments are used in clinical practice, which the company points out is the case for the treatments excluded for the first six months. Therefore, it is clear that the treatment received in the placebo arm of the INBUILD trial does not represent current best practice or best supportive care (BSC) in the UK.

The company did not include rituximab and infliximab as comparators despite NICE explicitly requesting to make this comparison (see NICE Response to comments on draft scope[3]).

2.4 Outcomes

The NICE final scope lists the following outcome measures:[2]

- Measures of disease progression such as:
  - lung function
  - physical function
  - exacerbation rate
  - lung transplantation
- Mortality
- Adverse effects of treatment
- Health-related quality of life.

The following outcomes were assessed in the INBUILD trial:[1]

- Rate of decline in FVC at 52 weeks (primary endpoint)
- Absolute change from baseline in total score on K-BILD questionnaire at 52 weeks
- Time until acute exacerbation of ILD or death at 52 weeks
- Death at 52 weeks
- Acute exacerbation of ILD or death up to DBL2
- Death up to DBL2
- AEs, serious AEs and severe AEs

ERG comment: The outcomes are generally in line with the NICE scope. However, physical function does not seem to be reported. The K-BILD questionnaire is a self-completed health status questionnaire that comprises 15 items and a seven-point Likert response scale.[4] It has three domains: psychological, breathlessness and activities and chest symptoms. The K-BILD domain and total score ranges are 0–100; 100 represents best health status. Therefore, the activities domain from the K-BILD questionnaire might cover physical function. However, only K-BILD total scores have been reported in the CS. Therefore, physical function is not reported in the CS.

2.5 Other relevant factors

According to the company, nintedanib is innovative because until the recent approval of nintedanib for SSc-ILD and PF-ILD, there were no licensed treatments for patients with PF-ILD other than IPF. In addition, the company states that nintedanib is the first pharmacological treatment to show clinical evidence of slowing disease progression in patients with PF-ILD (CS, Section B.2.12).[1]

A simple PAS is in place for nintedanib (applies to current both indications – as VARGATEF in non-small cell lung cancer (NSCLC) and OFEV in IPF) and the company [redacted] for the current appraisal.[5] The PAS price is a [redacted] discount to the list price of £2,151 for both the 100 mg and 150 mg units = [redacted] (CS, page 105).[1]

According to the company, nintedanib is not expected to meet the criteria for end-of-life use (CS, page 54).[1] This is also illustrated by the statement from the company that "it is expected that patients with PF-ILD who are not receiving an anti-fibrotic therapy would have a median post-diagnosis survival of 2 to 5 years" (CS, page 53).[1, 6, 7] Therefore, treatment is not indicated for patients with a short life expectancy (normally less than 24 months).

According to the company, no equality issues related to the use of nintedanib for the treatment of adults with progressive-fibrosing interstitial lung disease are expected (CS, Section B.1.4).[1]

3. CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

3.1.1 Searches

Appendix D.1.1 of the CS details a systematic literature review (SLR) conducted to provide evidence on the efficacy of treatments for PF-ILDs other than IPF. In D.1.1 it states that the SLR aimed to identify RCTs that have evaluated pharmacological treatments for ILD with a progressive phenotype.

Searches were conducted on 13 August 2019 and were limited to English language publications. Update searches were run on 29 October 2019 and also on 26 May 2020. Databases were searched from date of inception. A summary of the sources searched is provided in Table 3.1.

Table 3.1: Data sources for the clinical effectiveness systematic review (as reported in CS)

	Resource	Host/source	Date ranges	Dates searched
Electronic databases	Embase	Ovid	1974 - 26/5/20	13/8/19 29/10/19 26/5/20
	Cochrane CDSR Cochrane CENTRAL	Cochrane library.com	Inception - 26/5/20	13/8/19 29/10/19 26/5/20
	MEDLINE, MEDLINE In-Process and Other Non-Indexed Citations MEDALL	Ovid	1946 - 26/5/20	28/6/20
Clinical Trial Registries	ClinicalTrials.gov		01/01/2010 to 13/08/2019 13/08/2019 to 26/05/2020	13/8/19 26/5/20
	The WHO International Clinical Trials Registry Platform		01/01/2010 to 13/08/2019	13/8/19
Conference proceedings	American Thoracic Society (ATS)	Online abstracts	2019	28/6/20
	British Thoracic Society (BTS)	Online abstracts	2018 2019	13/8/19 26/5/20
	European League Against Rheumatism (EULAR) - European Congress of Rheumatology	Online PDF abstract book	2019 2020	13/8/19 26/5/20
	European Respiratory Society (ERS) International Congress	Not reported	Not reported	Not reported
CDSR = Cochrane Database of Systematic Reviews; CENTRAL = Central Register of Controlled Trials				

ERG comments:

- A single set of searches was undertaken to identify clinical effectiveness and adverse events data. The CS provided sufficient details for the ERG to appraise the literature searches. Several databases and a good range of conference proceedings were searched, and reference checking was conducted. Searches were generally well documented, making them transparent and reproducible.
- The ERG was concerned that limiting the searches to English language may have introduced potential language bias. Current best practice states that that "Whenever possible review authors should attempt to identify and assess for eligibility all possibly relevant reports of trials irrespective of language of publication" [8] and that "research related to language bias supports the inclusion of non-English studies in systematic reviews".[9, 10]
- Study design filters were appropriately used but were not referenced.
- Separate adverse events (AE) searches were not performed. The clinical effectiveness searches incorporated a methodological filter intended to limit the search to RCTs. Guidance by the Centre for Reviews and Dissemination (CRD)[11] recommends that if searches have been limited by a study design filter, additional searches should be undertaken to ensure that adverse events that are long-term, rare or unanticipated are not missed. The ERG considered that it was possible that some relevant evidence may not have been identified as a consequence of the study design limits used.
- MeSH terms were used in the initial Embase searches but these were corrected in subsequent updates and efforts were made to ensure no studies were missed from the mistakes in the previous searches.

3.1.2 Inclusion criteria

The eligibility criteria used in the search strategy for randomised controlled trials (RCTs) and non-RCTs which was guided by expert clinical opinion on PF-ILD is presented in Table 3.2.

Table 3.2: Eligibility criteria used in search strategy for RCT and non-RCT evidence

Criteria	Inclusion criteria	Exclusion criteria
Population	Patients with ILD and progressive fibrosing phenotype	Patients with IPF
Interventions	Any dose of the following: • Nintedanib • Pirfenidone • Azathioprine • Cyclophosphamide • Rituximab • Mycophenolate mofetil • Corticosteroids • Methotrexate • Tocilizumab • Abatacept • Infliximab • Etanercept • Adalimumab	None
Comparators	Any	None

Criteria	Inclusion criteria	Exclusion criteria
Outcomes	Primary outcomes: • FVC • Progression-free survival/time to progression • Overall survival • Disease-related survival • Acute exacerbation of fibrosis / acute respiratory worsening Secondary outcomes: • FEV1 • FEV1/FVC • VC • TLC • DLco • HRCT • Corticosteroid sparing/corticosteroid use • AEs • Hospitalisation • Activity measures including, but not restricted to 6MWD test • HRQoL measures including, but not restricted to: • SGRQ • K-BILD • EQ-5D • SF-36 • HAQ-DI • VAS	None
Study design	RCTs	All other types of study designs
Language restrictions	English Language only	None
Date	No limits	None

Source: CS, Appendix D, Table 74, pages 161-162.[1]
6MWD = 6-minute walk distance; AE = adverse effect; DLco = diffusing capacity of the lung for carbon monoxide; EQ-5D = EuroQol-5 dimensions questionnaire; FVC = forced vital capacity; HAQ-DI = health assessment questionnaire disability index; HRCT = high-resolution computed tomography; HRQoL = health related quality of life; ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis; K-BILD = King's brief interstitial lung disease questionnaire; RCT = randomised controlled trial; SF-36 = 36-item short form health survey; SGRQ = St George's respiratory questionnaire; TLC = total lung capacity; VAS = visual analogue scale; VC = Vital Capacity; TLC = total lung capacity.

ERG comment: Given the final scope issued by NICE, the PICO (patients, interventions, comparators, and outcomes) inclusion criteria seem appropriate. However, it must be noted that two restrictions were placed on study design and language, respectively. Although an RCT is the gold standard for evaluating the effectiveness of an intervention or device, observational studies can contribute to the evidence base

for effective interventions, of a condition that has no current market authorisation. Additionally, the restriction to English language studies only, could mean that all relevant studies may not have been retrieved.

3.1.3 Critique of data extraction

Data extraction was carried out by one reviewer, and checked for consistency and accuracy by another reviewer.[1]

ERG comment: To minimise error during data extraction, it is usually advised that data extraction is carried out independently by two reviewers.

3.1.4 Quality assessment

Quality assessment of included studies was carried out by one reviewer, and checked by another.[1] The INBUILD trial was subjected to risk of bias assessment and judged to be of a low risk of bias.[1] Cost utility studies were assessed using the Drummond checklist and the NICE Decision Support Unit Recommendations were used to assess the quality of studies reporting utilities.[1]

ERG comment: The formal scale used to assess the risk of bias for the INBUILD trial was not described explicitly. However, we assume the company used the University of York, Centre for Reviews and Dissemination criteria.[11]

3.1.5 Evidence synthesis

The company notes and justifies the unfeasibility of conducting a quantitative evidence synthesis, despite there being the possibility of an indirect comparison between nintedanib and pirfenidone. This was due to the heterogeneity of patient and trial characteristics, and lack of comparable outcome reporting of pirfenidone vs. placebo, and nintedanib vs. placebo trials.[1] In addition, pirfenidone was not listed as a comparator in the NICE scope.

ERG comment: The ERG has no further comment regarding evidence synthesis (see also Section 3.3 in this report).

3.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

3.2.1 Details of the included trial: the INBUILD trial

The main evidence for the clinical effectiveness of nintedanib was from the INBUILD trial.[1, 12, 13] This trial (n=663) was a phase III, multicentre, randomised, double-blind, placebo-controlled, parallel-group study with follow-up at 52 weeks followed by a variable treatment period, where patients continued on blinded, randomised assigned treatment until the end of the trial or until a reason for treatment withdrawal was met. In both arms, patients could not be taking any immunomodulatory treatment at randomisation and for the first six months of the trial, but could do so for the remainder of the trial after six months. Immunomodulatory treatments included: azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids. The INBUILD trial was undertaken in 15 countries in North America, South America, Western Europe, and East Asia, including five centres (22 patients) in the UK. The purpose the INBUILD trial was to investigate the efficacy and safety of nintedanib for treating progressive-fibrosing lung disease.

Patients aged ≥18 years were eligible for enrolment if they had a physician-diagnosed fibrosing interstitial lung disease (ILD, such as connective tissue disease-associated ILD, rheumatoid arthritis-

associated ILD, systemic sclerosis–associated ILD, chronic fibrosing hypersensitivity pneumonitis, idiopathic non-specific interstitial pneumonia, unclassifiable idiopathic interstitial pneumonia, environmental/occupational lung disease, sarcoidosis and other ILDs), present with features of diffuse fibrosing lung disease of ≥10% extent on high-resolution computed tomography, and met the protocol criteria for progression within 24 months of screening as assessed by the investigator. In addition, patients were also required to have a forced vital capacity (FVC) >45% of predicted value and a diffusing capacity of the lungs for carbon monoxide (DLco) of >30% and <80% of predicted at randomisation. Patients who had taken immunomodulatory treatments as outlined above could participate in the trial if they observed a washout period before randomisation. Full inclusion and exclusion criteria are available in the company submission (CS, Table 5).[1]

Primary efficacy endpoint was the rate of decline in FVC as assessed over 52 weeks.

A summary of the methodology of the INBUILD trial is presented in Table 3.3 below.

Table 3.3: Summary of the methodology of the INBUILD trial

Trial design	Phase 3, multicentre, international, randomised, double-blind, placebo-controlled, 52-week study.
Participant eligibility criteria	Patients aged ≥18 years if they had a physician-diagnosed fibrosing ILD (such as connective tissue disease-associated ILD, rheumatoid arthritis- associated ILD, systemic sclerosis – associated ILD , chronic fibrosing hypersensitivity pneumonitis, idiopathic non-specific interstitial pneumonia, unclassifiable idiopathic interstitial pneumonia, environmental/occupational lung disease, sarcoidosis and other ILDs) present with features of diffuse fibrosing lung disease of ≥10% extent on HRCT, and met the protocol criteria for progression within 24 months of screening as assessed by the investigator.
Settings and locations where the data were collected	15 countries in North America, South America, Western Europe, and East Asia. The trial was run in the UK (22 patients enrolled in five centres).
Intervention	Oral nintedanib 150 mg twice daily (n=332)
Comparator	Oral placebo twice daily (n=331)
Primary outcome	Primary endpoint: annual rate of decline in FVC (mL/year) over 52 weeks in two co-primary populations (overall population and patients with UIP-like pattern on HRCT). Main secondary endpoints: change from baseline K-BILD total score at week 52; time to first acute ILD exacerbation or death over 52 weeks; time to death over 52 weeks.
Other outcomes used in the economic model / specified in the scope	• Acute exacerbation of ILD or death up to DBL2 • Death up to DBL2 • AEs, serious AEs and severe AEs Safety endpoints: • AEs over 52 weeks • Physical examination over 52 weeks • Vital signs over 52 weeks • Bodyweight over 52 weeks
Source: company submission [1]	

AEs = adverse events, DBL2 = database lock 2, FVC = forced vital capacity, HRCT = high-resolution computed tomography, HRQoL = health-related quality of life, ILD = interstitial lung disease, K-BILD = King's Brief Interstitial Lung Disease Questionnaire, L-PF = living with pulmonary fibrosis, UIP = usual interstitial pneumonia

ERG comment: The CS states that the INBUILD trial is likely to be reflective of clinical practice in England and Wales, given the trial endpoints, study population and comparators, and that five centres (22 patients) were located in the UK.[1] The primary endpoint, rate of decline in FVC, is a validated endpoint for studies of IPF.[14]

There are few registries for PF-ILD: in the UK, there is only the BTS ILD registry, which includes the UK IPF registry.[15] There were some differences between patients in the INBUILD trial and patients in the UK IPF registry: 54% of patients in INBUILD were male vs 79% in the registry; mean age was 66 years in INBUILD and 73.5 years in the registry; and 51% of INBUILD were former or current smokers vs 66% in the registry. However, the effects of these differences on the cost effectiveness analysis is unknown, and there is limited evidence of subgroup differences in the INBUILD trial, though there is a lack of power to detect even large differences. The UK IPF registry includes patients other than those with PF-ILD, so some differences are expected. Additionally, 22 patients in INBUILD (3.5%) were from the UK. As such, the cost effectiveness analysis is unlikely to be materially affected by the differences between INBUILD and the UK PF-ILD population.

However, one issue with the generalisability of results to a UK population is that the INBUILD trial did not allow off-label use of immunomodulatory treatments for the first six months of the trial in either arm. From six months into the trial, all participants were allowed to have immunomodulatory treatments in addition to nintedanib or placebo, and some patients were prescribed these. The CS states this "*reflects clinical opinion that treatment for worsening CTD or ILD was required and is reflective of the underlying treatment that would be seen in UK clinical practice*".[1] As such, while the INBUILD trial reflects UK clinical practice after six months, it does not necessarily reflect it during the first six months. However, as there is little evidence from trials for the effectiveness of off-label treatments for PF-ILD it is unknown how much this could affect the cost effectiveness analysis.

3.2.2 Statistical analyses of the INBUILD trial

The INBUILD trial was a superiority trial designed to demonstrate that nintedanib 150 mg twice daily was superior to placebo. The primary endpoint was reduction in FVC from baseline to 52 weeks, see Table 3.4. The initial 52 weeks of the trial were followed by a variable treatment period, where patients continued their blinded, randomised assigned treatment until the end of the trial or until a reason for treatment withdrawal was met. There were two primary co-populations: all patients, and patients with high-resolution computed tomography (HRCT) with usual interstitial pneumonia-like (UIP-like) fibrotic pattern only.

The analysis used all observations over 52 weeks and a random coefficient regression model. The analysis was performed on the intention to treat population, defined as all patients who were randomised and received at least one dose of study treatment. Continuous secondary endpoints were analysed using mixed effects models for repeated measures. Time-to-event secondary endpoints were analysed using Cox proportional hazards models and Kaplan-Meier plots; binary secondary endpoints were analysed using logistic regressions.

Table 3.4: Summary of statistical analyses in the INBUILD trial

Hypothesis objective	**Null hypothesis**: There is no difference in either of the co-primary populations (all patients and patients with HRCT with UIP-like fibrotic pattern only) in the annual rate of decline in FVC from baseline until 52 weeks between nintedanib 150 mg bid and placebo. **Alternative hypothesis**: There is a difference in the annual rate of decline in FVC between nintedanib 150 mg bid and placebo over 52 weeks, in either or both co-primary populations.
Statistical analysis	Primary analysis of the primary endpoint was based on all measurements taken over 52 weeks using a random coefficient regression model. Continuous secondary endpoints were analysed using Mixed Effects Models for Repeated Measures. Time-to-event secondary endpoints were analysed using Cox proportional hazards models and Kaplan-Meier plots; binary secondary endpoints were analysed using logistic regressions. Formal statistical testing was performed on both co-primary populations, and statistical significance declared if the analysis in both populations was significant at the two-sided 5% level, or if the analyses in either population were statistically significant at the two-sided 2.5% level. A Hochberg procedure was used to maintain an overall type 1 error rate of 5%.
Sample size, power calculation	A sample size of 600 patients (300 per randomised treatment group with 400 patients with UIP-like HRCT pattern) was expected to provide adequate power to demonstrate a clinically important treatment benefit on the primary endpoint, according to three scenarios (see CS, Table 13).[1] This included a scenario where the effect on the primary endpoint in both co-primary populations is lower than observed for IPF patients in the INPULSIS trials.
Data management, patient withdrawals	To reduce the amount of missing data, patients who discontinued the trial drug prior to completing the 52 week treatment period were asked to attend all visits as planned. In addition, for patients who prematurely discontinued trial medication and were unable to complete the scheduled visits, every attempt was made to collect information on vital status at week 52, at the time of data cut-off for the primary analysis and at the end of the trial. All aspects of data handling were performed according to guidelines and safety procedures established by the company for safety, completeness, consistency, accuracy, plausibility, legibility and adherence to the Clinical Trial Plan.

Source: company submission [1]
Bid = twice daily, FVC = forced vital capacity, HRCT = high-resolution computed tomography, IPF = idiopathic pulmonary fibrosis, UIP = usual interstitial pneumonia

ERG comment: The analysis of the INBUILD trial used appropriate statistical methods and the ERG has no concerns.

3.2.3 Baseline characteristics of the INBUILD trial

Table 3.5 shows the baseline characteristics of the participants in the INBUILD trial.

Briefly, the INBUILD trial had a total of 663 participants, n=332 received nintedanib and n=331 received placebo. The mean age of participants in the trial was 66 years. Both female and male

participants were included, and 54% of participants were male. The trial was conducted in 15 countries in North America, South America, Western Europe, and East Asia, and 74% of participants were white, 25% were Asian, and 1.5% were Black of African American. Fifty-one per cent of participants were former or current smokers, and 62% had UIP-like fibrotic pattern on HRCT while 38% had other fibrotic patterns. All participants matched at least one criterion for disease progression in the 24 months prior to screening: approximately 50% of participants had a relative decline in FVC ≥10% predicted; 31% had a relative decline in FVC ≥5–<10% predicted combined with worsening of respiratory symptoms and/or increased extent of fibrosis on HRCT; and 19% had worsened respiratory symptoms and increased extent of fibrosis on HRCT only. At baseline, participants had an average of 69% of their predicted FVC.

Table 3.5: Baseline characteristics in the INBUILD trial

	Nintedanib (n=332)	**Placebo (n=331)**
Male – no. (%)	179 (53.9)	177 (53.5)
Age – years	65.2±9.7	66.3±9.8
Former or current smoker – no. (%)	169 (50.9)	169 (51.1)
UIP-like fibrotic pattern on HRCT – no. (%)	206 (62.0)	206 (62.2)
Criteria for disease progression in 24 months before screening (grouped) – no. (%)		
Relative decline in FVC ≥10% predicted	160 (48.2)	172 (52.0)
Relative decline in FVC ≥5–<10% predicted combined with worsening of respiratory symptoms and/or increased extent of fibrosis on HRCT	110 (33.1)	97 (29.3)
Worsened respiratory symptoms and increased extent of fibrosis on HRCT only	62 (18.7)	61 (18.4)
FVC		
Mean value – mL	2,340±740	2,321±728
% of predicted value	68.7±16.0	69.3±15.2
DLco, mmol/min/kPa†	3.5±1.2	3.7±1.3
DLco, % of predicted value†	44.4±11.9	47.9±15.0
K-BILD questionnaire total score‡	52.5±11.0	52.3±9.8

Source: CS, Table 10, page 31.[1]

DLco = diffusion capacity of the lungs for carbon monoxide, FVC = forced vital capacity, HRCT = high-resolution computed tomography, K-BILD = King's Brief Interstitial Lung Disease, kPa = kiloPascal, UIP = usual interstitial pneumonia.

* Plus–minus values are means ± SD. † The DLco value was corrected for the haemoglobin level. ‡ K-BILD questionnaire total score ranges from 0–100, with higher scores representing better health status.

ERG comments: There was a balanced number of men and women in the INBUILD trial. Despite this, there was limited statistical power to detect differences in the effectiveness of nintedanib between genders. As such, although there was little evidence of a difference in effect between genders, there could still be a meaningful difference in the effectiveness of nintedanib between the genders. This may be relevant if the gender distribution of PF-ILD is not balanced in the UK: in the UK IPF registry 79% of the patients were male, though this includes patients who do not have PF-ILD. It should be noted, however, that patients on nintedanib had smaller declines in FVC at 52 weeks in both genders compared with placebo (male = 145.2 ml, 95% CI: 88.5 ml to 201.9 ml; female: 64.2 ml, 95% CI: 3.9 ml to 124.6

ml), and assuming that females are generally smaller than males, the relative rather than absolute changes in FVC may be more equal.

Further subgroup analyses showed little evidence for differences by age (<65 years versus ≥65 years, with patients ≥65 years having a slightly higher point estimate), by baseline FVC percentage predicted (≤70% versus >70%), by underlying ILD diagnosis (hypersensitivity pneumonitis, idiopathic nonspecific interstitial pneumonia, unclassifiable idiopathic interstitial pneumonia, autoimmune ILDs or other ILDs), or by race (White, Asian or Black or African American; though there was very little evidence for African Americans: 222.5 ml, 95% CI: -143.1 ml to 588.1 ml). As such, even though the participants were younger in the INBUILD trial compared with the UK IPF registry, this is unlikely to substantially affect the cost effectiveness analysis. Lung function (percentage FVC predicted) at presentation in the UK IPF registry was similar to INBUILD at recruitment, with 38% of patients having a predicted FVC of >80%, 57% of patients having a predicted FVC of 50 to 80%, and 5% of patients having a predicted FVC of <50%.[16] Race and underlying ILD diagnosis were not available in the UK IPF registry.

3.2.4 Risk of bias assessment of the INBUILD trial

The company assessed the quality of the INBUILD trial using the University of York, Centre for Reviews and Dissemination criteria.[11] Elements assessed were randomisation, allocation concealment, baseline comparability, care provider, participant and outcome assessor blinding, dropout imbalances, selective outcome reporting, use of intention to treat analysis and conflicts of interest. No information was provided on the number of reviewers who assessed the quality of the INBUILD trial, although it seems likely only one reviewer assessed the quality given the use of "*reviewer's judgement*" rather than "*reviewers' judgement*". The company concluded that all elements had been appropriately addressed in all three of the trials.

Table 3.6: Quality assessment of the INBUILD study

	How is the question addressed in the study?	Company	ERG
Was randomisation carried out appropriately?	Randomisation was performed using an IRT system.	Yes	Yes
Was the concealment of treatment allocation adequate?	Randomisation was performed by IRT, and trial packaging and labelling were identical. Colour, size and shape of nintedanib and placebo capsules were indistinguishable within dose strength but were different between dose strengths.	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Participants in all populations had similar baseline characteristics and treatment arms were well balanced.	Yes	Yes
Were the care providers, participants, and outcome assessors blind to treatment allocation?	Patients, investigators and everyone involved in trial conduct or analysis or with any other interest in this double-blind trial remained blinded with regard to the randomised treatment assignments until after DBL1.	Yes	Yes

	How is the question addressed in the study?	Company	ERG
Were there any unexpected imbalances in dropouts between groups?	Although there were some differences, these were consistent with the known safety profile of nintedanib in IPF and other indications.	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	All pre-specified outcomes have been reported.	No	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Efficacy and safety analyses were performed based on the treated set, which included all randomised patients who received ≥1 dose of trial medication; however, since all patients who were randomised received treatment with nintedanib or placebo this included all randomised patients. To reduce the amount of missing data, patients who discontinued trial drugs for any reason prior to completing the 52-week treatment period were asked to attend all visits and undergo all examinations as previously planned. In addition, for all patients who prematurely discontinued trial medication and were unable to complete the scheduled visits, every attempt was made to collect information on vital status at week 52, at the time of data cut-off for the primary analysis and at the end of the trial. The statistical model used for the primary analysis allowed for missing data, assuming they were missing at random.	Yes	Yes
Did the authors of the study publication declare any conflicts of interest?	All authors have clearly declared any conflicts of interest, and these are not considered to have biased the reporting or results of the study.	Yes	Yes

Source: CS, Appendix D, Table 90, page 224.[1]
DBL1 = database lock 1; IRT = Interactive Response Technology.

ERG comments:

- It is normally recommended that two reviewers are involved in the assessment of study quality to avoid bias and error.
- The ERG examined the clinical study report for the INBUILD trial and assessed it against the above criteria.[1, 12, 13] Randomisation and allocation concealment procedures appeared to be appropriate. Methods to ensure blinding of care providers, participants and outcome assessors also appeared to be appropriate. All outcomes appeared to be reported. Data from all participants who received at least one treatment dose were included, which is appropriate. The patients in the nintedanib and placebo arms appear similar, based on baseline demographics. Therefore, the ERG agrees the INBUILD trial was well conducted.

3.2.5 Efficacy results of the INBUILD trial

The results presented in the CS have been taken from two published manuscripts (Flaherty et al, 2019[12] and Wells et al, 2020[17]) and the clinical trial report[18]. Data from database lock 2 (DBL2) of INBUILD have been taken from a poster developed for the European Respiratory Society International Congress, 7-9th September 2020.[19]

The analysis of the INBUILD trial considered two co-primary analysis populations, the overall population (including all patients) and all patients with high-resolution computed tomography (HRCT) with usual interstitial pneumonia (UIP)-like fibrotic pattern only. In this report, we will only present data for the overall population.

The primary endpoint, annual rate of decline in FVC over 52 weeks, was met (see Table 3.7). Treatment with nintedanib reduced the adjusted annual rate of decline in FVC by 107.0 mL (p<0.001) in the overall population vs. placebo.

Table 3.7: Efficacy endpoint results in the INBUILD trial

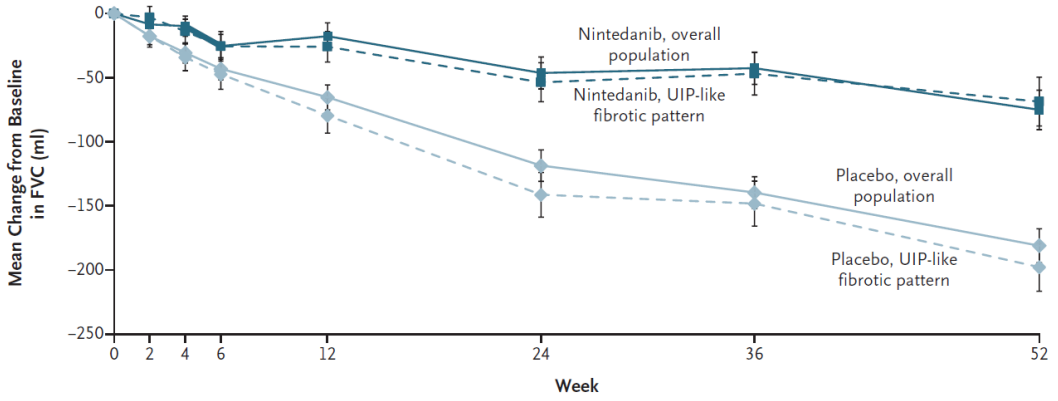
Endpoint	Nintedanib (N = 332)	Placebo (N = 331)	Difference vs. placebo (95% CI; p-value)
Primary endpoint			
Rate of decline in FVC at 52 weeks (mL/year)[†]			
Overall population	−80.8±15.1	−187.8±14.8	107.0 (65.4, 148.5; p<0.001)
Annual rate of decline in FVC (mL/ year) over the whole trial period up to DBL2			
Overall population	−118.14±11.4	−175.67±11.2	57.5 (26.1–89.0)
Main secondary endpoints			
Absolute change from baseline in total score on K-BILD questionnaire at 52 weeks[§]			
Overall population	0.55±0.60	−0.79±0.59	1.34 (−0.31, 2.98; p=0.1115)[‡]
Acute exacerbation of ILD or death at 52 weeks (no. with event/total no. [%])			
Overall population	26/332 (7.8)	32/331 (9.7)	0.80 (0.48, 1.34; p=0.3948)[‡¶]
Time to first acute ILD exacerbation or death over the whole trial period up to DBL2 (no. with event/total no. [%])			
Overall population	46/332 (13.9)	65/331 (19.6)	0.67 (0.46 to 0.98)[¶]
Death at 52 weeks (no. with event/total no. [%])			
Overall population	16/332 (4.8)	17/331 (5.1)	0.94 (0.47, 1.86; p=0.8544)[‡¶]
Time to death over the whole trial period up to DBL2 (no. with event/total no. [%])			
Overall population	36/332 (10.8)	45/331 (13.6)	0.78 (0.50 to 1.21)[¶]
Other secondary endpoints assessed until DBL2 in the overall population (no. with event/total no. [%])			
Time to progression (≥10% absolute decline in FVC % predicted) or death	134/332 (40.4)	181/331 (54.7)	0.66 (0.53 to 0.83)[¶]
Time to death due to a respiratory cause	21/332 (6.3)	30/332 (9.1)	0.68 (0.39 to 1.18)[¶]

Source: CS, Table 15, page 39-40.
FVC = forced vital capacity; ILD = interstitial lung disease; K-BILD = King's Brief Interstitial Lung Disease Questionnaire; NR = not reported; UIP = usual interstitial pneumonia.

Endpoint	Nintedanib (N = 332)	Placebo (N = 331)	Difference vs. placebo (95% CI; p-value)
† For the primary end point, the patients with a UIP-like fibrotic pattern included 206 in each treatment group. The patients with other fibrotic patterns included 126 in the nintedanib group and 125 in the placebo group. ‡ The widths of the confidence intervals have not been adjusted for multiple comparisons, so the intervals should not be used to infer definitive treatment effects. § For the analysis of the scores on the K-BILD questionnaire, 332 patients were included in the nintedanib group and 330 in the placebo group in the overall population; among the patients with a UIP-like fibrotic pattern, included were 206 patients and 205 patients, respectively. ¶ The difference was assessed as a hazard ratio. Data are taken from Flaherty 201912 and the Clinical Trial Report18. DBL2 data have been taken from the Clinical Trial Report18 and a poster developed by Flaherty et al for the European Respiratory Society International Congress, 7-9th September 2021.19			

The curves of observed change from baseline in FVC in the nintedanib and placebo groups separated early and continued to diverge up to 52 weeks follow-up (Figure 3.1).

Figure 3.1: Decline from baseline in FVC at 52 weeks



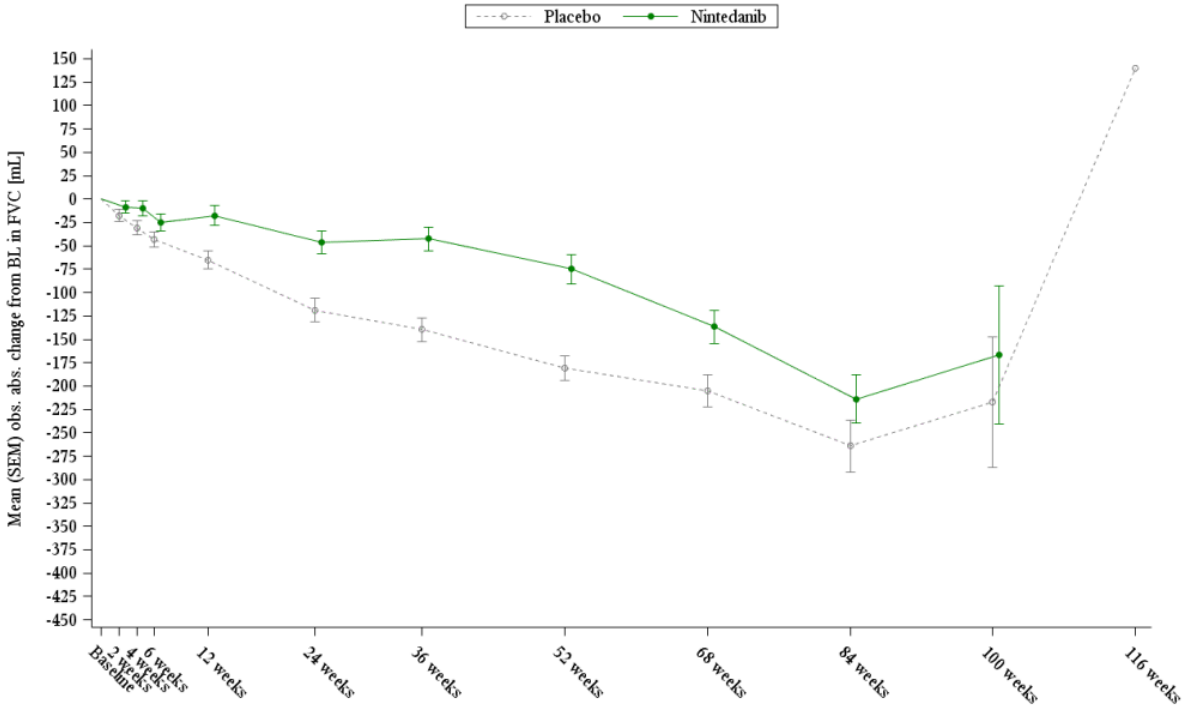
No. of Patients								
Overall population								
Nintedanib	332	326	320	322	314	298	285	265
Placebo	331	325	326	325	320	311	296	274
Patients with UIP-like fibrotic pattern								
Nintedanib	206	203	200	199	193	180	171	160
Placebo	206	202	202	201	197	190	176	162

Source: CS, Figure 6, page 41.[1]
Abbreviations: FVC, forced vital capacity; UIP, usual interstitial pneumonia

As can be seen from Table 3.7, the difference in the annual rate of decline in FVC (mL/year) between nintedanib and placebo is smaller over the whole trial period up to DBL2 (difference vs. placebo: 57.5 (95% CI: 26.1to 89.0)) than it is at 52 weeks (difference vs. placebo: 107.0 (95% CI: 65.4 to 148.5)). Therefore, it is likely the curves converge after 52 weeks. In order to see what happens to the curves after 52 weeks, the ERG asked the company to provide a figure such as Figure 6 in the CS for the 'Annual rate of decline in FVC (mL/year) over the whole trial period up to DBL2' (Response to clarification, Question A5, page 11).[3] In response, the company provided Figure 3.2 below. As can be seen from Figure 3.2, the curves of observed change from baseline in FVC in the nintedanib and placebo groups separated early and continued to diverge up to 52 weeks follow-up. However, after 52 weeks follow-up the curves move closer together again. The company does warn that "The analysis of annual rate of decline in FVC (mL/year) including data over the whole trial should be interpreted with caution.

Because of the trial design with a variable duration of Part B, many patients had missing FVC assessment values after week 52" (Response to Clarification, Question A5, page 11).[3]

Figure 3.2: Mean of observed absolute change from baseline in FVC (mL) over the whole trial – treated set, overall population



Source: Response to Clarification, Question A5, Figure 2, page 12.[3]
Abbreviations: FVC, forced vital capacity

In the overall population, treatment with nintedanib did not show a significant difference in health-related quality of life (HRQoL) as measured by the King's Brief Interstitial Lung Disease (K-BILD) questionnaire compared with placebo (adjusted mean difference 1.34; 95% CI: -0.31 to 2.98); the change from baseline total score was small in both treatment groups.

The hazard ratio (HR) for time to first acute ILD exacerbation or death also showed no significant difference between nintedanib and placebo (HR 0.80; 95% CI: 0.48 to 1.34); nor did the HR for time to death over 52 weeks (HR 0.94; 95% CI: 0.47 to 1.86).

Over the whole trial (up to DBL2), in the overall population, a lower proportion of patients in the nintedanib group (13.9%) than in the placebo group (19.6%) had an event of first acute ILD exacerbation or death; this difference was statistically significant (HR 0.67; 95% CI: 0.46 to 0.98) (Table 3.7).

In the overall population, the percentage of patients who died over 52 weeks was similar between treatment groups (%; n/N, nintedanib: 4.8%; 16/332, placebo: 5.1%, 17/331). The HR for time to death over 52 weeks was 0.94 (95% CI: 0.47 to 1.86). Over the whole trial (up to DBL2), in the overall population, a lower proportion of patients died in the nintedanib group (10.8%) than in the placebo group (13.6%). However, this difference was not statistically significant (HR 0.78; 95% CI: 0.50 to 1.21).

In the overall population, over the whole trial period (up to DBL2), a lower proportion of patients in the nintedanib group (40.4%; n/N, 134/332) than in the placebo group (54.7%; n/N, 181/331) progressed

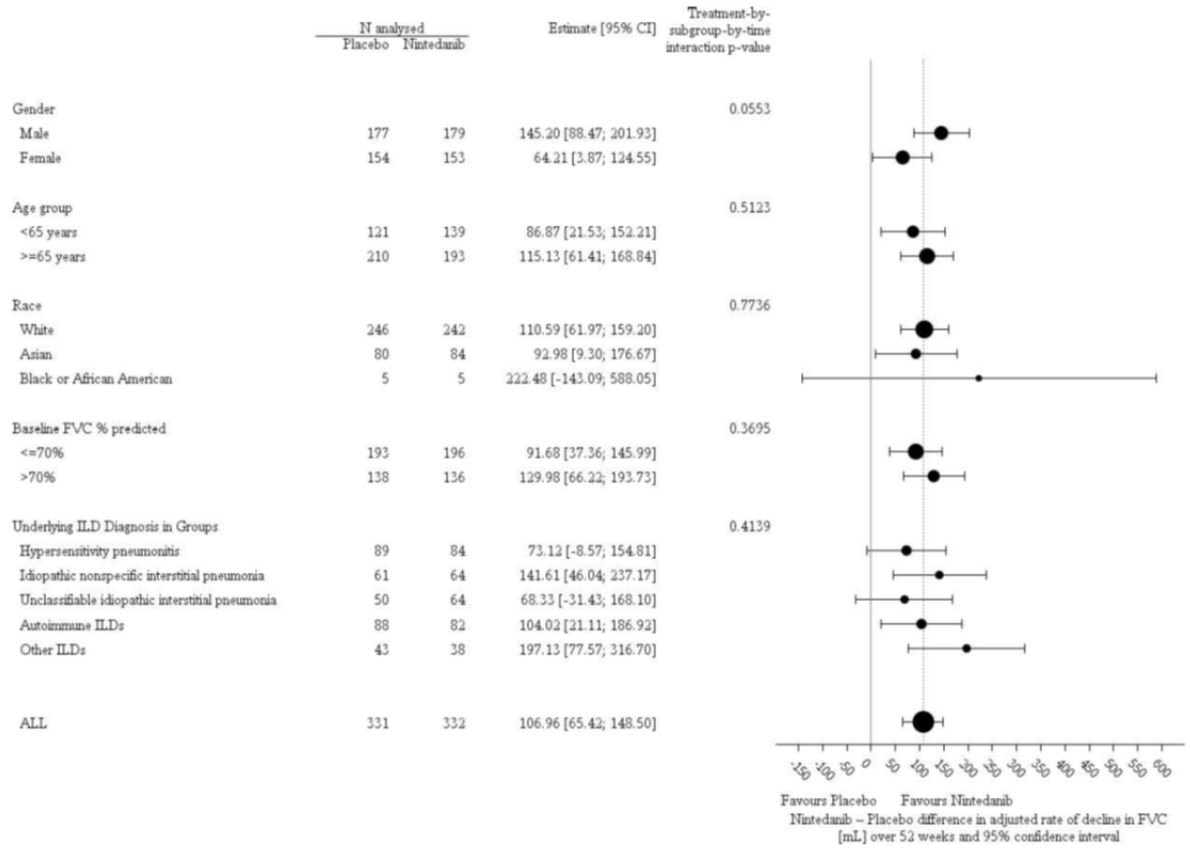
(defined as ≥10% absolute decline in FVC % predicted) or died. Most of these patients had an event of progression (34.3% nintedanib vs. 48.3% placebo). Treatment with nintedanib reduced the risk of progression or death by 34% compared with placebo, as indicated by the HR of 0.66 (95% CI: 0.53 to 0.83). In the overall population, over the whole trial period (up to DBL2), a lower proportion of patients died due to respiratory cause in the nintedanib group (6.3%; n/N, 21/332) than in the placebo group (9.1%; n/N, 30/331). However, this difference was not statistically significant (HR 0.68; 95% CI: 0.39 to 1.18).

3.2.5.1 Subgroup analyses

The NICE scope specified that if the evidence allows subgroup analyses by ILD type, these should be considered.[2] The company performed subgroup analyses for the description of the trial population, the primary endpoint and safety endpoints in the following pre-planned groups: gender, age (<65 years vs. over 65 years), race, baseline FVC percentage predicted (≤70% vs >70%) and underlying clinical ILD diagnosis in groups.

According to the company, none of the demographics nor clinical characteristics had a substantial influence on the treatment effect of nintedanib vs. placebo in the overall population (Figure 3.3). All point estimates were in favour of nintedanib vs. placebo. An additional analysis investigated the impact of the underlying ILD diagnoses by employing the method of excluding ILD diagnosis groups one by one, thus exploring the influence of the excluded ILD diagnosis group on the overall treatment effect. The point estimates and CIs were very similar in these analyses, showing that the treatment effect was not driven by one of the ILD diagnosis groups.

Figure 3.3: Forest plot for the primary endpoint analysis in subgroups (overall population)



3.2.6 Adverse events

The CS reported adverse events (AEs) that occurred in both the nintedanib and placebo groups over the course of 52 weeks in the INBUILD trial (CS, page 47, Table 19 – see also Table 3.8 below) and data presented was consistent with that in the published study.[13] The CS reported that overall, the percentages of patients with any AEs (nintedanib: 95.5% v placebo: 89.4%) and serious AEs (nintedanib: 32.2% v placebo: 33.2%) were similar in both groups.

Table 3.8: AEs in the INBUILD trial (overall population, 52 weeks)

AE	Nintedanib	Placebo
Any (n [%])	317 (95.5)	296 (89.4)
Any except for progression of interstitial lung disease	317 (95.5)	295 (89.1)
Most frequent AEs		
Diarrhoea	222 (66.9)	79 (23.9)
Nausea	96 (28.9)	31 (9.4)
Bronchitis	41 (12.3)	47 (14.2)
Nasopharyngitis	44 (13.3)	40 (12.1)
Dyspnoea	36 (10.8)	44 (13.3)
Vomiting	61 (18.4)	17 (5.1)
Cough	33 (9.9)	44 (13.3)
Decreased appetite	48 (14.5)	17 (5.1)
Headache	35 (10.5)	23 (6.9)
Alanine aminotransferase increased	43 (13.0)	12 (3.6)
Progression of ILD	16 (4.8)	39 (11.8)
Weight loss	41 (12.3)	11 (3.3)
Aspartate aminotransferase increased	38 (11.4)	12 (3.6)
Abdominal pain	34 (10.2)	8 (2.4)
Severe AEs	60 (18.1)	73 (22.1)
Serious AEs	107 (32.2)	110 (33.2)
Fatal AE		
Any	11 (3.3)	17 (5.1)
Any except for progression of ILD	10 (3.0)	14 (4.2)
AE leading to discontinuation	65 (19.6)	34 (10.3)
AE leading to permanent dose reduction	110 (33.1)	14 (4.2)

Source: CS, Table 19, pages 47-48.[1]
AE = adverse event; ILD = interstitial lung disease

AEs which were most frequently reported by System Organ Class (SOCs with a frequency >20% in either treatment group) were described in the CS. These included gastrointestinal disorders (nintedanib: 80.7%; placebo: 45.0%); infections and infestations (53.3% vs. 55.9%); respiratory, thoracic and mediastinal disorders (38.6% vs. 43.5%); investigations (34.3% vs. 16.9%); general disorders and administration site conditions (25.9% vs. 25.7%); musculoskeletal and connective tissue disorder (23.2% vs. 26.3%); nervous system disorders (20.8% vs. 16.3%); and metabolism and nutrition disorders (20.8% vs. 11.5%).

It is of note that gastrointestinal disorders occurred more frequently (80.7% vs 45.0%) in the nintedanib group than the placebo group while respiratory, thoracic and mediastinal disorders (38.6% vs. 43.5%) occurred more frequently in the placebo group.

The CS provided frequency detail on occurrence of specific AEs in each treatment group and where a >5%-point difference between groups exists it is noteworthy and included here. The following AEs were more frequent in the nintedanib group than the placebo group; diarrhoea (66.9% versus 23.9%); nausea (28.9% versus 9.4%); vomiting (18.4% versus 5.1%); decreased appetite (14.5% versus 5.1%); alanine aminotransferase increases (13.0% versus 3.6%); weight loss (12.3% versus 3.3%); aspartate aminotransferase increases (11.4% versus 3.6%); and abdominal pain (10.2% versus 2.4%). Furthermore, there was an increased frequency of AEs leading both to discontinuation (19.6% versus 10.3%) and to permanent dose reduction (33.1% versus 4.2%) in the nintedanib group; however, progression of ILD occurred more frequently in the placebo group (11.8% versus 4.8%).

The CS elaborated on the frequency of reported AEs leading both to discontinuation and dose reduction and data demonstrated that diarrhoea (nintedanib: 5.7%, placebo: 0.3%), was the most frequently reported AE leading to treatment discontinuation, while the most frequently reported AEs leading to permanent dose reduction were diarrhoea (nintedanib: 16.0%, placebo: 0.9%) and alanine aminotransferase increased (5.4% vs. 0.6%). The CS also reported that these were the most common other significant AEs (diarrhoea: 19.9% vs. 1.2%, alanine aminotransferase increased: 6.6% vs. 0.6%, and aspartate aminotransferase increased: 5.4% vs. 0.3%).

Investigator-defined drug related AEs were more frequently reported in the nintedanib group and were consistent with increased reporting by SOC of gastrointestinal disorder, these included diarrhoea (nintedanib: 59.0%, placebo: 17.8%), nausea (23.8% vs. 5.7%), and vomiting (12.3% vs. 2.1%).

There were broadly similar results (<5%-point difference) in the frequency of serious adverse events (SAEs) with the noticeable exception of interstitial lung disease which was more common in the placebo group (9.4% vs. 3.3%).

Overall, the data presented in the CS demonstrated that in the described 52 weeks, the groups are similar with respect to frequency of any and serious adverse events. Gastrointestinal discomfort, and in particular diarrhoea, was the most common adverse event and was most frequently reported in those who had taken nintedanib. Administration of nintedanib was associated with increased frequency of indicators of hepatic injury, and gastrointestinal disorder that required a permanent reduction in dosage.

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

The company state that "as an exercise of due diligence, the feasibility of a quantitative evidence synthesis, such as an NMA or Bucher's indirect comparison with available treatments used in clinical practice, was assessed based on evidence identified in the SLR described in Appendix D" (of the CS).[1]

Six studies were explored by the company in the feasibility assessment as they met the criteria for inclusion in the SLR and reported results. Only one of these studies was deemed suitable for an indirect comparison according to the company.

Therefore, the company concluded that "an indirect comparison at 24 weeks was technically possible between nintedanib and pirfenidone, based on INBUILD[12] and NCT03099187[20]. However, since PF-ILD is a chronic condition, this comparison is expected to be immature. As a result, no indirect treatment comparisons were undertaken."[1]

ERG comments: The ERG agrees with the company that none of the studies identified in the systematic literature review performed by the company are suitable for an indirect comparison; mainly because pirfenidone is not a relevant comparator according to the NICE scope.

However, as described in Section 2.3 of this report, this means that none of the comparators described in the NICE scope have been included in the CS.

3.4 Critique of the indirect comparison and/or multiple treatment comparison

The company concluded that "it was not possible to conduct any indirect or mixed treatment comparisons due to lack of published evidence for comparator treatments".[1] Therefore, no indirect comparison and/or multiple treatment comparison have been described in the CS.

3.5 Additional work on clinical effectiveness undertaken by the ERG

No additional work on clinical effectiveness was undertaken by the ERG

3.6 Conclusions of the clinical effectiveness section

The population is not completely in line with the NICE scope but is in line with the main trial (the INBUILD trial) described in the company submission, which included patients aged ≥18 years if they had a physician-diagnosed fibrosing ILD present with features of diffuse fibrosing lung disease of ≥10% extent on HRCT and met the protocol criteria for progression within 24 months of screening as assessed by the investigator.

The company only included one comparator, referred to as placebo. The comparator (placebo) in the CS was defined as the treatment patients received in the control arm of the INBUILD trial. As stated by the company, "Due to the lack of availability of specific targeted therapies, immunomodulatory treatments (including azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids) have routinely been used in clinical practice for the treatment of ILD. However, their benefit-risk profiles in PF-ILD have not been established and they are not licensed for the treatment of PF-ILD. In order to avoid the potential impact of these drugs on the assessment of nintedanib in PF-ILD, their use was not allowed at randomisation and during the first 6 months of the treatment period. Patients who had taken these drugs could only participate in the trial if a wash-out period was observed before randomisation" (CS, pages 25-26).[1] Therefore, it is doubtful that the placebo group in the INBUILD trial represents current best practice or best supportive care (BSC) in the UK.

The company did not include rituximab and infliximab as comparators despite NICE explicitly requesting to make this comparison (see NICE Response to comments on draft scope[3]).

The main evidence for the clinical effectiveness of nintedanib was from the INBUILD trial.[1, 12, 13] This trial (n=663) was a phase III, multicentre, randomised, double-blind, placebo-controlled, parallel-group study with follow-up at 52 weeks followed by a variable treatment period, where patients continued on blinded, randomised assigned treatment until the end of the trial or until a reason for treatment withdrawal was met. In both arms, patients could not be taking any immunomodulatory treatment at randomisation and for the first six months of the trial, but could do so for the remainder of the trial after six months. Immunomodulatory treatments included: azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids. The INBUILD trial was undertaken in 15 countries in North America, South America, Western Europe, and East Asia, including five centres (22 patients) in the UK. The purpose the INBUILD trial was to investigate the efficacy and safety of nintedanib for treating progressive-fibrosing lung disease.

The primary endpoint, annual rate of decline in FVC over 52 weeks, was met. Treatment with nintedanib reduced the adjusted annual rate of decline in FVC by 107.0 mL ($p<0.001$) in the overall population vs. placebo. In the overall population, treatment with nintedanib did not show a significant difference in HRQoL as measured by the K-BILD questionnaire compared with placebo (adjusted mean difference 1.34; 95% CI: -0.31 to 2.98); the change from baseline total score was small in both treatment groups.

The hazard ratio (HR) for time to first acute ILD exacerbation or death also showed no significant difference between nintedanib and placebo (HR 0.80; 95% CI: 0.48 to 1.34); nor did the HR for time to death over 52 weeks (HR 0.94; 95% CI: 0.47 to 1.86).

Over the whole trial (up to DBL2), in the overall population, a lower proportion of patients in the nintedanib group (13.9%) than in the placebo group (19.6%) had an event of first acute ILD exacerbation or death; this difference was statistically significant (HR 0.67; 95% CI: 0.46 to 0.98).

In the overall population, the percentage of patients who died over 52 weeks was similar between treatment groups (%; n/N, nintedanib: 4.8%; 16/332, placebo: 5.1%, 17/331). The HR for time to death over 52 weeks was 0.94 (95% CI: 0.47 to 1.86). Over the whole trial (up to DBL2), in the overall population, a lower proportion of patients died in the nintedanib group (10.8%) than in the placebo group (13.6%). However, this difference was not statistically significant (HR 0.78; 95% CI: 0.50 to 1.21).

Overall, the data presented in the CS demonstrated that over the 52 weeks follow-up, the groups were similar with respect to frequency of any and serious adverse events. Gastrointestinal discomfort, and in particular diarrhoea, was the most common adverse event and was most frequently reported in those who had taken nintedanib. Administration of nintedanib was associated with increased frequency of indicators of hepatic injury, and gastrointestinal disorder that required a permanent reduction in dosage.

4. COST EFFECTIVENESS

4.1 ERG comment on company's review of cost effectiveness evidence

This section pertains mainly to the review of cost effectiveness analysis studies. However, the search section (5.1.1) also contains summaries and critiques of other searches related to cost effectiveness presented in the company submission. Therefore, the following section includes searches for the cost effectiveness analysis review, measurement and evaluation of health effects as well as for cost and healthcare resource identification, measurement and valuation.

4.1.1 Searches performed for cost effectiveness

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness presented in the company submission.

Appendix G.1.1 of the CS details an SLR which was conducted to identify published cost-effectiveness studies, health-related quality-of-life studies, and costs and healthcare resource use.

Searches were conducted on 9 June 2020. and were limited to English language publications. Databases were searched from date of inception. A summary of the sources searched is provided in Table 4.1.

Table 4.1: Data sources for the cost effectiveness systematic review (as reported in CS)

	Resource	Host/source	Date range	Date searched
Electronic databases	Embase	Ovid	1974 - 9/6/20	9/6/20
	Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R)	Ovid	1946 - 9/6/20	9/6/20
	Cochrane CDSR Cochrane CENTRAL	Cochranelibrary.com	Inception - 29/6/20	9/6/20
	NIHR Centre for Reviews and Dissemination (CRD; including NHS EED, DARE, and HTA)	CRD website	Inception - 29/6/20	9/6/20
	Econlit	Ovid	1886-9/6/20	9/6/20
Conference proceedings	ATS	via database searches	2018 onwards	9/6/20
	BTS			
	ISPOR			
	ERS			
	EULAR	Online abstract archive		
Additional resources	Clinicaltrials.gov	No details provided	No details provided	9/6/20
	The WHO International Clinical Trials Registry			

	Resource	Host/source	Date range	Date searched
	Tufts Medical Center Cost Effectiveness Analysis registry			
	SCHARR health utilities database			
	HERC utilities database			
NHS EED = NHS Economic Evaluation Database; HTA Database = Health Technology Assessment database; CRD - Centre for Reviews and Dissemination; ATS = American Thoracic Society; BTS = British Thoracic Society; EULAR = European League Against Rheumatism; ERS = European Respiratory Society; ISPOR = International Society for Pharmacoeconomics and Outcomes Research				

ERG comments:

- A single set of searches were undertaken for economic evaluations and healthcare resource use and cost studies, quality of life and health state utility value studies.
- Several databases and a good range of conference proceedings were searched, and reference checking was conducted. Searches were well documented, making them transparent and reproducible. There were no searches of health technology assessment organisation websites.
- The ERG was concerned that limiting the searches to English language may have introduced potential language bias (please see comments in Section 3.1.1 of this report regarding language bias.
- Study design filters were appropriately used but were not referenced.

4.1.2 Inclusion/exclusion criteria

In- and exclusion criteria for the review on cost effectiveness studies, utilities and costs and resource use are presented in Table 4.2.

Table 4.2: Eligibility criteria for the systematic literature reviews

	Inclusion criteria	Exclusion criteria
Patient population	Studies including any proportion of patients with ILD and progressive fibrosing phenotype defined as: • FVC – any decline in FVC percentage predicted at baseline • DLco – any decline in DLco at baseline • HRCT – worsening of fibrotic features on imaging; images identifying progression of disease • Reference to the progression of lung fibrosis (without any disease specific criteria) are to be included.	Patients with IPF
Intervention	No limits applied in searching. No limits applied during screening for costs, HCRU, or utilities.	
Comparator		

	Inclusion criteria	Exclusion criteria
	Economic evaluation studies limited to the following specific treatments during screening: • Nintedanib • Pirfenidone • Azathioprine • Cyclophosphamide • Rituximab • Mycophenolate mofetil • Prednisone • Prednisolone • Tocilizumab • Abatacept • Methotrexate • Etanercept • Infliximab • Adalimumab	
Outcomes - Economic evaluations	Cost utility analysis.	
Outcomes - Utility studies	• Utility values. • Mapping algorithms.	
Outcomes -Cost/resource use studies	• Direct and indirect costs. • Direct and indirect resource use.	
Study design	Any	• Case reports and case studies. • Editorials. • Retracted studies/ data.
Geography	No geographic limits.	Studies not conducted in Ireland and England will be considered only where no data specific to Ireland and England are identified.
Language	English Language abstracts	Non-English language

Source: Table 97 of the CS.[1]
DLco = Diffusing capacity of the lung for carbon monoxide; FVC = forced vital capacity; HCRT = high-resolution computed tomography; HCRU = healthcare resource use; ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis.

ERG comment: The ERG agrees that the eligibility criteria are suitable to fulfil the company's objective to identify cost effectiveness studies. The restriction to only include consider cost utility analyses (CUAs) in the economic evaluation SLR may have caused some relevant literature to have been missed.

4.1.3 Conclusions of the cost effectiveness review

Appendices G-I of the CS provide an overview of the results of the cost effectiveness, utility and resource use and costs SLRs. No cost effectiveness or HRQoL studies were included in the review. Four publications reporting on two studies were included for cost and resource use, but these were not used in the model.

Eligibility criteria were suitable for the SLR performed and the review was conducted appropriately. However, the English language restriction may have caused relevant literature to be missed.

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

4.2.1 NICE reference case checklist

Table 4.3 provides the ERGs comments on how well this submission aligns with the NICE reference case.

Table 4.3: NICE reference case checklist

Element of health technology assessment	Reference case	ERG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	As per the reference case
Perspective on costs	NHS and PSS	As per the reference case
Type of economic evaluation	Cost utility analysis with fully incremental analysis	As per the reference case
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	As per the reference case
Synthesis of evidence on health effects	Based on systematic review	As per the reference case
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.	Health effects are expressed in QALYs. HRQoL was measured in the INBUILD trial using the EQ-5D.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	HRQoL was measured directly in patients in the INBUILD trial.
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	The UK cross-walk value set was used to value the EQ-5D HRQoL data collected in INBUILD
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	As per the reference case
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be	As per the reference case

Element of health technology assessment	Reference case	ERG comment on company's submission
	valued using the prices relevant to the NHS and PSS	
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	As per the reference case

Source: Information provided in the CS.[1]
EQ-5D = European Quality of Life-5 Dimensions; HRQoL = health-related quality of life; NHS = national health service; PSS = personal social services; QALY = quality-adjusted life year.

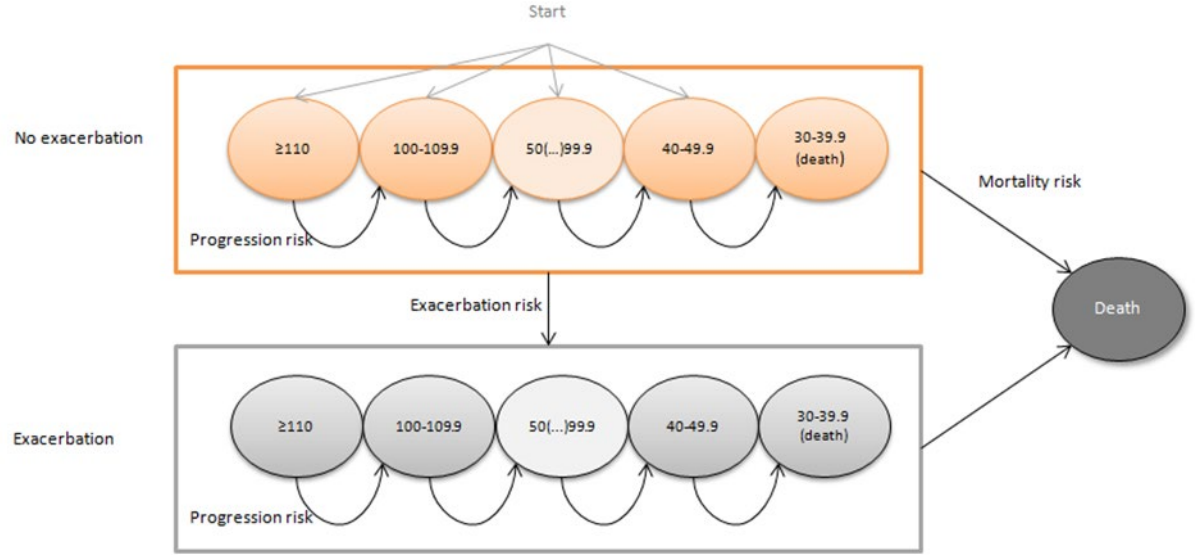
4.2.2 Model structure

The company developed a Markov model in Microsoft Excel and adopted the same model structure as for the nintedanib submission for IPF in TA379.[21] The company considered this appropriate given the equivalent disease trajectories for IPF and PF-ILD and because it was previously considered to be appropriate by the NICE committee and ERG in TA379.[21]

In preparation for their submission for TA379 in 2015,[21] the company performed a targeted review of the literature that identified no other relevant economic analyses within IPF and consulted with Irish clinicians who validated the model structure for IPF.[22] The model structure for PF-ILD was validated by UK clinicians in 2020.[23] For the development of the model for IPF in TA379, the company considered FVC percentage predicted (FVC%Pred) as the most appropriate outcome for incorporation in the Markov model as an indicator of disease progression. FVC is commonly used as a measure of disease status and as an endpoint in clinical trials in IPF and ILD, whilst FVC%Pred is considered as a better indicator of general disease status than FVC since it does not reflect patient heterogeneity in terms of body capacity, age, gender and height that are determinants of absolute FVC. Analogous to TA379,[21] FVC%Pred was also used to define the model health states in the current submission for PF-ILD. Also in line with TA379,[21] a 10-point categorisation of FVC%Pred was used to define the model health states in the current submission for PF-ILD.

In addition to lung function, acute exacerbations of ILD are dramatic, singular events that are often fatal and a major cause of mortality and morbidity in ILD. In line with the model for IPF in TA379,[21] the model structure for PF-ILD in the current submission was designed with health states that describe the patient condition as a combination of lung function, as indicated by FVC%Pred, and exacerbation. The structure of the model is shown in Figure 4.1.

Figure 4.1: Schematic representation of the model structure



Source: Figure 8 in the CS.[1]
Note: numbers in diagram relate to FVC%Pred.

The model structure is thus the same as the one used for IPF in TA379,[21] with its input parameter values updated to correspond to PF-ILD and to the application of nintedanib in this population.

The cohort of patients enters the model at different FVC%Pred health states without exacerbation. Patients can then either remain in the same health state or transition to one of the following other health states: health state with the same FVC%Pred with exacerbation, health state with 10-point lower FVC%Pred without exacerbation, health state with 10-point lower FVC%Pred with exacerbation, or Death. It is assumed that patients cannot transition to a health state with higher FVC%Pred. Similarly, it is assumed that following an exacerbation, patients cannot transition to a health state without exacerbation for the remainder of the time horizon. Transitions to Death can occur from any health state based on survival analysis of clinical trial data, or by reaching a level of FVC%Pred below 40% at which point it is assumed that the level of lung function is unsustainable. The latter was provided as an option in the model that was not used by the company.

The model uses a cycle length of three months, consistent with the clinical trial intervals between observations. The company considers this to be a balanced interval for model outcomes. The same cycle length was also used in TA379 and was considered as appropriate by the ERG of that appraisal.[21]

ERG comments: The company's description of the model provides two routes for patients to transition to Death; one is mortality based on OS, the other is the transition to an FVC%Pred lower than 40%, which the company assumed to be an unsustainable level of lung function. However, in the model, only the first of these two options were used. This implies that mortality is modelled as independent from lung function decline, even for patients with the lowest level of lung function which is assumed to be unsustainable. A similar independence between mortality and rate of acute exacerbations is also assumed in the model, despite the fact that the company report that acute exacerbations are often fatal and a major cause of mortality in ILD.[1] The ERG assumes that this decision was made to avoid double counting, as the overall survival (OS) data already includes all deaths, and obviously agrees that deaths should not be double counted. Therefore, no change was made to these assumptions in the model, but the ERG notes that this can produce implausible results in relation to discontinuation in the model as further discussed in section 4.2.6.5.

The ERG considers the other aspects of model structure appropriate given the similarities between IPF and PF-ILD, validation by UK clinicians, and the ERG and committee in TA379 having considered it appropriate.[21]

4.2.3 Population

Nintedanib has marketing authorisation for adults with chronic fibrosing ILD with a progressive phenotype, i.e. PF-ILD, based on the results of INBUILD. The model population was based on this trial and included patients within the marketing authorisation. The baseline characteristics of this patient population and the extent to which these match the characteristics of the relevant UK population are reported in Section 3.2.3.

4.2.4 Interventions and comparators

The intervention under investigation is continuous treatment with nintedanib oral capsules, in a dosage of 150 mg twice daily (i.e. 300 mg per day). In case of tolerability issues, the dosage can be reduced to 100 mg twice daily. The latter dosage is also recommended for patients with mild hepatic impairment (Child Pugh A).

The company considered that there are no relevant comparators for the treatment of adults with PF-ILD in the UK, therefore the model implements a comparison of nintedanib versus BSC. In the model, BSC was based on the placebo arm of INBUILD that the company considered as a close match to BSC for adults with PF-ILD in UK clinical practice.

ERG comments: The ERG cannot confirm that that there are no relevant comparators for the treatment of adults with PF-ILD in the UK, considering the consensus among the UK clinicians that were consulted by the company during the advisory board meeting of 11 November 2020 that there are other treatment options: steroids, immunosuppressants (i.e. both can be used as part of current best supportive care in clinical practice, but not in INBUILD; see below) and possibly off-license use of pirfenidone, especially when it goes off patent (class effect).[23]

As noted in Sections 2.3 and 3.6, the ERG has concerns regarding the representativeness of the placebo arm of the INBUILD trial for best supportive care. This is because patients in INBUILD were not permitted to receive immunomodulatory treatments (including azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids) at randomisation and during the first six months of the treatment period in INBUILD.

4.2.5 Perspective, time horizon and discounting

The model was constructed from the perspective of the NHS and Personal Social Services (PSS), in line with the NHS Reference case.[24] A lifetime horizon was adopted to capture all relevant costs and health-related utilities, with all costs and utilities discounted at a rate of 3.5% per year, in line with the NHS Reference case.[24]

4.2.6 Treatment effectiveness and extrapolation

The INBUILD trial was the main source of evidence for model parameters including: overall survival, time-to-first acute ILD exacerbation, loss of lung function, time-to-treatment discontinuation, utility values and healthcare resource use.[1] A 52-week analysis of INBUILD has previously been published, however a second database lock, taken approximately three months after the first lock was used to populate the parameters listed above as it provides longer follow-up.

The model requires evidence for three types of transitions related to treatment efficacy: mortality, acute ILD exacerbations and decline in lung function (based on FVC%Pred).

4.2.6.1 Overall survival

The mortality risk in the model is based on parametric extrapolation of OS data and is applied irrespective of health state or model events. OS extrapolation was undertaken using two different approaches: a standard frequentist approach with standard parametric distributions fitted independently to each arm and an exploratory Bayesian approach, undertaken with the aim of improving the accuracy and precision of the extrapolated OS estimates by estimating priors using available long-term data from other sources.

Goodness of fit was assessed using the Akaike information criterion and Bayesian information criterion (AIC and BIC), with models considered to be suitable candidates for inclusion in the economic model if they were within three points of the parametric model with the lowest AIC or BIC.[1] After excluding any models which did not meet this criteria, the results of the remaining parametric models were compared with evidence from the literature (visual inspection/face validity and comparison with published cohorts).

For the standard frequentist extrapolation approach, six parametric distributions were explored, as shown in Table 4.4. The exponential, lognormal and generalised gamma were considered to have a poor fit and were excluded. Therefore, the loglogistic, Gompertz and Weibull distributions were adopted for the frequentist approach.

Table 4.4: Goodness of fit frequentist OS

FVC%Pred Health state	Distribution	AIC	BIC	Decision
Placebo	Exponential	842.1154	845.9175	Excluded
	Weibull	822.3554	829.9597	
	Lognormal	825.7844	833.3886	Excluded
	Loglogistic	822.5821	830.1864	
	Gompertz	823.3835	830.9878	
	Generalised gamma	824.2238	835.6302	Excluded
Nintedanib	Exponential	690.9068	694.712	Excluded
	Weibull	687.0584	694.6687	
	Lognormal	690.5765	698.1868	Excluded
	Loglogistic	687.4335	695.0438	
	Gompertz	685.4074	693.0177	
	Generalised gamma	688.7022	700.1176	Excluded

Source: Table 25 of the CS.[1]
AIC = Akaike information criterion; BIC = Bayesian information criterion; FVC%Pred = forced vital capacity % predicted; OS = overall survival.

For the Bayesian OS analysis, additional data sources were required to generate informative priors.[1] The company used data from several trials conducted in IPF patients. The company stated that "*While IPF is the classic fibrosing ILD, PF-ILD patients demonstrate a number of similarities to IPF, with their disease being defined by the presence of progressive pulmonary fibrosis, worsening respiratory symptoms, declining lung function, resistance to immunomodulatory therapies and, ultimately, early mortality.*"[1] Given these similarities the company hypothesised that the trajectory of the survival of IPF patients could be used to inform survival estimates for PF-ILD patients.

Long term survival data were available from one phase 2 study (TOMORROW), two phase 3 IPF trials (INPULSIS I and INPULSIS II) and a combined long-term extension of these studies, known as INPULSIS-ON which monitored OS for more than eight years in IPF patients taking nintedanib.[25-27]

These IPF data were used to generate informative priors to inform the Bayesian survival analysis of the PF-ILD data. The IPF patients were matched to PF-ILD patients using propensity score matching to ensure that these patients had similar baseline characteristics. Survival data were then generated for the matched, weighted IPF patients.

Study linking and cleaning

The following data from the aforementioned trials were used:

- TOMORROW (phase II) study: patients receiving nintedanib (300mg) or placebo; patients from TOMORROW who did not receive the 300mg dose of nintedanib were excluded.[25]
- INPULSIS 1 and 2 (phase III studies): all patients.[26]
- INPULSIS-ON (open-label extension [OLE] from phase II and III studies): patients previously receiving nintedanib (300mg) who continue treatment; patients who were on placebo and then went on to receive nintedanib in the OLE were censored on initiation of nintedanib.[27]

These data were merged for the purpose of this analysis using the following censorship rules:

- Placebo patients were censored at the last contact date recorded in the phase II/III studies, or on the date they entered the OLE study, whichever happened first.
- Nintedanib patients who did not enter the OLE study were censored at the last contact date recorded in phase II/III.
- Nintedanib patients who entered the OLE study were censored at the last contact date recorded in the OLE.

A total of 1,239 IPF patients were included in this global dataset; 726 patients were treated with nintedanib and 513 with placebo. Data from the INBUILD trial were used in this analysis to incorporate PF-ILD patients. The INBUILD dataset contained 663 patients with PF-ILD; 332 patients were treated with nintedanib and 331 with placebo.

Propensity score matching

Patients from the IPF dataset were matched to PF-ILD patients from the INBUILD trial using propensity score matching, with the aim of ensuring that the IPF patients used to inform the Bayesian priors had similar baseline characteristics and disease severity to the PF-ILD patients.[1]

Baseline characteristics were assessed to determine which patient characteristics reported across the PF-ILD and IPF trials would be most relevant in the propensity score matching analysis. Baseline characteristics were assessed according to whether they were widely reported and clinically meaningful. The following baseline characteristics were used in the patient matching:

- Age
- Gender
- Race (coded in this analysis as Asian versus other)
- Time since IPF or PF-ILD diagnosis
- Percent predicted diffusing capacity for carbon monoxide (DLco) corrected for haemoglobin
- Percent predicted forced vital capacity (FVC) at baseline
- Smoking status (coded in this analysis as never smoked, used to smoke, currently smokes)

This selection of variables led to the upfront exclusion of nine PF-ILD patients with a missing baseline percent predicted DLco, and 140 IPF patients (129 had missing race, three missing baseline percent predicted DLco and eight had no baseline characteristics). The final analysis dataset therefore contained 654 PF-ILD patients (326 nintedanib patients and 328 placebo patients) and 1,099 IPF patients (640 nintedanib patients and 459 placebo patients).

Kernel and Radius matching algorithms with radii of 0.1 and 0.05 were considered. Balance was checked and the common support assumption was assessed after patients' propensity scores had been generated to determine whether there was overlap between the scores generated by the IPF and PF-ILD patients to enable matching.

The validity of the matching was assessed using common diagnostic statistics and plots.[1] The balance of covariates after the matching and weighting of control observations was checked by examining standardised differences and a summary of the mean and median bias across all covariates before and after matching, as well as Rubin's B (absolute standardised difference of the means of the linear index of the propensity scores between the two groups) and Rubin's R (ratio of the variances of the propensity score index in the two groups) indicators. Ideally, the bias (expressed as a percentage) should be below 5, Rubin's B less than 25 and Rubin's R between 0.5 and 2. The distribution of the propensity scores was also plotted. Separate analyses were conducted for each treatment arm.

Generating survival data

IPF patients who received nintedanib in both a clinical trial and (optionally) an open-label extension were of interest in this analysis. IPF patients who received placebo at the start of a clinical trial and then went on to receive nintedanib in an open-label extension were censored on initiation of the open-label extension when they started treatment with nintedanib. Overall survival was estimated as time from a patient's first baseline visit to the date of the last recorded visit. Patients were censored on their last visit if they had not been recorded as having died during the trial period. The survival analysis was performed using OpenBUGS (version 3.2.3 rev 1012).[28]

Generating informative priors

Standard frequentist survival models were fit to the matched, weighted IPF patient data using the "flexsurv" package in R (version 3.6.1).[29, 30] The three models with the lowest AIC and BIC (i.e. the best fitting models of the matched IPF data) were used to generate informative priors for the shape parameter of the Bayesian PF-ILD model. The best fitting model of the IPF data dictated the extrapolation models that were fit to the PF-ILD data.

The distribution of the shape parameter generated using the matched IPF data was used to inform the shape parameter of the PF-ILD model. Following the methodology outlined in Soikkeli 2019,[31] the Bayesian shape parameter prior was modelled using a gamma (α,β) distribution. A vague (noninformative) prior was used for the scale parameter throughout all analyses. Convergence was assessed, and a sufficient number of iterations for burn-in selected, for all analyses conducted in OpenBUGS. Autocorrelation was also evaluated and a thinning factor was applied when required.

OS estimates informing Bayesian priors

The AIC and BIC of the IPF survival models are presented in Table 4.5. Across the nintedanib and placebo cohorts, the Weibull, log-logistic and gamma distributions produced the lowest overall AICs and BICs. Given the small differences in fit between these models, all three were considered in Bayesian survival analysis. The exponential distribution produced the lowest BIC value for the nintedanib group but produced unrealistic long-term survival estimates for the placebo cohort and was therefore not considered further.

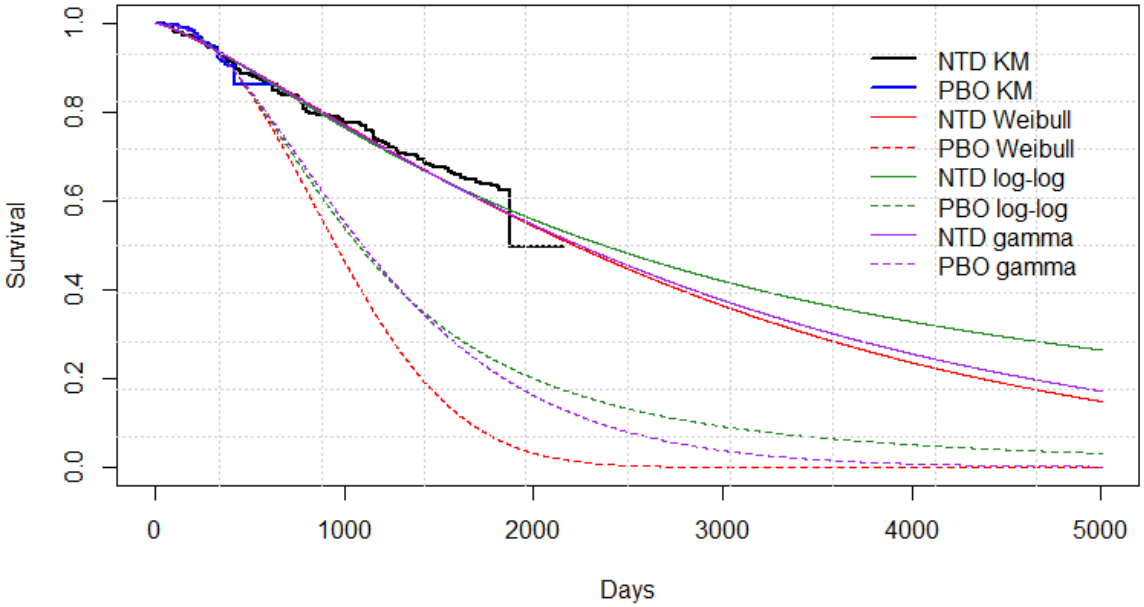
Table 4.5: AIC and BIC values for matched IPF survival models used to generate analysis prior

	Nintedanib		Placebo	
Distribution	**AIC**	**BIC**	**AIC**	**BIC**
Weibull	1468.961	1476.535	567.0736	574.6227
Exponential	1471.934	1475.721	580.1805	583.9613
Generalised gamma	1470.677	1482.037	569.1665	580.4714
Log-logistic	1469.346	1476.920	567.0456	574.5948
Log-normal	1470.437	1478.010	568.6821	576.2312
Gompertz	1470.285	1477.859	568.4749	576.0240
Gamma	1468.814	1476.388	567.2287	574.7778

Note: The three lowest AIC and BIC values are shaded in grey.
Source: Table 28 of the CS.[1]
AIC = Akaike information criterion; BIC = Bayesian information criterion; IPF = idiopathic pulmonary fibrosis.

The three survival models that produced the lowest overall AIC and BIC across the nintedanib and placebo cohorts were plotted against the corresponding Kaplan-Meier curves produced by the matched IPF data in Figure 4.2.

Figure 4.2: Matched IPF Kaplan-Meier curves for placebo and nintedanib plotted alongside the three best survival models



Source: Figure 12 of the CS.[1]
IPF = idiopathic pulmonary fibrosis; KM = Kaplan-Meier; log-log = log-logistic; NTD = nintedanib; PBO = placebo.

The three best fitting survival models of the matched IPF data were used to inform the shape parameter priors in the Bayesian analysis of the PF-ILD data for both nintedanib and placebo. For each IPF model, the same survival model was fit to the PF-ILD data. The results from fitting the gamma, log-logistic and Weibull models are described below. The standard frequentist results produced by modelling survival using the matched IPF data and the PF-ILD data (with no informative prior) were also plotted against the Bayesian survival analysis results for comparison.

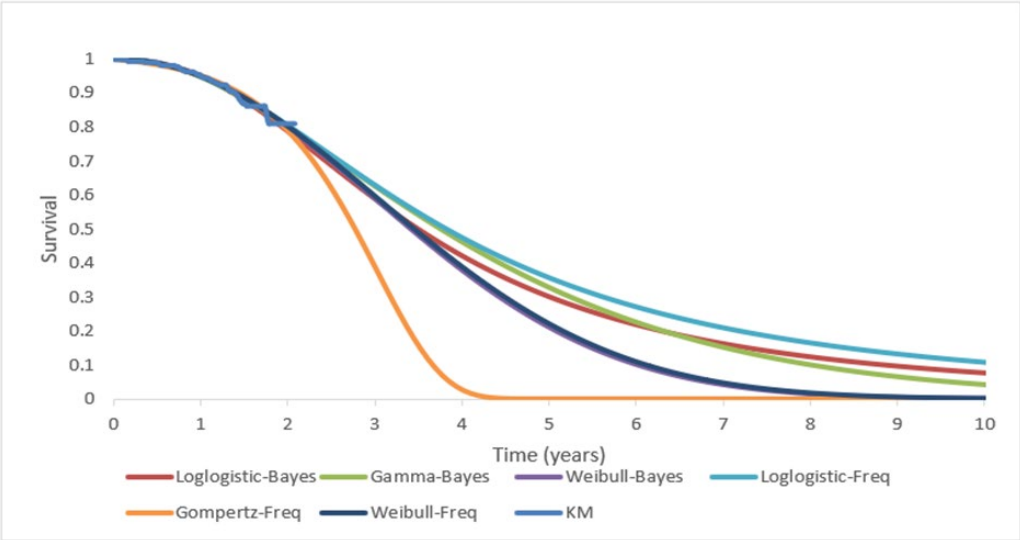
The company included three frequentist distributions (i.e. based on PF-ILD data alone) and three Bayesian survival curve distributions in the model. Figures 4.3 and 4.4 present all six distributions, and the Kaplan-Meier (KM) curves from the INBUILD trial, for placebo and nintedanib, respectively. The OS estimates produced by the three included Bayesian survival models are displayed in Table 4.6.

Table 4.6: OS estimates produced by Bayesian survival models

	Median OS (years)		Five-year survival (%)	
Distribution	**Nintedanib**	**Placebo**	**Nintedanib**	**Placebo**
Log-logistic	6.39	3.51	59	30
Gamma	6.50	3.76	60	32
Weibull	6.45	3.42	60	21

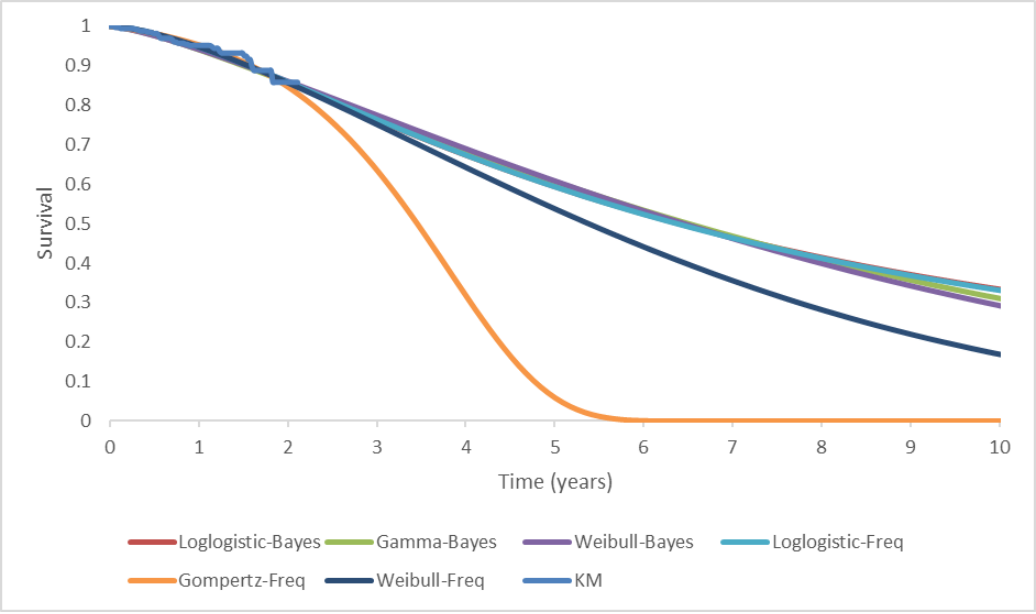
Source: Table 29 of the CS.[1]
OS = overall survival.

Figure 4.3: OS models fit versus INBUILD clinical trial KM – placebo arm



Source: Figure 16 of the CS.[1]
Bayes = Bayesian; Freq = frequentist; KM = Kaplan-Meier; OS = overall survival.

Figure 4.4: OS models fit versus clinical trial KM – nintedanib arm



Source: Figure 17 of the CS.[1]
Bayes = Bayesian; Freq = frequentist; KM = Kaplan-Meier; OS = overall survival.

External validation

Five clinical experts were approached to validate the assumptions within the model during a two-hour teleconference held on 11 November 2020. The advisory board was facilitated by company representatives and details of the attendees are available in Section B3.3 of the CS. During the teleconference, the clinical assumptions of the model were checked and discussed between the clinicians, with a particular focus on the long-term overall survival predictions of the model for PF-ILD patients.

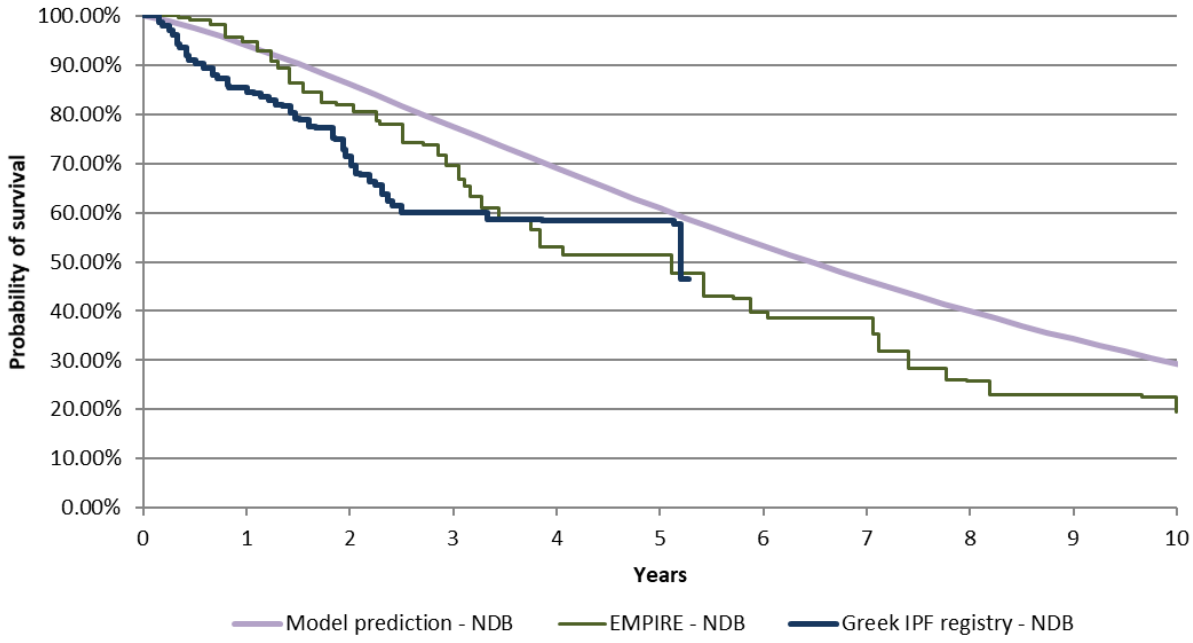
The clinicians were presented with the overall survival extrapolations presented in Figures 4.3 and 4.4 and were able to provide more commentary on the curves for BSC given the limited knowledge on the

long-term impact of nintedanib in the PF-ILD population. The clinicians agreed that for both curves the frequentist Gompertz curve was likely to underestimate survival as they would expect a proportion of patients to live beyond five years; these were therefore removed from further consideration. They also considered that both loglogistic curves appeared to overestimate survival as nearly all ILD patients with the progressive fibrosing phenotype would be dead by 10 years without any anti-fibrotic treatment. The clinicians agreed that either of the Weibull (frequentist or Bayesian) curves could be plausible for BSC.

When choosing between the Weibull curves, the company expected that the Bayesian analysis should provide more robust estimates of long-term survival, given the inclusion of longer-term IPF data to support to use of immature PF-ILD data. Therefore, the Bayesian Weibull curves were adopted for both nintedanib and BSC in the base-case.

The company used two sources of real-word data, both in IPF populations, in an attempt to validate the Weibull Bayesian curve for nintedanib, The EMPIRE study provides approximately 10 years of follow-up in 637 IPF patients taking nintedanib and a study by Antoniou et al, 2020 reports five-year survival data in 244 Greek IPF patients receiving nintedanib.[32, 33] The survival data from these studies were compared to the Weibull Bayesian extrapolation for nintedanib by the company in Figure 4.5 below. The company recognised that, in comparison to the EMPIRE study, the Weibull Bayesian extrapolation follows the KM curve for the first year or so, but then overpredicts survival and survival is consistently overpredicted by the extrapolation compared to the Greek IPF registry study.

Figure 4.5: OS models fit versus clinical trial KM – nintedanib arm

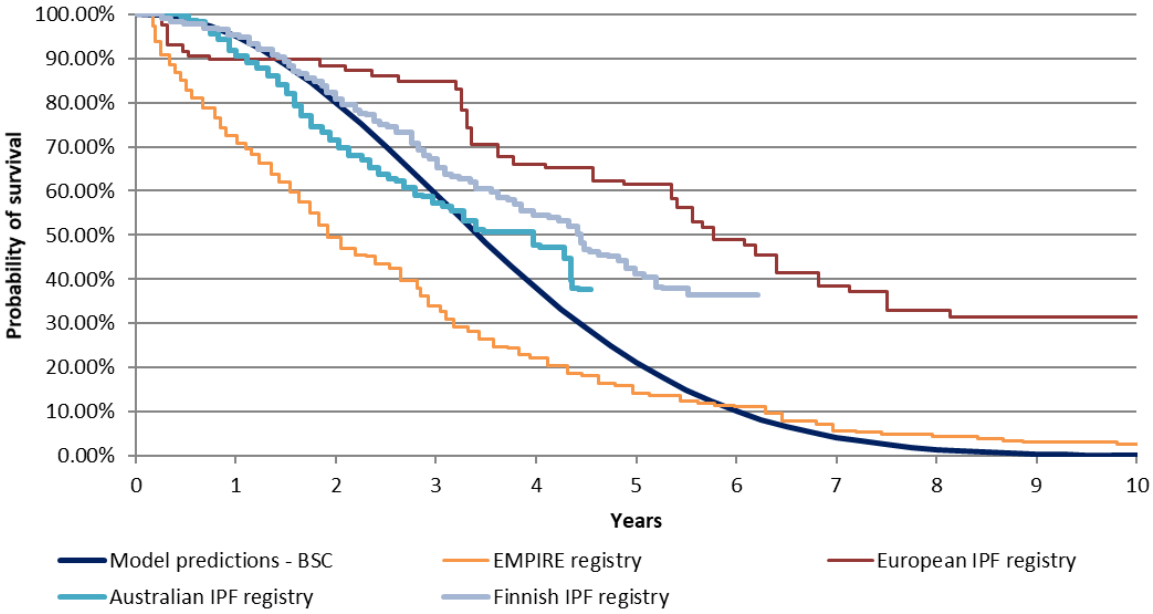


Source Figure 20 of the CS.[1]
IPF = idiopathic pulmonary fibrosis; NDB = nintedanib.

KM data from the treatment arms with no anti-fibrotic treatment in the EMPIRE study, Australian IPF registry, European IPF registry and Finnish IPF registry were used to validate the BSC survival extrapolations. Figure 4.6 shows a lack of consistency in survival between these sources. The clinicians considered the Australian registry most appropriate due to similarities between UK and Australian

clinical practice.[34] However it should be noted that as shown in Table 35 of the CS, patients in the INBUILD study were younger with lower FVC percentage than in the Australian study.

Figure 4.6: OS models fit versus clinical trial KM – BSC arm



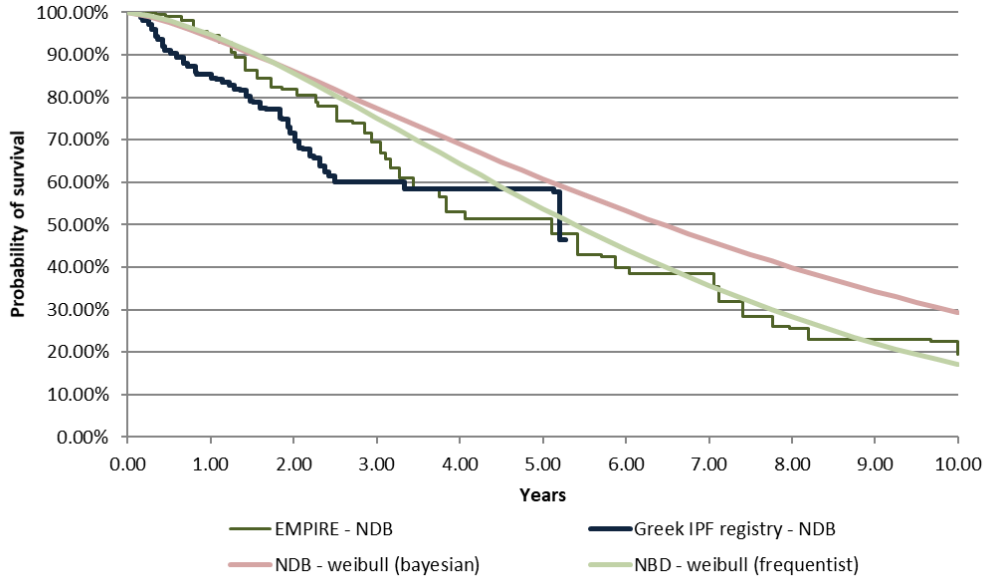
Source: Figure 21 of the CS.[1]
BSC = best supportive care; IPF = idiopathic pulmonary fibrosis; KM = Kaplan-Meier; OS = overall survival.

ERG comments: The ERG does not agree with the immediate exclusion of survival curves which may produce plausible long-term extrapolations due to arbitrary AIC and BIC difference cut-offs. Therefore, at clarification the ERG requested to see all extrapolations and have them included in the model for potential use. The company complied with this request.[3] Figures including all tested extrapolations can be seen in Figures 8 and 9 of the clarification response.[3]

The use of the Bayesian analysis adds uncertainty by requiring the use of propensity score matching and an assumption that IPL and PF-ILD patients have equivalent survival. It is not clear whether the benefits of having long-term data with which to generate priors and guide the extrapolations outweighs the additional uncertainty incorporated into the survival analysis when using the Bayesian method.

The clinicians consulted by the company to validate the survival curves considered that either of the Weibull curves (frequentist or Bayesian) could be plausible for BSC.[1] Given that the company's external validation in Figure 4.5 above shows that the Bayesian curve appears to overpredict survival compared to real-world data, the ERG requested that the company add the Weibull frequentist curve to this external validation figure, which resulted in Figure 4.7 below. This shows that the Weibull frequentist provides a better fit to the long-term KM data from the real-world data. Therefore, the ERG considers the Weibull frequentist curve more appropriate and hence has included it in their base-case.

Figure 4.7: Comparison of data on long-term survival with nintedanib in the IPF population (EMPIRE study and Greek IPF registry) versus the model predictions



Source Figure 11 of the Clarification Response.[3]
IPF = idiopathic pulmonary fibrosis; NDB = nintedanib.

4.2.6.2 Time to first acute exacerbation

Time to first acute exacerbation (TTFAE) was a secondary endpoint in the INBUILD trial. Standard parametric models were also considered to extrapolate TTFAE, resulting in AIC scores as shown below in Table 4.7.

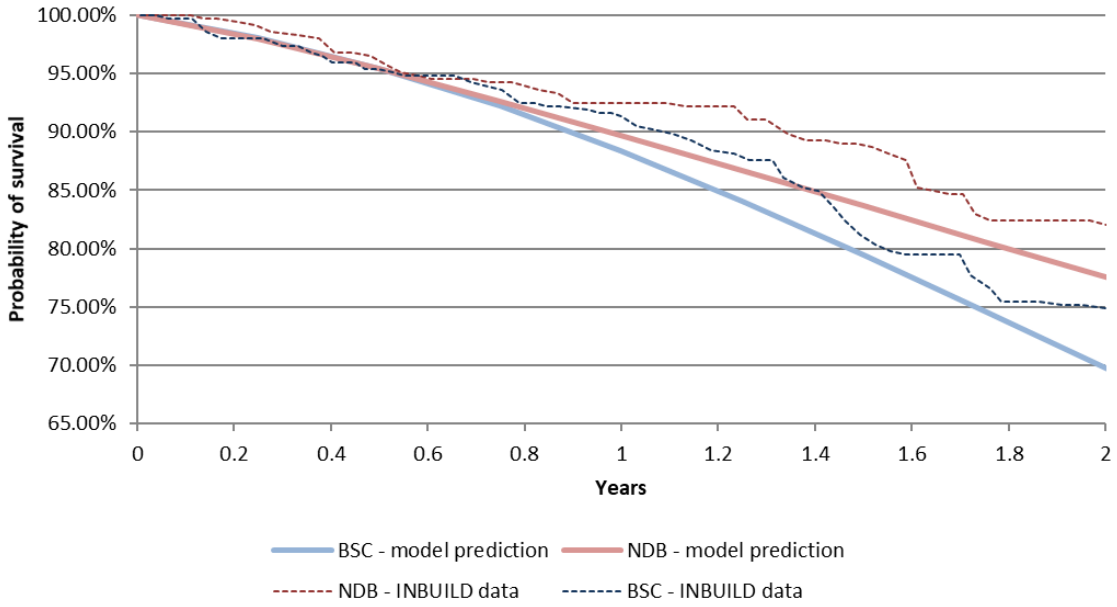
Table 4.7: Goodness of fit: time to first acute exacerbation

Treatment arm	Exponential	Generalised Gamma	Gompertz	Log logistic	Log normal	Weibull
Nintedanib	461.81	458.98	463.48	463.64	462.02	463.79
Placebo	670.14	673.82	672.14	672.15	671.82	672.11

Source: Table 36 of the CS[1]
Grey highlighted values represent the best fit

The exponential curve was associated with the lowest AIC score for the placebo arm and the second lowest for the nintedanib arm. Use of the exponential curve also facilitated a simpler modelling approach allowing the use of a fixed transition probability. Therefore, the exponential curve was used in the model. The coefficients for each arm are shown in Table 37 of the CS.[1] These coefficients resulted in a per-cycle risk of exacerbation of 1.76% and 1.12% for patients receiving BSC and nintedanib respectively. The company presented Figure 4.8 below, to demonstrate the fit of the exponential curves to the INBUILD KM data for TTFAE.

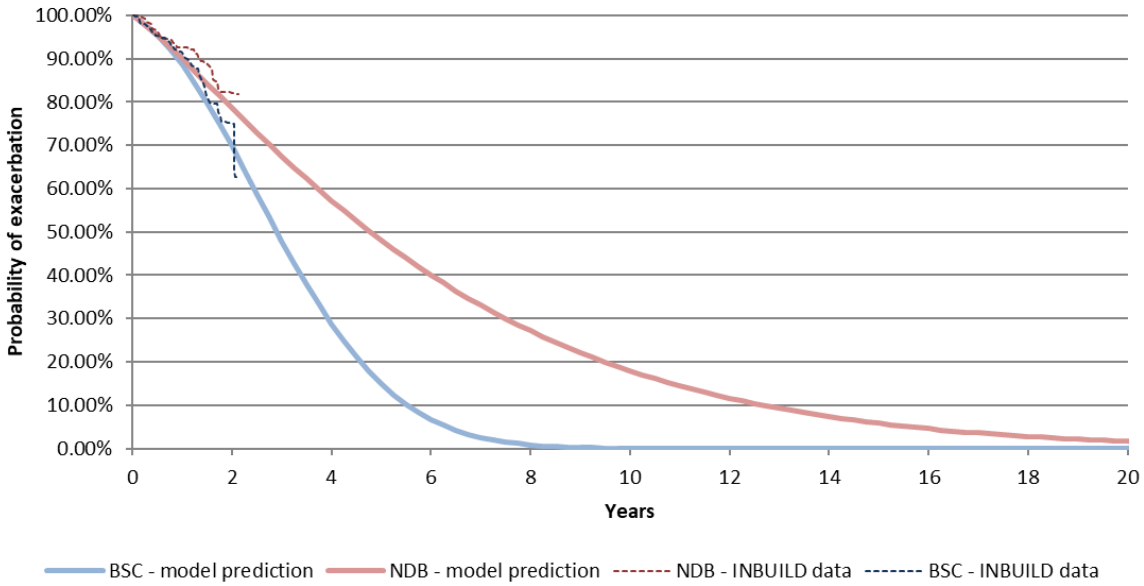
Figure 4.8: Exacerbation model fit vs. clinical trial Kaplan-Meier



Source: Figure 23 of the CS.[1]
BSC = best supportive care; NDB = nintedanib.

ERG comment: Figure 4.8 above suggests that the model is overpredicting the risk of acute exacerbation after approximately eight months, but the extrapolations beyond two years are not shown, so the long-term plausibility could not be examined. The ERG requested to see the long-term extrapolations and these were provided in the clarification response and are displayed below in Figure 4.9.

Figure 4.9: Amended exacerbation model fit vs. clinical trial Kaplan-Meier



Source: Figure 17 of the clarification response.[3]
BSC = best supportive care; NDB = nintedanib.

The ERG considered that this updated Figure provides quite a different view on the long-term difference between nintedanib and BSC which is modelled using these exponential extrapolations. The sharp drop

in the KM observed in BSC towards the end of follow-up, which is likely to be quite uncertain at the tail of the KM, has a substantial influence on the BSC extrapolation, substantially increasing the difference observed between the treatments.

The company did not include any other extrapolation options in the model or include an option for time varying risks of exacerbation which may better reflect the KM data. The company reported that they ran a scenario analysis where the rate of exacerbation with nintedanib was varied from 1.12% to 20% per cycle, which resulted in only a small increase to the incremental cost effectiveness ratio (ICER) of £3,000 per quality adjusted life year (QALY) and therefore exacerbations were not a driver of results. The ERG considered that this is likely due to the fact that mortality is not directly linked to the occurrence of acute exacerbation in the model. The ERG will explore scenarios regarding the assumed constant risk of exacerbation to explore the impact that this overprediction in both arms and the potential overestimation of the difference between arms has on results.

4.2.6.3 Recurrent exacerbations

The company base-case in the company submission assumed that patients could experience one acute exacerbation in the model. They reported that since the outlook of patients with an acute ILD exacerbation is generally very poor, this is probably a conservative assumption and the low overall frequency of exacerbations combined with the limited remaining lifetime of the patients in the model results in a very low risk for recurrent exacerbation.

ERG comment: At clarification, the ERG requested data on the occurrence of recurrent exacerbations in the INBUILD trial. The company responded that 1.5% and 1.2% of placebo and nintedanib patients experienced a recurrent exacerbation during the 52-week follow-up period of INBUILD, equating to 9/663 patients (1.36%) with a recurrent exacerbation overall. The breakdown of the number of exacerbations experienced per patient is shown in Table 4.8.[3]

Table 4.8: Exacerbations reported in the INBUILD trial up to 52 weeks

Number of exacerbation episodes	Nintedanib		Placebo	
0	311	93.7%	297	89.7%
1	17	5.1%	29	8.8%
2	1	0.3%	3	0.9%
3	3	0.9%	2	0.6%
>=4	0	0.0%	0	0.0%
Source: Table 4 of the clarification response.[3]				

The company added functionality to the model to allow the inclusion of recurrent exacerbations according to the rates of 1.5% and 1.2% for placebo and nintedanib respectively, converted to three-month probabilities. This had a limited impact of <£100 on the ICER and was included in their post-clarification base-case. The ERG agrees with the inclusion of the risk of recurrent exacerbation in the model. The ERG notes that the impact of recurrent exacerbation on patients in the model is limited to utility and costs but does not further increase the probability of loss of lung function beyond that of the first exacerbation.

4.2.6.4 Loss of lung function

Patients start the model in different FVC%Pred health states, according to the distribution of patients at baseline in the INBUILD trial, as shown in Table 4.9 below.[1, 3]

Table 4.9: Patient distribution at the start of the model

FVC%Pred Health state	Distribution (%)
110 and above	1.25%
100-109.9	1.88%
90-99.9	7.34%
80-89.9	13.59%
70-79.9	20.16%
60-69.9	25.00%
50-59.9	21.41%
40-49.9	9.38%
Source: Table 38 of the CS FVC%Pred = forced vital capacity % predicted	

Probabilities of decline in lung function per cycle for the BSC arm were estimated from the INBUILD data using a multivariate mixed effects logistic regression model including predictors of lung function decline.[1] This allowed for the analysis of recurrent events and the incorporation of additional covariates that could influence the probability of decline. Candidate predictors were:

- Age (continuous)
- Gender (male or female)
- Race (white, Asian, or other)
- Methotrexate use at baseline (yes or no)
- High-resolution computed tomography (HRCT) results (i.e. UIP-like pattern only, other fibrosis patterns)
- Underlying ILD diagnosis (e.g. autoimmune ILDs, hypersensitivity pneumonitis)
- Group criteria for progressive ILD [PGGR1] (i.e. clinically significant decline in FVC%Pred >=10%, marginal decline in FVC %Pred (>=5-<10%) combined with worsening of respiratory symptoms or increasing extent of fibrotic changes on chest imaging, worsening of respiratory symptoms and increasing extent of fibrotic changes on chest imaging only)
- FVC%Pred at the start of the time period (continuous)
- Exacerbation during the analysed three-month period (whether it occurred or not)

A p-value of 0.2 was used to determine which variables had a univariate association. The final model included the following variables: age, HRCT pattern, group criteria for progressive ILD, FVC at start of interval, and exacerbation variable. Further details of the model coefficients are available in Table 39 of the CS.[1]

The resulting three-monthly probabilities of progressing for each FVC%Pred category are shown in Table 4.10. Separate values are used for patients prior to and after an acute exacerbation as exacerbation was found to be a statistically significant predictor of lung function, with lung function decline expected

to occur more quickly after exacerbation and a diminishing effect in progression as lung function was lost observed.

Table 4.10: Three-month probabilities of progression, placebo (i.e. BSC)

FVC%Pred at start of interval	No exacerbation at start of interval	Intervals starting after first exacerbation
115	7.35%	41.14%
105	5.34%	33.19%
95	3.85%	26.10%
85	2.77%	20.07%
75	1.99%	15.14%
65	1.42%	11.26%
55	1.01%	8.27%
45	0.72%	6.02%

Source: Table 41 of the CS.[1]
BSC = best supportive care; FVC%Pred = forced vital capacity % predicted.

The risk of loss of lung function for nintedanib was informed by an odds ratio applied to the baseline placebo risk, assuming a constant relationship over time.[1] This odds ratio (shown in Table 4.11) was estimated using a mixed effect logistic regression of data from INBUILD, in which treatment was included as the only predictor. The company note that the 95% confidence interval for the odds ratio contains the value of 1 at the very upper limit of the interval, indicating that there is no statistically significant difference in effect between nintedanib and placebo at the 95% level. However, given this occurs at the highest end of the range it was judged appropriate to model a difference in lung function decline between nintedanib and placebo (or BSC) and explore this uncertainty further in a sensitivity analysis. The modelled three-month probabilities of progression for nintedanib patients are displayed in Table 4.12.

Table 4.11: OR values for loss of lung function

Fixed effects:	Estimate	SE	p-value	Odds ratio	95% CI
Intercept	0.654	0.2405	<0.01		
NDB coefficient	-0.4248	0.226	0.0602	0.654	0.420 – 1.1018

Source: Table 43 of the CS.[1]
CI = confidence interval; NDB = nintedanib; OR = odds ration; SE = standard error.

Table 4.12: Three-month probabilities of progression, nintedanib

FVC%Pred at start of interval	No exacerbation at start of interval	Intervals starting after first exacerbation
115	4.93%	31.37%
105	3.56%	24.52%
95	2.55%	18.76%
85	1.83%	14.10%

FVC%Pred at start of interval	No exacerbation at start of interval	Intervals starting after first exacerbation
75	1.31%	10.45%
65	0.93%	7.66%
55	0.66%	5.57%
45	0.47%	4.02%

Source: Table 41 of the CS.[1]
FVC%Pred = forced vital capacity % predicted.

ERG comment: It is not clear to the ERG why the impact of treatment on the probability of progression was not included in the full model used to estimate the probability of progression in BSC, but instead estimated in a separate model. The ERG requested this to be included in the full model at clarification. The company conducted the requested analysis, which resulted in the following probabilities of loss of lung function shown in Table 4.13 below. The ERG notes that these two different methods produce very different probabilities of loss of lung function after first exacerbation in both placebo and nintedanib patients. The company allowed for the use of these updated probabilities in the model, stating that this had a minimal impact on the ICER (<£20). The ERG was somewhat surprised that changes to the probability of progression had such a small impact on results, but this is likely due to the fact that while the absolute values differ substantially the relative differences between pre and post-exacerbation and between nintedanib and placebo do not differ substantially between the two models. The ERG also notes that in both methods the coefficient for treatment was not statistically significant, with confidence intervals crossing one.

From a methodological point of view the ERG would have preferred that the impact of treatment on the probability of progression was included in the full model, but given the minimal impact on the ICER, no change was made. It is worth noting that both methods assume a lifetime treatment effect while on nintedanib treatment.

Table 4.13: Three-month probabilities of progression (based on new regression output)

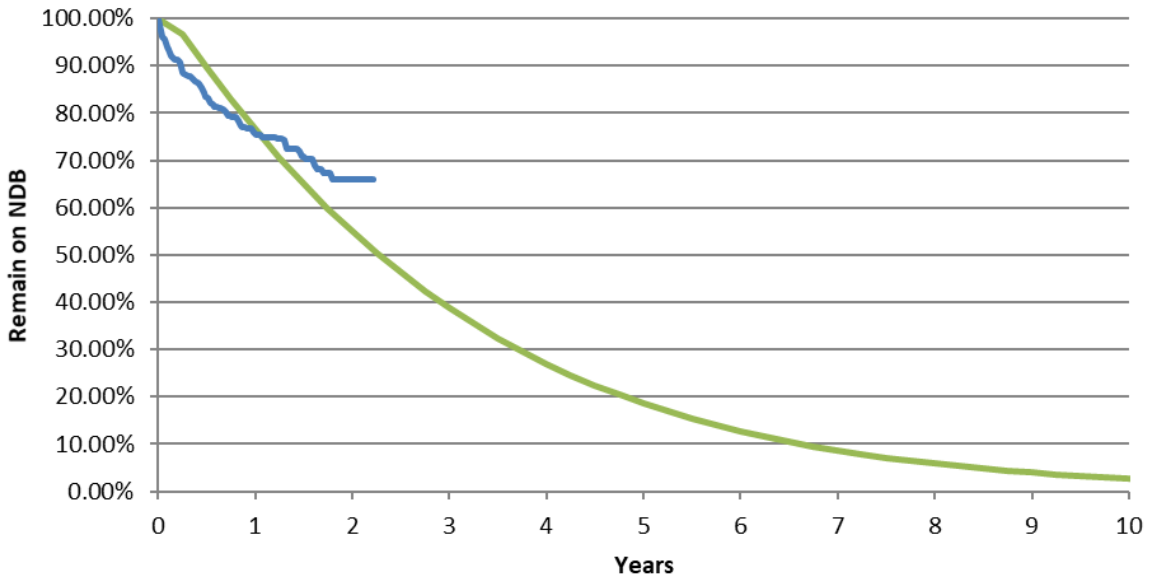
FVC%Pred at start of interval	Nintedanib		Placebo	
	No exacerbation at start of interval	Intervals starting after first exacerbation	No exacerbation at start of interval	Intervals starting after first exacerbation
115	5.57%	16.81%	8.26%	23.56%
105	4.29%	13.31%	6.41%	18.98%
95	3.30%	10.45%	4.94%	15.11%
85	2.53%	8.15%	3.80%	11.92%
75	1.93%	6.31%	2.92%	9.32%
65	1.47%	4.87%	2.23%	7.25%
55	1.12%	3.75%	1.71%	5.61%
45	0.86%	2.87%	1.30%	4.32%

Source: Tables 6 and 7 of the clarification response.[3]
FVC%Pred = forced vital capacity % predicted.

4.2.6.5 Treatment discontinuation

The company reported that up to DBL2, approximately 34% of patients had discontinued treatment in the nintedanib arm of the clinical trial.[1] Overall nintedanib discontinuation risk was estimated by extrapolating INBUILD discontinuation data using an exponential model, as it assumes a constant hazard and therefore a fixed discontinuation rate allowing for simple model implementation.[35] The company noted that this approach was also taken in TA379.[21] Discontinuation due to death was excluded from analysis. The coefficient for the exponential model was 7.270 (SD 1.737, 95% CI 7.083-7.457). This resulted in an overall discontinuation risk for nintedanib of 5.97% per month. The model predictions for time to discontinuation based on this risk, compared to available KM data from INBUILD, are presented in Figure 4.10.

Figure 4.10: Time on treatment with nintedanib



Source: Figure 27 in the CS.[1]
NDB = nintedanib.

This figure shows that the model underestimates discontinuation in the first year, but from approximately 15 months onwards the model appears to overestimate discontinuation. The company validated these predictions using data from Lancaster et al. 2019, which provides long-term data on the safety and efficacy of nintedanib in the IPF population.[36] Lancaster et al. 2019, reported that the median exposure to nintedanib, based on the long-term follow-up data from the nintedanib trials, was 22.5 months with a maximum exposure time of 93.1 months. The exponential model fitted to the INBUILD data predicts median survival of approximately 2.3 years (or 27-28 months), with a proportion of patients remaining on nintedanib after eight years (96 months), which was past the maximum exposure point measured by Lancaster et al. 2019. Therefore, the company acknowledged that the model may underestimate the true rate of discontinuation for nintedanib and conducted a scenario analysis in which a higher rate of discontinuation was applied to more closely match the data reported by Lancaster et al. 2019.

ERG comments: Given that the company's base-case exponential extrapolation of time to discontinuation does not appear to reflect the underlying KM data well, at clarification the ERG requested that the company consider alternative plausible extrapolations, or constant or time dependent discontinuation rates which better represent the INBUILD KM data, for possible use in the model.[37]

The company responded that the inclusion of alternative extrapolations or time dependent discontinuation rates would have required a more complicated and less transparent model structure and therefore these options were not included in the model.[3] Instead they conducted further sensitivity analyses using constant rates of discontinuation determined by the upper and lower bounds of the confidence interval from INBUILD (5.13% – 7.37%) as well as an alternative analysis where the exponential coefficient was varied until a curve was generated that was more consistent with that reported by Lancaster et al. 2019 and lastly a scenario which generated the long-term predictions to more closely match the tail of the INBULD KM curve.

The ERG noted a plausibility concern in the model regarding the impact of discontinuation on model results. When the discontinuation rate from nintedanib is increased in the company's model, the ICER decreases due to substantial treatment cost savings, with the optimal ICER observed when discontinuation is 100%. However, increasing the discontinuation rate had zero impact on life years in the nintedanib group and a minimal impact on QALYs (5.97% discontinuation = [redacted] vs. 100% discontinuation = [redacted]). This would imply that the optimal course of treatment according to the model, would be for all patients to take nintedanib for the first three months and then discontinue. The lack of difference in LYs is due to two modelling aspects: a) the company assumed that patients who had discontinued from nintedanib continued to be represented by the nintedanib survival analysis post-discontinuation, as most patients who discontinued treatment were included in the trial survival analysis; and b) exacerbation events were not directly linked to mortality in the model, meaning the increased risk of exacerbation events after discontinuation (when patients are assumed to have the same risk as BSC patients), does not translate into any difference in LYs. This results in a lifetime treatment effect in terms of OS in the model. Given that a high proportion of patients who discontinued nintedanib in the trial continued to be followed-up, the ERG consider that the OS is likely to reflect the weighted efficacy of patients on and off-treatment over the observed follow-up. However, the impact on efficacy in the longer-term remains uncertain as it is not clear whether the trial follow-up is sufficiently long to fully capture the impact of discontinuation on OS. It is important to note that the way discontinuation has been incorporated into the survival analysis makes it impossible to assess the impact of changes in the discontinuation rate on the ICER, as a new OS curve would be needed.

4.2.7 Adverse events

Data on the frequency of AEs were obtained from the INBUILD trial CSR.[18] The company included those AEs which:

- Had an incidence of >10% in either treatment arm
- Were treatment-related/treatment-emergent.
- Had an incidence at least 1.5 times higher in the treatment arm than in the control arm.

Based on these criteria the AEs shown in Table 4.14 were included in the model.

Table 4.14: Adverse events included in the model

	Nintedanib		Placebo	
AE	**N (%)**	**Risk per cycle**	**N (%)**	**Risk per cycle**
Patients	332 (100.0)	N/A	331 (100.0)	N/A
GI events				
Diarrhoea	196 (59.0)	20.05%	59 (17.8)	4.8%

	Nintedanib		Placebo	
AE	**N (%)**	**Risk per cycle**	**N (%)**	**Risk per cycle**
Nausea	79 (23.8)	6.59%	19 (5.7)	1.47%
Vomiting	41 (12.3)	3.25%	7 (2.1)	0.53%
Investigations				
Alanine aminotransferase increased	36 (10.8)	2.84%	8 (2.4)	0.61%

Source: Table 44 of the CS.[1]
AE = adverse events; GI = gastrointestinal.

ERG comment: At clarification, the ERG requested that the company provide an option in the model to include AEs with an incidence of > 5%, and AEs with an incidence of > 5% or 1.5 times greater than in the comparator arm and to justify their choice of a 10% cut-off. The company clarified that they had chosen an incidence cut-off of > 10% because adverse events of all severities were included and not just serious or severe adverse events. They did not provide a 5% incidence option in the model because no severe or serious adverse events occurred in greater than 5% of patients receiving nintedanib and therefore the overall impact on costs of extending the criteria from a 10% to 5% incidence was expected to be negligible.

4.2.8 Health-related quality of life

The literature review conducted to identify relevant health state utility values (HSUVs) did not identify any values specific to PF-ILD. Therefore the HRQoL data collected from the INBUILD trial was used to estimate HSUVs in the model.[1] EQ-5D HSUVs were estimated for each FVC%Pred health state. Acute exacerbation and AEs were included as utility decrements.[1]

In INBUILD, HRQoL was measured using the EQ-5D-5L on day 1 of treatment and then at weeks 12, 24, 36 and 52 of treatment as well as the end of treatment visit.[35] This HRQoL data was valued using the EQ-5D cross walk value set for the UK to obtain utility values. Table 4.15 shows the mean EQ-5D-5L utility used in the model for each FVC%Pred health state. The analysis only used data before exacerbations so that these events would not affect the HSUVs as the impact of exacerbations is considered separately. The analysis resulted in a lower estimated utility in patients with an FVC%Pred ≥110 than those patients in the 100-109.9 category (0.7028 vs 0.7521). This was considered implausible by two clinicians consulted by the company and given that the ≥110 estimate was based on only 10 patients, utility in the ≥110 category was assumed equal to utility in the 100-109.9 category in the model. It was assumed that the utility was 0 (dead) for FVC%Pred values < 40%.

Table 4.15: EQ-5D utility values used in the model by FVC%Pred group

FVC%Pred Health state	Mean EQ-5D utility	SD	Number of patients
≥110	0.7521	NA.	NA.
100-109.9	0.7521	0.2570	30
90-99.9	0.7287	0.2278	76

FVC%Pred Health state	Mean EQ-5D utility	SD	Number of patients
80-89.9	0.7333	0.2051	148
70-79.9	0.7242	0.2113	214
60-69.9	0.6750	0.2349	271
50-59.9	0.6453	0.2240	256
40-49.9	0.6045	0.2457	137

Source: Table 46 of the CS.[1]
EQ-5D = European Quality of Life-5 Dimensions; FVC%Pred = forced vital capacity % predicted; SD = standard deviation; NA = not applicable

When patients experience an acute exacerbation, this is associated with a utility decrement of 0.167 (SE = 0.050).[1] This decrement was estimated from regression analysis using the EQ-5D collected in the INBUILD trial. Reduction in utility due to acute exacerbation was assumed to last for one month and therefore this disutility was adjusted to 0.0556 per three month cycle, after which utility returned to the relevant FVC%Pred HSUV. The company report that the disutility value estimated from the INBUILD data was likely to be a conservative estimate because it is likely that the worst patients were missing not-at-random from the dataset (as they were unable or unwilling to attend the next study visit).

Disutilities for gastrointestinal (GI) event were based on estimates from TA379 based on the assumption that nintedanib has a similar safety profile regardless of the indication.[1, 21] Post hoc analysis of INPULSIS safety data showed that the EQ-5D change in patients that experienced a serious GI event was -0.068 (-0.201 to 0.065).[38, 39] The company assumed half of this value (-0.034) in this model for GI disutility in patients that experienced a non-serious GI event. The company validated this assumption against results from a phase III trial in recurrent non-small cell lung cancer which estimated a disutility for grade 3/4 diarrhoea of -0.042.[40] If 0.042 is a reasonable disutility for a serious diarrhoea, the company considered their assumed value of 0.034 for any GI event to be plausible. For alanine aminotransferase (ALT) increase the company assumed no disutility as this event is of mild to moderate severity and therefore considered asymptomatic.

ERG comments: The ERG was pleased to see base-case utility values based on EQ-5D trial data from INBUILD. The reversal in the trend that patients with lower FVC%Pred have lower utility for the 80-89 FVC%Pred category is not particularly plausible. Therefore, the ERG requested that the company make some adjustment to this value so that the trend remained consistent. The company responded that this was possible within the model structure but had a minimal impact on the ICER. To ensure that plausible values were used, the ERG incorporated a utility value of 0.7265 for the 80-89 FVC%Pred health state, which equates to a linear decline in utility from the 90-99 and 70-79 health states.

The company updated the model to allow for the age-adjustment of utilities during the clarification stage as the request of the ERG. This was done using UK population norms calculated by Kind et al, 1999.[41]

The ERG identified two other estimates for the impact of acute exacerbations in the first month in TA379.[21] These were estimated from EQ-5D data from the INPULSIS trial in IPF.

The validity of the assumed disutility for all GI events included in the model, estimated as half the value of serious GI events in TA379 is unclear. However, given the limited impact of AEs on model results

this is not a key issue. The disutility estimated from investigator ruled exacerbations was -0.14 in the first month, while the disutility estimated from adjudication committee ruled exacerbations was -0.274. These estimates will be explored as scenarios to examine the impact of the assumed disutility.

4.2.9 Resources and costs

The company included the following costs in the cost effectiveness analysis: drug acquisition costs for nintedanib, liver function test costs, health care resource use costs corresponding to each of the health states in the model, acute exacerbation costs, end of life costs, and costs in relation to adverse events.

4.2.9.1 Drug acquisition costs

The list price for nintedanib is £2,151.10 per pack of 60 capsules, for both the 100 mg and 150 mg formulations. The price that is used in the model includes a Patient Access Scheme (PAS) discount of [redacted] and is [redacted] per pack of 60 capsules. This amounts to a cost of [redacted] per capsule, or a cost of [redacted] per daily dose of two capsules of either 100 or 150 mg. Based on prescription records of nintedanib for IPF, the company assumed that 79% of patients receive the 150 mg formulation and 21% of patients receive the 100 mg formulation. Since the same price applies to both formulations, this has no implications for the calculation of drug acquisition costs. Administration costs are not applicable, because nintedanib is an oral treatment.

4.2.9.2 Liver function test costs

The nintedanib Summary of Product Characteristics states that hepatic transaminase and bilirubin levels should be investigated before treatment initiation and during the first month of treatment, and should be monitored at regular intervals thereafter.[42] The company assumed that all patients on active treatment would incur the cost of a liver function test at a quarterly frequency (i.e. once every three months). The cost per liver panel blood test was estimated at £2.79 (NHS Reference Costs 2018/19, Direct Access: Pathology Services: DAPS05 Haematology).[43]

4.2.9.3 Health state resource use costs

The company used individual patient data from INBUILD on the frequencies of use for the following health care resources: hospitalisations, emergency room (ER) visits, general practitioner visits, specialist visits, nurse visits, physiotherapy visits, occupational therapist visits, other visits, and use of oxygen. These data were grouped into the same 10-point FVC%Pred categories as used to define the model health states. Within each category, the number of observations corresponds to the number of patients multiplied by the number of months spent in that category. These numbers of observations and the three month probabilities of resource use are provided in Table 4.16.

Table 4.16: Three monthly probabilities of resource use for each FVC%Pred group

Health care resource	FVC%Pred group							
	≥110	100 - 109.9	90 - 99.9	80 - 89.9	70 - 79.9	60 - 69.9	50 - 59.9	40 - 49.9
Number of observations	124	274	599	1,215	1,958	2,566	2,386	1,497
Hospitalisation	0.12	0.05	0.05	0.05	0.05	0.09	0.09	0.14
Emergency room visit	0.10	0.04	0.02	0.04	0.03	0.05	0.06	0.05
GP visit	0.12	0.10	0.16	0.18	0.31	0.19	0.17	0.15
Specialist visit	0.07	0.14	0.25	0.21	0.26	0.23	0.18	0.17

Health care resource	FVC%Pred group							
	≥110	100 - 109.9	90 - 99.9	80 - 89.9	70 - 79.9	60 - 69.9	50 - 59.9	40 - 49.9
Number of observations	124	274	599	1,215	1,958	2,566	2,386	1,497
Nurse visit	0.02	0.01	0.03	0.02	0.03	0.03	0.02	0.02
Physiotherapy visit	0.01	0.01	0.01	0.02	0.03	0.00	0.00	0.00
Other visits	0.02	0.01	0.02	0.04	0.01	0.01	0.02	0.03
Occupational therapy visit	0.02	0.01	0.02	0.04	0.01	0.01	0.02	0.03
Oxygen use	0.14	0.16	0.17	0.15	0.27	0.33	0.47	0.57

Source: the electronic model from the CS / Figure 32, Tables 48, 51, 53 and 54 of the CS.[1]
CS = company submission; FVC%Pred = forced vital capacity % predicted; GP = general practitioner.

The estimated cost of hospitalisation was composed of the following: the average number of hospitalisations per patient with at least one hospitalisation (1.35, SE 0.22), the average duration of hospitalisation (10.74 days, SE 0.62), the proportion of hospitalisations associated with an ICU stay (5.1%, SE 1.1%), the proportion of hospitalisations associated with mechanical ventilation use (2.1%, SE 0.8%), the proportion of hospitalisations associated with an ER overnight stay (7.8%, SE 1.4%), and the proportion of hospitalisation associated with ambulance use (18.5%, SE 2.0%). The company considered the number of observations for each of these components too low for an analysis by FVC%Pred group, therefore the averages for each component over all groups was used to calculate the cost per hospitalisation that was applied to all groups. The unit costs and average values for each component of the hospitalisation cost estimate as well as the total cost estimate per hospitalisation are provided in Table 4.17.

Table 4.17: Hospitalisation cost estimate

Health care resource	Unit cost		Number of visits (per patient)	
	Value	Source	Average value (SE)	Source
Hospitalisation	£324	National Schedule of Reference Costs - Year 2017-18 - NHS trusts and NHS foundation trusts; Weighted average of DZ27S, DZ27T and DZ27U (Respiratory Failure without Intubation with CC score 11+, 6-10, 0-5 respectively. Inflated to 2018/2019 price year. Excess bed days are not reported within 2018/2019 NHS reference costs.[44]	Number of visits: 1.35 (0.22); Duration 10.74 days (0.62)	INBUILD trial post hoc analysis [35]
ICU stay	£1,073	Weighted average of XC06Z (Adult Critical Care, 1 organ supported) and XC07Z (Adult Critical Care, 0 organs supported), Adult Critical Care Unit National Schedule of Reference Costs Year 2018/19 - NHS trusts and NHS foundation trusts; Critical Care.[43]	5.1% (1.1%)	

Health care resource	Unit cost		Number of visits (per patient)	
	Value	Source	Average value (SE)	Source
Mechanical ventilation	£1,735	Non-Invasive Ventilation Support Assessment, 19 years and over, Non-Elective Long Stay, DZ37A; NHS Reference Costs 2018/2019.[43]	2.1% (0.8%)	
ER overnight stay	£268	Weighted average across all types (admitted only). Excludes patients that are dead on arrival, dental services and patients with no treatment/investigations. National Schedule of Reference Costs Year 2018/19 - NHS trusts and NHS foundation trusts; Accident and Emergency Services.[43]	7.8% (1.4%)	
Ambulance use	£224	Weighted average of ASH1 (hear and treat or refer), ASS01 (see and treat or refer), ASS02 (see and treat and convey); National Schedule of Reference Costs Year 2018/19 - All NHS trust and NHS foundation trusts - ambulance services.[43]	18.5% (2.0%)	
Total cost per hospitalisation	£4,815			

Based on Table 49 in the CS.[1]
CS = company submission; ER = emergency room; ICU = intensive care unit; NHS = national health service.

The estimated cost of an emergency room visit was composed of the average number of emergency room visits (1.21, SE 0.113), and the proportion of emergency room visits associated with ambulance use (19.4%, SE 2.724%). The unit costs and average values for each component of the emergency room visit cost estimate as well as the total cost estimate per emergency room visit are provided in Table 4.18.

Table 4.18: Emergency room visit cost estimate

Health care resource	Unit cost		Number of visits (per patient)	
	Value	Source	Average value (SE)	Source
ER visit	£182.85	Weighted average across all types. Excludes patients that are dead on arrival, dental services and patients with no treatment/investigations; National Schedule of Reference Costs - Year 2018/19.[43]	1.21 (0.113)	INBUILD trial post hoc analysis[35]
Ambulance use	£224.39	Same as hospitalisation, Table 4.17.[43]	19.4% (2.724)	

Health care resource	Unit cost		Number of visits (per patient)	
	Value	Source	Average value (SE)	Source
Total cost per ER visit	£264			
Based on Table 50 in the CS.[1] CS = company submission; ER = emergency room.				

For general practitioner visits, specialist visits, nurse visits, physiotherapy visits, occupational therapist visits and other visits, the unit costs and average number of visits per patient are provided in Table 4.19.

Table 4.19: Outpatient visits unit costs and average number of visits

Health care resource	Unit cost		Number of visits (per patient)	
	Value	Source	Average value (SE)	Source
GP	£39 per visit	PSSRU 2019[45]	1.497 (0.507)	INBUILD trial post hoc analysis[35]
Specialist	£158.02	Consultant led, weighted average between respiratory physiology and respiratory medicine (codes 340 and 341)[43]	1.613 (0.344)	
Nurse	£124.37	Non-consultant led, weighted average between respiratory physiology and respiratory medicine (codes 340 and 341)[43]	0.181 (0.051)	
Physiotherapist	£57.66	Physiotherapy, weighted average between consultant led and non-consultant led (code 650)[43]	0.068 (0.088)	
Occupational therapy	£70.96	Occupational therapy, weighted average between consultant led and non-consultant led (code 651)[43]	0.133 (0.105)	
Other visits	£158.02	Assumed to be the same as a specialist visit.	0.133 (0.105)	
Source: Table 52 in the CS,[1] and Table 12 in the response to clarification questions.[3] CS = company submission; GP = general practitioner.				

The analysis also included the costs of supportive long-term oxygen supplementation in case of resting hypoxemia. The cost of oxygen supplementation was estimated at £0.21 per hour, based on a £1,600 annual cost (sourced from the UK National Guideline on diagnosis and management of suspected IPF,[46] which was based on NHS Reference Costs 2010/2011[47] and inflated to 2018/2019 costs). The average hours of oxygen use per day and days of oxygen use (per patient) were 12.86 (SE 1.25) and 51.21 (SE 3.89), respectively.[35]

4.2.9.4 Acute exacerbation costs

The unit cost associated with each acute exacerbation was estimated using patient-level data from patients with IPF in INPULSIS who experienced an exacerbation, based on the three month probabilities of visiting the hospital (63.49%, which was combined with an average number of 1.3 hospitalisations and an average duration of 16.3 days), visiting an emergency room (7.49%), visiting a general practitioner (7.94%, which was combined with an average number of 1.59 visits), and visiting a specialist (15.87%, which was combined with an average number of 1.3 visits).[21, 48] The resulting estimate of £4,134 (2012/2013 cost year) was also used in TA379 and Rinciog et al, 2017, [21, 48] and inflated to 2018/2019 it was £4,424 using the NHSCII from PSSRU 2019.[45]

4.2.9.5 End of life costs

The company included end of life costs in the analysis, which were sourced from Georghiou and Bardsley, 2014 and consisted of the costs of secondary (acute) hospital care, local authority-funded social care, district nursing, and GP contacts that were based on patients without a cancer diagnosis.[49] Since the original estimate was largely based on costs from the cost year 2010, the end of life cost estimate was inflated to 2018/2019 values. This resulted in a cost estimate for end of life costs of £6,045.

4.2.9.6 Adverse event costs

The company assumed that all adverse events were resolved without treatment other than a visit to the general practitioner. A unit cost of £39 was sourced from PSSRU 2019 for this, referring to a per patient contact visit lasting 9.22 minutes.[45] The company also noted (in Section B.3.5 'Cost and healthcare resource use identification, measurement and valuation') that the frequencies of patients with adverse events related to increased hepatic enzymes were about four times higher in the nintedanib group (22.6%) than in the placebo group (5.7%).[18] This was not reported in Section B.2.10 'Adverse reactions'.

ERG comments: The ERG considers the health care resource use and costs that were included in the analysis as appropriate. The same approach was used in TA379 and deemed appropriate by the ERG of that appraisal. The CS did not state which source was used to inflate costs from previous cost years, but the ERG can confirm that the inflated costs were in line with those when applying the NHS Cost Inflation Index values from PSSRU 2019.[45]

5. COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company's post-clarification base-case deterministic cost effectiveness results are presented in Table 5.1. The total costs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with incremental costs of £[redacted] associated with nintedanib. The total QALYs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an incremental cost effectiveness ratio (ICER) of [redacted] per QALY gained.

Table 5.1: Company post-clarification base-case deterministic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Nintedanib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
BSC	[redacted]	[redacted]	[redacted]				

Source Post-clarification company's base-case results provided in response to additional ERG requests on March 2nd 2021.[3]
BSC = best supportive care; ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY(s) = quality adjusted life year(s).

For consistency with the company's sensitivity analyses results as reported in the original CS that are reported in the next section, the company's original submission deterministic cost effectiveness results are reported in Table 5.2 below as well.

Table 5.2: Company original submission deterministic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Nintedanib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
BSC	[redacted]	[redacted]	[redacted]				

Source: Table 58 in the CS.[1]
BSC = best supportive care; ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY(s) = quality adjusted life year(s).

5.2 Company's sensitivity analyses

The results of the company's sensitivity analyses based on the post-clarification version of the model were not provided to the ERG. The ERG could also not reproduce these results using the post-clarification version of the model, due to an issue that became apparent from the results of the probabilistic sensitivity analysis (PSA) and an issue with the functionality of the one-way sensitivity analyses (OWSA) in the post-clarification model. Therefore, the ERG reports below the results of the sensitivity analyses that were provided by the company in their original (i.e. pre-clarification) CS.[1]

5.2.1 Probabilistic sensitivity analysis

A PSA with 1,000 iterations was performed to assess the sensitivity of the cost effectiveness results to the uncertainty associated with model input parameters. Random samples were drawn simultaneously

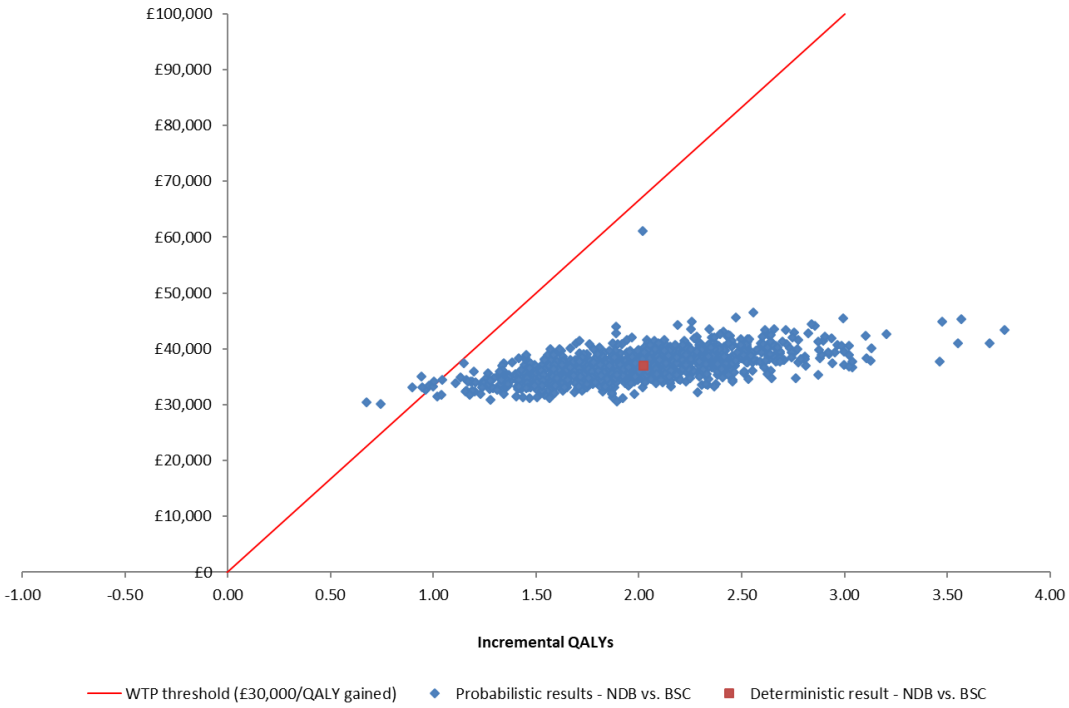
from the probability distributions that were assumed for each input parameter, which are detailed in Table 59 in the CS.[1]The company's PSA results are presented in Table 5.3. The total costs for the nintedanib and BSC arms were £[redacted] and £[redacted] respectively, with incremental costs of £[redacted] associated with nintedanib. The total QALYs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an ICER of £[redacted] per QALY gained. The probability that nintedanib is cost effective in comparison to BSC is 66% and 98% at cost effectiveness thresholds of £20,000 and £30,000 per QALY gained, respectively. The cost effectiveness plane and cost effectiveness acceptability curves are shown in Figures 5.1 and 5.2, respectively.

Table 5.3: Company first submission probabilistic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Nintedanib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
BSC	[redacted]	[redacted]	[redacted]				

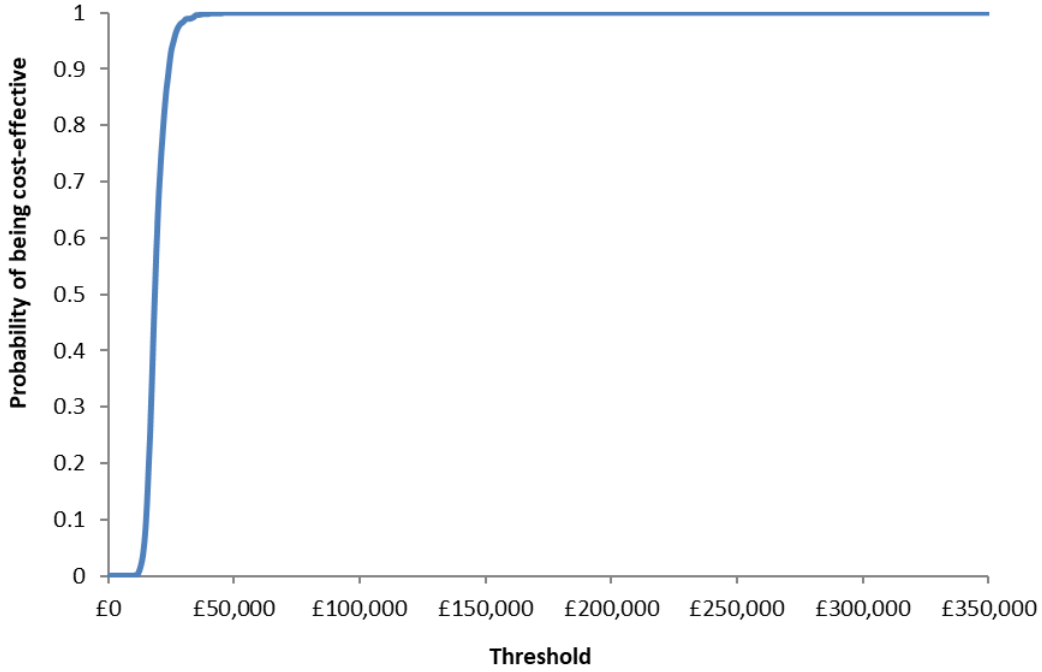
Source: Table 62 in the CS.[1]
BSC = best supportive care; ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY(s) = quality adjusted life year(s).

Figure 5.1: Cost effectiveness plane



Source: Figure 35 in the CS.[1]
BSC = best supportive care; NDB = nintedanib; PSA = probabilistic sensitivity analysis; QALY = quality-adjusted life year; WTP = willingness to pay.

Figure 5.2: Cost effectiveness acceptability curve

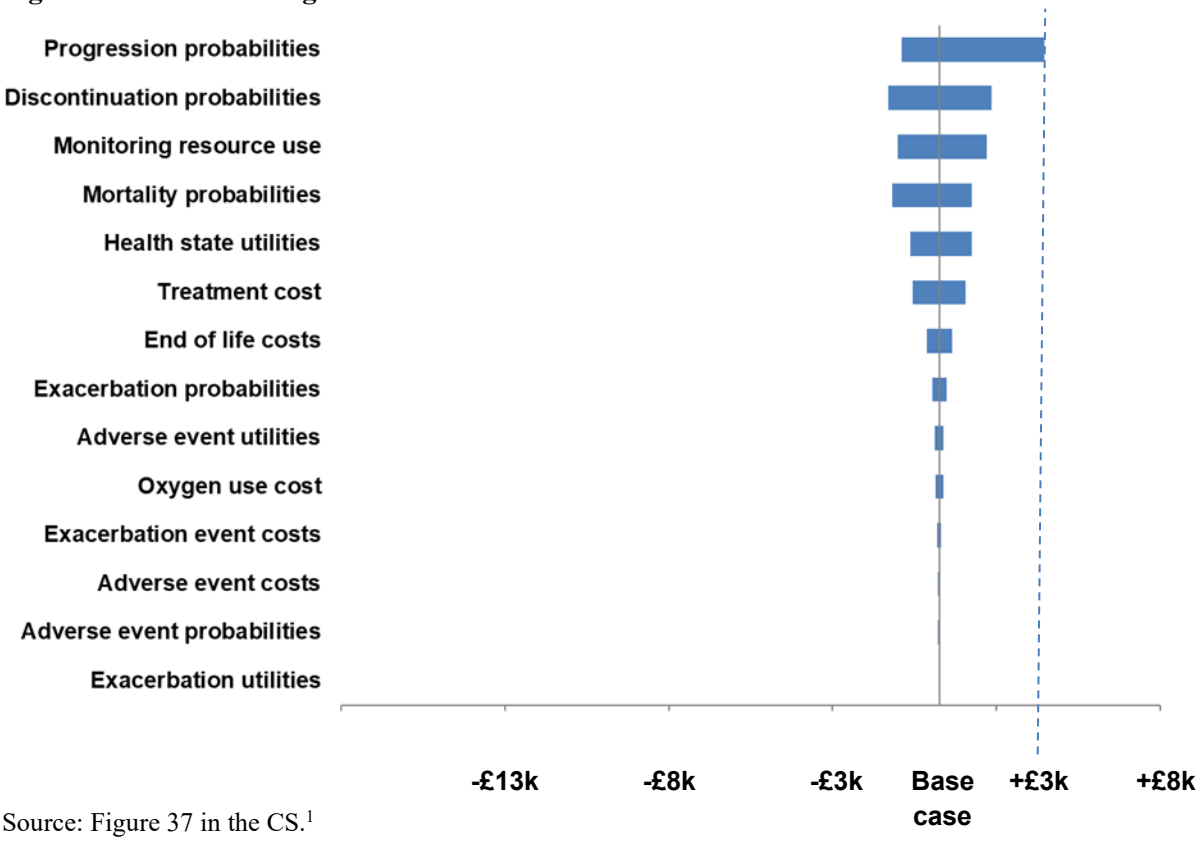


Source: Figure 36 in the CS.[1]

5.2.2 Deterministic sensitivity analysis

The company performed a deterministic, one-way sensitivity analysis (OWSA) to assess the impact of varying each parameter independently at both the upper and lower bounds of the 95% confidence interval that surrounds its mean estimate. The results of the OWSA are shown in Figure 5.3. Varying the progression probabilities caused the most substantial impact on the ICER, increasing it by approximately £3,000 per QALY gained to approximately [redacted] when varied to the highest confidence interval. The discontinuation and mortality probabilities, resource use associated with patient monitoring and health state utilities also cause some variation in the model results. None of the variations in inputs caused the ICER to increase to values higher than £30,000 per QALY gained.

Figure 5.3: Tornado diagram



Source: Figure 37 in the CS.[1]

5.2.3 Scenario analysis

The company performed a series of scenario analyses to assess the impact of alternative parameter inputs and assumptions on the cost effectiveness results. Three sets of scenarios were explored, relating to 1) alternative parametric distributions for OS extrapolations, 2) alternative utility inputs, and 3) alternative discontinuation rates.

For the first set of scenario analyses (i.e. scenarios 1 – 5), alternative parametric distributions were used for the extrapolation of OS. Specifically, the company replaced the Bayesian Weibull OS curves that were used in the base-case for both nintedanib and BSC with Bayesian Gamma OS curves in Scenario 1, Bayesian loglogistic OS curves in Scenario 2, frequentist Weibull OS curves in Scenario 3, frequentist loglogistic OS curves in Scenario 4, and frequentist Gompertz OS curves in Scenario 5.

For Scenario 6, the company replaced the utility values from INBUILD that were used in the base-case with utility values from patients with IPF in INPULSIS. These values were higher for all health states and are shown in Table 5.4.

Table 5.4: Alternative utility values used in scenario 6

FVC%Pred	Utility value	SD
≥110	0.8380	0.1782
100-109.9	0.8380	0.1782
90-99.9	0.8380	0.1782
80-89.9	0.8105	0.2051
70-79.9	0.7800	0.2244
60-69.9	0.7657	0.2380

FVC%Pred	Utility value	SD
50-59.9	0.7387	0.2317
40-49.9	0.6634	0.2552

Source: Table 64 in the CS.[1]
FVC%Pred = forced vital capacity percentage predicted; SD = standard deviation.

For the third set of scenario analyses (i.e. scenarios 7 and 8), the company replaced the discontinuation rate of 5.97% per cycle that was used in the base-case with a discontinuation rate of 7.67% per cycle to match the median time on treatment from the study by Lancaster et al, 2019 in Scenario 7,[36] and with a discontinuation rate of 3.97% in Scenario 8. The latter was considered by the company to provide a better fit to the tail of the INBUILD KM curve (i.e. only the last few months of available data), while noting that it did not fit the first two years of those data well.

The results of the scenario analyses are provided in Table 5.5.

Table 5.5: Results of the company's scenario analyses

Scenario #	Description	Incremental costs	Incremental QALYs	ICER
1	Bayesian gamma OS curves	[redacted]	[redacted]	<£25,000 [redacted]
2	Bayesian loglogistic OS curves	[redacted]	[redacted]	<£20,000 [redacted]
3	Frequentist Weibull OS curves	[redacted]	[redacted]	<£30,000 [redacted]
4	Frequentist loglogistic OS curves	[redacted]	[redacted]	<£20,000 [redacted]
5	Frequentist Gompertz OS curves	[redacted]	[redacted]	>£30,000 [redacted]
6	Alternative utility values	[redacted]	[redacted]	<£20,000 [redacted]
7	Discontinuation to match Lancaster et al, 2019[36]	[redacted]	[redacted]	<£20,000 [redacted]
8	Discontinuation to match tail of INBUILD KM data	[redacted]	[redacted]	<£25,000 [redacted]

Source: Table 65 in the CS.[1]
ICER = incremental cost effectiveness ratio; KM = Kaplan-Meier; OS = overall survival; QALYs = quality adjusted life years.

Nintedanib was associated with higher incremental costs and incremental QALYs than BSC in all of the scenarios considered. Scenarios two, six and seven resulted in a reduction in the ICER compared to the company's base case results. Nintedanib is not cost effective compared to BSC when the frequentist Gompertz OS curves are used, as this scenario produced an ICER >£30,000 ([redacted]) per QALY gained. However, based on clinician input the results from using the frequentist Gompertz OS curves were considered as implausible since they resulted in overly pessimistic extrapolations for both treatment arms.

5.3 Model validation and face validity check

5.3.1 Face validity assessment

The face validity of the model was examined during the UK Advisory Board.[1] This was achieved by describing the model structure and inputs to UK clinical experts to ensure the suggested approach appropriately captured costs and outcomes for UK clinical practice. Specific revisions were made to the model upon the advice received.

As described in Section 4.2.6.1, five clinical experts were asked to validate the model assumptions during a teleconference held on 11 November 2020. The company stated that clinicians validated the overall survival extrapolations and agreed that the Weibull Bayesian may be the most appropriate choice for both treatment arms. The overall survival curves were also compared with relevant data identified in the wider literature.

Due to a lack of previous economic models in this indication, it was not possible to examine the external validity of the model by comparing the results.

5.3.2 Technical verification

The company examined the internal validity of the model via a two-step process. First, they performed a cell-by-cell check of all model formulae to ensure they were both correct and appropriately applied. Second, a model verification checklist including a range of tests and sense checks, for instance, changing certain inputs to zero and checking that the observed effect was as expected (i.e. illogical results were not generated) was used. This internal validation process was undertaken by a health economist who was not directly involved in the conceptualisation and development of the model.

5.3.3 Comparisons with other technology appraisals

The company stated that due to a lack of previous economic models in this indication, it was not possible to examine the external validity of the model by comparing the results.

5.3.4 Comparison with external data

Extrapolations were compared with external data for OS and discontinuation as described in Sections 4.2.6.1. and 4.2.6.5.

ERG comments: The company report that clinicians validated the overall survival extrapolations and agreed that the Weibull Bayesian may be the most appropriate choice for both treatment arms, but in fact clinicians could not choose between the two Weibull options. The company stated that the model was sense checked during technical verification, but this did not pick up the implausible relationship between discontinuation, the ICER and LYs.

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the ERG

6.1.1 Explanation of the company adjustments after the request for clarification

In response to the clarification letter, the company supplied an updated version of the model with the following changes:

- The company updated/corrected several costs at the request of the ERG during clarification, including the cost of mechanical ventilation, cost of outpatient visits and cost of acute exacerbation in response to clarification questions B24, B27 and B28.[3]
- Recurrent exacerbations were included in the model and in the company base-case
- The company incorporated age-adjustment of utilities into the model and included these in their base-case.
- The baseline distribution of patients, baseline age and AE incidences were included in the PSA.

6.1.2 Explanation of the ERG adjustments

Based on these model updates and all considerations in the preceding Sections of this ERG report, the ERG defined an alternative base-case. The ERG base-case included the above changes made during the clarification stage. Further adjustments made by the ERG were subdivided into three categories (derived from Kaltenthaler 2016)[50]:

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model where the ERG considers that reasonable alternative assumptions are preferred)

Adjustments made by the ERG, to derive the ERG base-case (using the post-clarification base-case as starting point) are listed below.

6.1.2.1 Fixing errors

After clarification no further errors were identified.

6.1.2.2 Fixing violations

After clarification no further violations were identified.

6.1.2.3 Matters of judgement

1. Extrapolation of OS (Key Issue 3, Section 4.2.6.1)
   The ERG preferred to extrapolate OS using the frequentist Weibull curve, given that it appeared to fit long-term nintedanib IPF survival data used for external validation better than the company's preferred Bayesian Weibull curve and clinicians considered both curves plausible.

2. Adjustment of HSUV for 80-89 FVC%Pred health state (Section 4.2.8)
   The ERG adjusted this value (assuming a linear decline between the neighbouring categories) to maintain the consistent trend between decline in lung function and decline in HRQoL.

6.1.3 ERG exploratory scenario analyses

The ERG performed the following exploratory scenario analyses to explore the impact of alternative assumptions conditional on the ERG base-case.

6.1.3.1 Exploratory scenario analyses

1. Extrapolation of OS (Key Issue 3, Section 4.2.6.1)
   The ERG compared results obtained from extrapolating OS using their preferred frequentist Weibull curves, compared to the company's preferred Bayesian Weibull approach.

2. Time to first acute exacerbation, recurrent exacerbations and loss of lung function (Sections 4.2.6.2-4.2.6.4)
   The ERG examined the impact of adjusting the time to first acute exacerbation and removing the possibility of recurrent exacerbations. The ERG also examined the impact of using the probabilities of loss of lung function generated using the coefficients of the alternative model provided at clarification.

3. Health state utility values and disutilities (Section 4.2.8)
   A scenario will be conducted showing the HSUVs applied as they are in the company base-case. Several scenarios exploring alternative disutilities for acute exacerbations and AEs were conducted to examine the impact of these disutilities on results.

6.1.4 ERG subgroup analyses

No subgroup analyses were performed by the ERG.

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

6.2.1 Results of the ERG preferred base-case scenario

The ERG's base-case deterministic cost effectiveness results are presented in Table 6.1. The total costs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with incremental costs of [redacted] associated with nintedanib. The total QALYs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an incremental cost effectiveness ratio (ICER) of [redacted] per QALY gained.

Table 6.1: ERG base-case deterministic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incr. costs (£)	Incr. LYG	Incr. QALYs	ICER (£/QALY)
Nintedanib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
BSC	[redacted]	[redacted]	[redacted]				

Source: ERG preferred base case, applied in electronic model from the response to the clarification letter.[3]
BSC = best supportive care; ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY(s) = quality adjusted life year(s).

The ERG's probabilistic cost effectiveness results are presented in Table 6.2. The total costs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with incremental costs of £[redacted] associated with nintedanib. The total QALYs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an incremental cost effectiveness ratio (ICER) of [redacted] per QALY gained. These PSA results are not

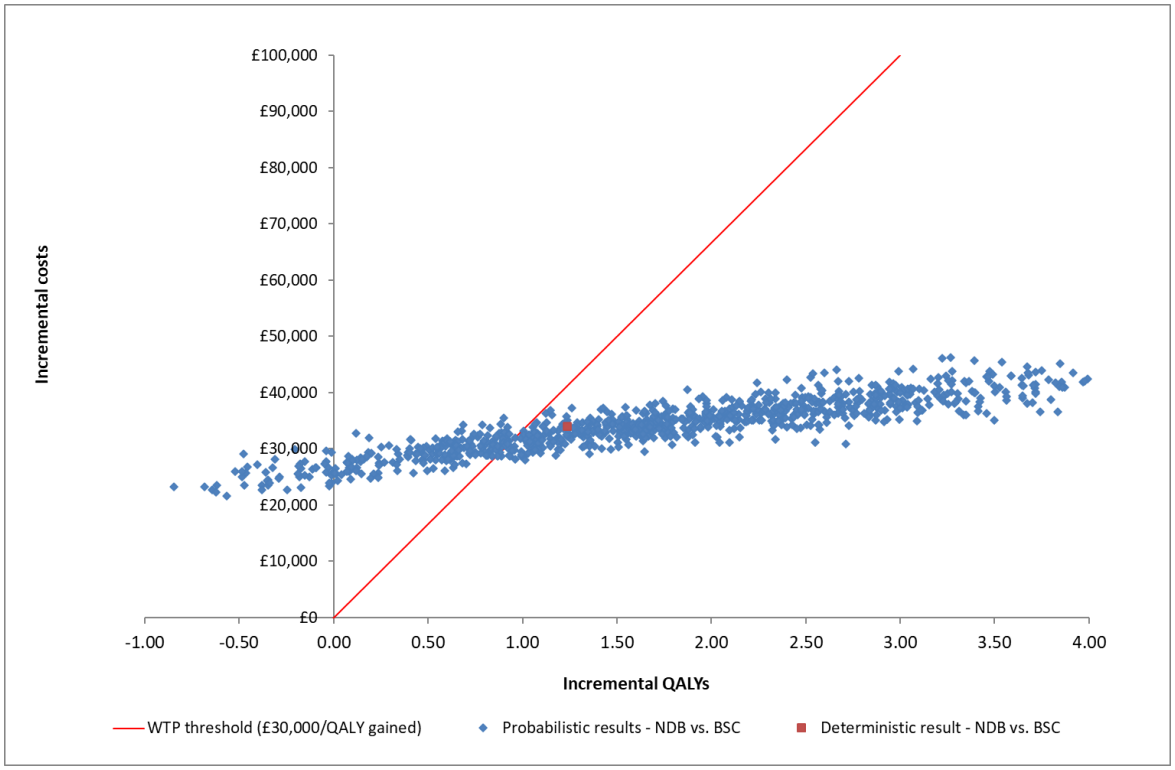
in line with the deterministic base-case results, with the difference being due to a discrepancy in the estimates for total QALYs per treatment. This misalignment is presumably caused by an issue with the PSA that the ERG could not resolve within the time that was available to them. Therefore, the ERG advises that this issue is resolved by the company at Technical Engagement so that the correct probabilistic results can be provided. The CE-plane and CEAC that are provided below in Figures 6.1 and 6.2 also pertain to the results from the PSA that includes this technical issue.

Table 6.2: ERG base-case probabilistic cost effectiveness results (discounted)

Technologies	Total costs (£)	Total LYG	Total QALYs	Incr. costs (£)	Incr. LYG	Incr. QALYs	ICER (£/QALY)
Nintedanib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
BSC	[redacted]	[redacted]	[redacted]				

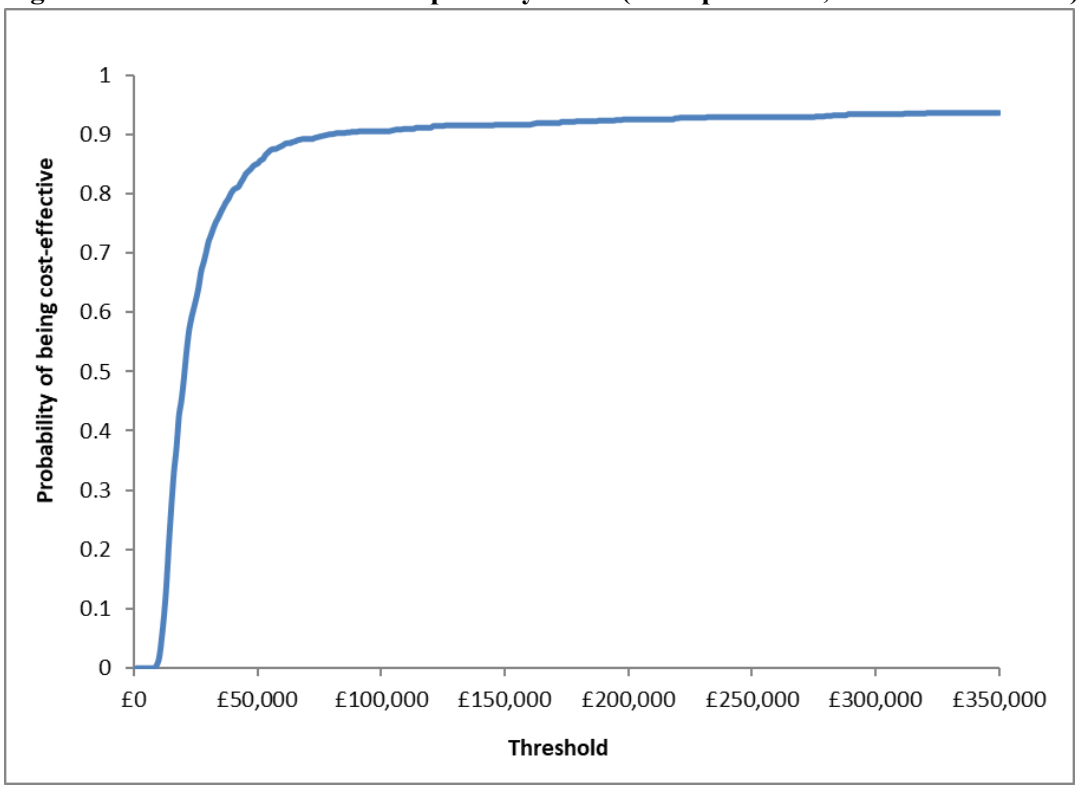
Source: The ERG preferred version of the electronic model provided in response to clarification questions.[3]
BSC = best supportive care; ICER = incremental cost effectiveness ratio; LYG = life years gained; QALY(s) = quality adjusted life year(s).

Figure 6.1: Cost effectiveness plane (ERG preferred, includes PSA issue)



Source: The ERG preferred version of the electronic model provided in response to clarification questions.[3]
BSC = best supportive care; ERG = evidence review group; NDB = nintedanib; PSA = probabilistic sensitivity analysis, QALY(s) = quality-adjusted life year(s); WTP = willingness to pay.

Figure 6.2: Cost effectiveness acceptability curve (ERG preferred, includes PSA issue)



Source: The ERG preferred version of the electronic model provided in response to clarification questions.[3]
ERG = evidence review group; PSA = probabilistic sensitivity analysis.

Based on the PSA results that substantially underestimated the ICER due to a technical issue, the probability that nintedanib is cost effective relative to BSC is 48.8% and 72.0% at ICER thresholds of £20,000 and £30,000 per QALY gained respectively.

6.2.2 Results of the ERG scenario analyses

6.2.2.1 Scenario set 1: Overall survival

Table 6.3 shows that extrapolating OS with the Bayesian Weibull, as per the company's base-case, reduces the ICER by approximately £7,500.

Table 6.3: OS scenarios

OS	Nintedanib		BSC		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
Bayesian Weibull (company BC)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Frequentist Weibull (ERG BC)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]

Source: ERG preferred base case, applied in electronic model from the response to the clarification letter.[3]
BC = base-case; BSC = best supportive care; ERG = Evidence Review Group; ICER = incremental cost-effectiveness ratio; Incr. = incremental; OS = overall survival; QALYs = quality adjusted life years.

6.2.2.2 Scenario set 2: Time to first acute exacerbation, recurrent exacerbations and loss of lung function

The results in Table 6.4 demonstrate that assumptions regarding time to first acute exacerbation, the inclusion of recurrent exacerbations and the method used to estimate the decline in lung function in nintedanib patients have a small impact on results. The largest impact was seen for TTFAE, but this scenario assumed that nintedanib had no impact on TTFAE, which is likely to be overly conservative.

Table 6.4: Scenarios regarding TTFAE, recurrent exacerbations and loss of lung function

TTFAE, recurrent exac. and lung function	Nintedanib		BSC		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
TTFAE							
TTFAE BC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
TTFAE BSC= NDB	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Recurrent exacerbation							
Recurrent exacerbation included (ERG and Company post-CL BC)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
No recurrent exacerbation (CS BC)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Loss of lung function nintedanib							
Estimated from OR (BC)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Estimated directly from regression results	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]

Source: ERG preferred base case, applied in electronic model from the response to the clarification letter.[3]
BC = base-case; BSC = best supportive care; CL = clarification letter; CS = company submission ERG = Evidence Review Group; ICER = incremental cost-effectiveness ratio; Incr. = incremental; NDB = nintedanib; OR = odds ratio; QALYs = quality adjusted life years; TTFAE = time to first acute exacerbation.

6.2.2.3 Scenario set 3: Health state utility values and disutilities

Table 6.5 indicates that the adjustment to the HSUV for the 80-89 FVC%Pred health state had minimal impact on the ICER. Doubling the assumed disutility for GI AEs increased the ICER by approximate £1,500, but all other changes had less than £600 impact.

Table 6.5: Health state utility values and disutilities

HRQoL	Nintedanib		BSC		Incr. Costs (£)	Incr. QALYs	ICER (£)
	Costs (£)	QALYs	Costs (£)	QALYs			
Health state utility values							
HSUVs company BC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
HSUVs ERG BC	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Disutilities							
Disutility for GI AEs 0.068 (TA379)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Disutility for GI AEs 0.042	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Disutility for acute exacerbation 0.14	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Disutility for acute exacerbation 0.274	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]

Source: ERG preferred base case, applied in electronic model from the response to the clarification letter.[3]
AEs = adverse events; BC = base-case; BSC = best supportive care; ERG = Evidence Review Group; GI = gastrointestinal; HSUV = health state utility value; ICER = incremental cost-effectiveness ratio; Incr. = incremental; OS = overall survival; QALYs = quality adjusted life years.

6.3 ERG's preferred assumptions

Table 6.6 below displays the step-by-step changes made by the company during clarification, followed by the changes made by the ERG, alongside the cumulative impact of each change, added to the previous changes, on results. This clearly shows that the only change which had a substantial impact on the ICER was modelling OS using the frequentist Weibull rather than the Bayesian Weibull and (increased the ICER by approximately £8,000). All other changes had less than £1,000 impact on the ICER.

Table 6.6: ERG's preferred model assumptions (cumulative)

Preferred assumption	Section in ERG report	Nintedanib		BSC		Inc. Costs (£)	Inc. QALYs	Cumulative ICER (£/QALY)
		Total Costs (£)	Total QALYs	Total Costs (£)	Total QALYs			
Company CS original base-case	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Updating/correction of several costs	6.1.1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Inclusion of recurrent acute exacerbations	4.2.6.3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Age adjustment of utilities	4.2.8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Company post-clarification base-case	6.1.1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<20,000 [redacted]
Extrapolate OS using Weibull frequentist	4.2.6.1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
Adjustment of HSUV for 80-89 FVC%Pred to maintain consistent trend in decline.	4.2.8	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]
ERG base-case	6.1.2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	<30,000 [redacted]

Source: ERG preferred base case, applied in electronic model from the response to the clarification letter.[3]
AE = adverse event; BSC = best supportive care; ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; Inc. = incremental; OS = overall survival; PD = progressed disease; QALY = quality-adjusted life year.

6.4 Conclusions of the cost effectiveness section

The cost effectiveness analysis was based on a model with the same structure as the one used in TA379,[21] which was validated by UK clinicians and deemed appropriate by the ERG and committee in TA379. The current ERG notes that mortality is modelled, both in the company base-case and the ERG base-case, based solely on OS and independent of the rate of exacerbations and lung function decline even for patients with the lowest sustainable lung function who are at risk of a further decline. Although the model provides the option to also allow patients with the lowest sustainable lung function to transition to Death upon further decline, the ERG did not use this option since this would imply a double counting of mortality. Nevertheless, the ERG has concerns about the assumption that mortality is assumed to be independent of exacerbation rate and lung function decline.

The clinical effectiveness inputs for the model are based on the results of the INBUILD trial. As noted in Sections 2.3 and 3.6, the ERG has concerns regarding the representativeness of the placebo arm of the INBUILD trial for best supportive care in UK clinical practice. This is because patients in INBUILD were not permitted to receive immunomodulatory treatments (including azathioprine, cyclosporin, tacrolimus, rituximab, cyclophosphamide, mycophenolate mofetil and oral corticosteroids) at randomisation and during the first six months of the treatment period in INBUILD, though these treatments are part of mainstream best supportive care.

A key source of uncertainty in the model is survival. The survival data from INBUILD is fairly immature and therefore relies heavily on extrapolation. The company chose to extrapolate OS using a Bayesian Weibull curve given that: 1) the Bayesian analysis was guided by external long-term IPF data, which could increase the accuracy of long-term predictions; 2) clinicians considered the two Weibull options (frequentist or Bayesian) the most plausible in the long-term; 3) the Bayesian Weibull provided a reasonably good fit to external IPF data. However, the ERG considers that the incorporation of external long-term data into the survival analysis potentially added more uncertainty than it solved given that the long-term data was in an IPF rather than a PF-ILD population and required the use of matching. Additionally, while clinicians considered both extrapolations plausible, the frequentist Bayesian actually fit the long-term nintedanib IPF survival data better than the Bayesian. For these reasons, the ERG preferred the frequentist Bayesian for the extrapolation of OS.

Discontinuation from nintedanib treatment in the model was extrapolated using an exponential distribution to allow a simple constant risk of discontinuation. However, the extrapolation did not fit the data well. Additionally, the model structure and assumptions made created an implausible relationship between discontinuation and the ICER whereby a discontinuation rate of 100% produced the most cost effective ICER and had no impact on LYs. This is because in the company base-case, it was assumed that patients who had discontinued from nintedanib continued to be represented by the nintedanib survival analysis post-discontinuation, as most patients who discontinued treatment were included in the trial survival analysis, and mortality was modelled as independent from the rate of acute exacerbations and decline in lung function. Therefore, increasing the discontinuation rate had no impact on LYs and a very minor impact on QALYs, while leading to large cost savings in terms of treatment costs. Given that a high proportion of patients who discontinued nintedanib in the trial continued to be followed-up, the ERG consider that the OS is likely to reflect the weighted efficacy of patients on and off-treatment over the observed follow-up. However, the impact on efficacy in the longer-term remains uncertain as it is not clear whether the trial follow-up is sufficiently long to fully capture the impact of discontinuation on OS. The way discontinuation has been incorporated into the survival analysis makes it impossible to assess the impact of changes in the discontinuation rate on the ICER, as a new OS curve would be needed.

Other more minor uncertainties relating to treatment effectiveness relate to the estimation of time to first acute exacerbation, the inclusion of recurrent exacerbations in the model and the method used to estimate loss of lung function. The company included the risk of recurrent exacerbation in their base-case during clarification, which had a very minor impact on results. Uncertainties surrounding the extrapolation of TTFAE and the model used to estimate loss of lung function also have a minor impact on results as exacerbations are not drivers of results and therefore no base-case changes were made.

The company included those treatment related/emergent AEs which had an incidence of >10% in either treatment arm and an incidence at least 1.5 times higher in the treatment arm than in the control arm. The ERG requested that the 10% cut-off be amended to 5%, but the company refused stating that events of all severities were included and not just serious or severe adverse events and given that no severe AEs occurred in >5% of patients receiving nintedanib, the overall impact on costs of extending the criteria to 5% was expected to be negligible. HRQoL was measured using the EQ-5D-5L during the INBUILD trial and valued using the UK cross-walk value set, as preferred by NICE. FVC%Pred HSUVs were estimated from this data and resulted in a largely consistent trend between decline in lung function and lower HRQoL. The ERG adjusted one HSUV for the 80-89 FVC%Pred health state to ensure a plausible trend in their base-case, but this had minimal impact on results. Disutilities were applied for the GI AEs included in the model, assuming a utility value from TA379 and for acute exacerbations based on the estimated impact of acute exacerbations on utility from the INBUILD data.

The company used a similar approach as in TA379 for the inclusion of resource use and costs in the model, with the inclusion of the following costs: drug acquisition costs for nintedanib, liver function test costs, health care resource use costs corresponding to each of the health states in the model, acute exacerbation costs, end of life costs, and costs in relation to adverse events. The ERG agrees that the company's approach to model resource use and costs is appropriate, in line with the assessment performed by the ERG in TA379.

The company's base-case deterministic cost effectiveness results, based on their post-clarification model indicated total costs for nintedanib and BSC of [redacted] and [redacted] respectively, with incremental costs of [redacted] associated with nintedanib, and total QALYs for nintedanib and BSC of [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an incremental cost effectiveness ratio (ICER) of [redacted] per QALY gained. The same model was used by the ERG in an attempt to provide post-clarification PSA results, at which point a technical issue with the PSA surfaced due to the QALY results not being comparable to the deterministic results. The ERG advises that the company resolves the PSA issue at Technical Engagement.

The ERG base-case differed from the company's post-clarification base-case in two ways: 1) OS was extrapolated using the frequentist Weibull; and 2) the HSUV for the 80-89 FVC%Pred health state was adjusted to ensure a plausible trend in HSUVs. The ERG's base-case deterministic cost effectiveness results indicate total costs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with incremental costs of [redacted] associated with nintedanib. The total QALYs for the nintedanib and BSC arms were [redacted] and [redacted] respectively, with an incremental QALY gain of [redacted] associated with nintedanib. This resulted in an incremental cost effectiveness ratio (ICER) of [redacted] per QALY gained. The larger ICER in the ERG base-case is largely due to the different approach for OS extrapolation. Assumptions around the extrapolation of OS also had the largest impact on results of all scenarios explored by the ERG.

The key uncertainties in the model are the long-term efficacy of nintedanib and BSC. Short-term trial data resulted in immature survival data, and therefore the model relies heavily on extrapolation. Different potentially plausible extrapolations produce substantially different results, making the base-case ICER uncertain. Uncertainty in the treatment effect is further increased by uncertainty regarding the extent to which the comparator arm in the trial truly reflects BSC in clinical practice, particularly given the observed treatment restrictions in the first six months of INBUILD. Without more data in these areas, these uncertainties cannot be resolved, and the results of the model remain somewhat speculative.

7. REFERENCES

[1] Boehringer Ingelheim. *Nintedanib for the treatment of adults with progressive fibrosing interstitial lung disease (PF-ILD): Submission to National Institute of Health and Care Excellence. Single technology appraisal (STA)*: Boehringer Ingelheim, 2021. 259p.

[2] National Institute for Health and Care Excellence. *Nintedanib for treating progressive fibrosing interstitial lung disease excluding idiopathic pulmonary fibrosis [ID1599]. Final scope*: NICE, 2020. 5p.

[3] Boehringer Ingelheim. *Nintedanib for treating progressive fibrosing interstitial lung disease [ID1599] – Response to request for clarification from the ERG*: Boehringer Ingelheim, 2021. 54p.

[4] Patel AS, Siegert RJ, Brignall K, Gordon P, Steer S, Desai SR, et al. The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax* 2012;67(9):804-810.

[5] Boehringer Ingelheim. *Nintedanib for treating progressive fibrosing interstitial lung disease excluding idiopathic pulmonary fibrosis [ID1599]. Company decision problem form*: Boehringer Ingelheim, 2020. 14p.

[6] Meltzer EB, Noble PWC. Idiopathic pulmonary fibrosis. *Orphanet J Rare Dis* 2008;3:8.

[7] Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183(6):788-824.

[8] Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012;28(2):138-44.

[9] Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350(9074):326-9.

[10] Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf MI, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*: Cochrane, 2019. Available from: http://www.training.cochrane.org/handbook

[11] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 23.3.11] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[12] Flaherty KR, Brown KK, Wells AU, Clerisme-Beaty E, Collard HR, Cottin V, et al. Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. *BMJ Open Respir Res* 2017;4(1):e000212.

[13] Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019;381(18):1718-1727.

[14] Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet* 2011;377(9779):1760-9.

[15] British Thoracic Society. BTS ILD Registry Annual Report 2019. 2019.

[16] British Thoracic Society. *BTS ILD Registry Annual Report 2019*: British Thoracic Society, 2019. 31p. Available from: https://www.brit-thoracic.org.uk/document-library/quality-improvement/ild-registry/bts-ild-registry-annual-report-2019/

[17] Wells AU, Flaherty KR, Brown KK, Inoue Y, Devaraj A, Richeldi L, et al. Nintedanib in patients with progressive fibrosing interstitial lung diseases-subgroup analyses by interstitial lung disease diagnosis in the INBUILD trial: a randomised, double-blind, placebo-controlled, parallel-group trial. *Lancet Respir Med* 2020;8(5):453-460.

[18] Boehringer Ingelheim. *Clinical trial report for the INBUILD trial (study 1199.247)*, 2019

[19] Flaherty KR, Wells AU, Cottin V, Devaraj A, Inoue Y, Richeldi L, et al. Effects of nintedanib on progression of ILD in patients with fibrosing ILDs and a progressive phenotype: further analyses of the INBUILD® trial. Poster developed for the European Respiratory Society International Congress, 7–9 September 2020.

[20] Maher TM, Corte TJ, Fischer A, Kreuter M, Lederer DJ, Molina-Molina M, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Respir Med* 2019;27:27.

[21] National Institute for Health Care Excellence. *Nintedanib for treating idiopathic pulmonary fibrosis [TA379]*. Manchester, UK: NICE, 2016 Available from: https://www.nice.org.uk/guidance/ta379/

[22] Boehringer Ingelheim. *Advisory board minutes. Data On File NIN15-04*, 2015

[23] National Institute for Health and Clinical Excellence. *Boehringer Ingelheim (BIL) PF - ILD NICE Advisory Board. Wednesday 11 November 2020. Minutes of Meeting*, 2020. 5p.

[24] National Institute for Health Care Excellence. *Guide to the methods of technology appraisal 2013*. London: National Institute for Health and Care Excellence., 2013 Available from: http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf

[25] Richeldi L, et al. Long-term treatment of patients with idiopathic pulmonary fibrosis with nintedanib: results from the TOMORROW trial and its open-label extension. *Thorax* 2018;73(6):581.

[26] Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014;370(22):2071-82.

[27] Crestani B, Huggins JT, Kaye M, al e. Long-term safety and tolerability of nintedanib in patients with idiopathic pulmonary fibrosis: results from the open-label extension study, INPULSIS-ON. *Lancet Respir Med* 2019;7:60-68.

[28] Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: Evolution, critique and future directions. *Stat Med* 2009;28:3049-3067.

[29] R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019 Available from: https://www.R-project.org/

[30] Jackson CH. flexsurv: a platform for parametric survival modeling in R. *Journal of Statistical Software* 2016;70.

[31] Soikkeli F, Hashim M, Ouwens M, et al. Extrapolating survival data using historical trial-based a priori distributions. *Value Health* 2019;22:1012-1017.

[32] Vasakova M, Sterclova M, Mogulkoc M, 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 Respir J* 2019;54:PA4720.

[33] Antoniou K, Markopoulou K, Tzouvelekis A, Trachalaki A, Vasarmidi E, Organtzis J, 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):00172-2019.

[34] Jo HE, Glaspole I, Grainge C, Goh N, Hopkins PM, Moodley Y, et al. Baseline characteristics of idiopathic pulmonary fibrosis: analysis from the Australian Idiopathic Pulmonary Fibrosis Registry. *Eur Respir J* 2017;49(2).

[35] Boehringer Ingelheim. *INBUILD phase III trial (no. 1199.247) post-hoc analysis. Data on file*, 2019

[36] Lancaster L, Crestani B, Hernandez P, Inoue Y, Wachtlin D, Loaiza L, 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.

[37] National Institute for Health and Care Excellence. *Nintedanib for treating progressive fibrosing interstitial lung disease [ID1599]: Clarification letter*. London: NICE, 2021. 13p.

[38] Guenther A, Prasse A, Kreuter M, Neuser P, Rabe K, Bonella F, et al. Exploring Efficacy and Safety of oral Pirfenidone for progressive, non-IPF Lung Fibrosis (RELIEF). *European Respiratory Journal. Conference: 29th International Congress of the European Respiratory Society, ERS. Spain.* 2019;54.

[39] EMC. Summary of Product Characteristics Ofev 150mg. 2020.

[40] Homma S, Azuma A, Taniguchi H, Ogura T, Mochiduki Y, Sugiyama Y, et al. Efficacy of inhaled N-acetylcysteine monotherapy in patients with early stage idiopathic pulmonary fibrosis. *Respirology* 2012;17(3):467-77.

[41] Kind P, Hardman G, Macran S. *UK Population Norms for EQ-5D*. York: Centre for Health Economics, University of York, 1999

[42] EMA. *Ofev: EPAR product information*: European Medicines Agency, 2020 Available from: https://www.ema.europa.eu/documents/product-information/ofev-epar-product-information_en.pdf

[43] NHS Improvement. National schedule of NHS costs 2018/19. 2019.

[44] NHS Improvement. National schedule of NHS costs 2017/18. 2018.

[45] Curtis L, Burns A. *Unit Costs of Health and Social Care*, 2019

[46] National Clinical Guideline Centre. *Diagnosis and management of suspected idiopathic pulmonary fibrosis: idiopathic pulmonary fibrosis*. London: National Clinical Guideline Centre, 2013. 307p.

[47] Department of Health and Social Care. *NHS reference costs 2010-2011 [Internet]*. London: Department of Health and Social Care, 2011 [accessed 12.3.21] Available from: https://www.gov.uk/government/publications/2010-11-reference-costs-publication

[48] Rinciog C, et al. A cost-effectiveness analysis of nintedanib in idiopathic pulmonary fibrosis in the UK. *Pharmacoeconomics* 2017;35(4):479-491.

[49] Georghiou T, M B. *Exploring the cost of care at the end of life*. London, UK, 2014

[50] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.