

**NATIONAL INSTITUTE FOR HEALTH AND CARE  
EXCELLENCE**

**Appraisal consultation document**

**Nintedanib for treating progressive fibrosing  
interstitial lung diseases excluding idiopathic  
pulmonary fibrosis**

The Department of Health and Social Care has asked the National Institute for Health and Care Excellence (NICE) to produce guidance on using nintedanib in the NHS in England. The appraisal committee has considered the evidence submitted by the company and the views of non-company consultees and commentators, clinical experts and patient experts.

**This document has been prepared for consultation with the consultees.** It summarises the evidence and views that have been considered, and sets out the recommendations made by the committee. NICE invites comments from the consultees and commentators for this appraisal and the public. This document should be read along with the evidence (see the [committee papers](#)).

The appraisal committee is interested in receiving comments on the following:

Has all of the relevant evidence been taken into account?

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

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

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

**Note that this document is not NICE's final guidance on this technology. The recommendations in section 1 may change after consultation.**

After consultation:

The appraisal committee will meet again to consider the evidence, this appraisal consultation document and comments from the consultees.

At that meeting, the committee will also consider comments made by people who are not consultees.

After considering these comments, the committee will prepare the final appraisal document.

- Subject to any appeal by consultees, the final appraisal document may be used as the basis for NICE's guidance on using nintedanib in the NHS in England.

For further details, see [NICE's guide to the processes of technology appraisal](#).

The key dates for this appraisal are:

Closing date for comments: 25 August 2021

Second appraisal committee meeting: 9 September 2021

Details of membership of the appraisal committee are given in section 5.

## 1 Recommendations

- 1.1 Nintedanib is not recommended, within its marketing authorisation, for treating progressive fibrosing interstitial lung diseases in adults (except idiopathic pulmonary fibrosis, for which it is recommended in [NICE technology appraisal guidance 379](#)).
- 1.2 This recommendation is not intended to affect treatment with nintedanib that was started in the NHS before this guidance was published. People having treatment outside this recommendation may continue without change to the funding arrangements in place for them before this guidance was published, until they and their NHS clinician consider it appropriate to stop.

### Why the committee made these recommendations

Current treatment for fibrosing interstitial lung disease starts with immunosuppressants. If recommended, nintedanib would be given at the same time. The clinical trial evidence suggests that nintedanib slows the decline of lung function compared with placebo. But, follow up was short so it was unclear whether people live longer. It is also uncertain what its effects would be in clinical practice because the trial restricted use of some immunosuppressants in the first 6 months. The economic model was also uncertain because of the company's modelling of survival in the placebo arm, nintedanib's long-term benefits, and for how long people would take it.

Because of these uncertainties, it is not possible to estimate nintedanib's cost effectiveness. Therefore, nintedanib is not recommended.

## 2 Information about nintedanib

### Marketing authorisation indication

- 2.1 Nintedanib (Ofev, Boehringer Ingelheim) is indicated in adults for the treatment of chronic fibrosing interstitial lung diseases with a progressive phenotype.

### Dosage in the marketing authorisation

- 2.2 The dosage schedule is available in the [summary of product characteristics](#).

### Price

- 2.3 The price of nintedanib is £2,151.10 per pack of 60 capsules, each containing 150 mg (excluding VAT; BNF online accessed July 2021). The company has a commercial arrangement. This makes nintedanib available to the NHS with a discount and it would have also applied to this indication if the technology had been recommended. The size of the discount is commercial in confidence. It is the company's responsibility to let relevant NHS organisations know details of the discount.

## 3 Committee discussion

The [appraisal committee](#) considered evidence submitted by Boehringer Ingelheim, a review of this submission by the evidence review group (ERG), NICE's technical report, and responses from stakeholders. See the [committee papers](#) [\[Add link to website in-development page on 'committee papers'\]](#) for full details of the evidence.

### Nature of the condition

#### Progressive fibrosing interstitial lung disease is a debilitating condition that increases the risk of dying

- 3.1 Interstitial lung disease (ILD) encompasses a large and varied group of lung disorders characterised by inflammation and fibrosis of the lung parenchyma. Some ILDs become progressive fibrosing ILD (PF-ILD).

Appraisal consultation document – Nintedanib for progressive fibrosing interstitial lung disease excluding idiopathic pulmonary fibrosis

Page 4 of 30

Issue date: July 2021

© NICE 2021. All rights reserved. Subject to [Notice of rights](#).

Idiopathic pulmonary fibrosis is one of the most common types of ILD and is associated with early mortality (Cottin et al. 2019). One patient expert explained that he understood that most people with cancer have a better life expectancy than people with PF-ILD; although the committee was not presented with these data, the committee appreciated the increased risk of dying and the need for treatment options. Idiopathic pulmonary fibrosis may share some features of natural history with other types of PF-ILD. This appraisal addresses PF-ILD excluding idiopathic pulmonary fibrosis, for which NICE has recommended nintedanib in [NICE's technology appraisal guidance on nintedanib for treating idiopathic pulmonary fibrosis](#) (from now on referred to as TA379). Other types of PF-ILD include idiopathic interstitial pneumonias, autoimmune ILDs, hypersensitivity pneumonitis, and sarcoidosis. People with PF-ILD may also have underlying conditions such as rheumatoid arthritis. Clinical experts explained that some underlying diseases start with more inflammation than others, while some mainly have fibrosis (for example, asbestosis) but all end up with fibrotic lungs. The progressive phenotype in PF-ILD is characterised by a gradual decline in lung function, dyspnoea (breathlessness), worsening of physical performance and quality of life, as well as poor response to immunomodulatory therapies and early mortality. The patient expert explained that in some people symptoms evolve rapidly and, within a short time, breathlessness limits physical activity. Ultimately, some people become housebound and dependent on supplementary oxygen before dying within 3 to 4 years after being diagnosed. The committee concluded that PF-ILD is a debilitating condition that increases the risk of dying relative to the general population.

## **Diagnosis and progression of PF-ILD**

### **NHS clinicians use multiple criteria to identify and diagnose PF-ILD**

3.2 Clinical experts explained that, in the NHS, people with PF-ILD can be seen by different medical specialists, particularly respiratory and rheumatology teams. Clinicians identify disease progression in people

with ILD using multiple criteria. These include a relative decline of at least 10% on spirometry in forced vital capacity percentage (FVC%) predicted, based on algorithms that are adjusted for age, sex and height. A worsening of fibrosis (shown on CT scan) and an increase in dyspnoea are also used. Clinical experts explained that these criteria are comparable to the diagnostic criteria of PF-ILD in INBUILD, the company's main trial (see section **Error! Reference source not found.**). This trial included patients who met at least 1 of the following criteria in the past 24 months, despite treatment with medications used in clinical practice to treat ILD:

- A relative decline in FVC% predicted of at least 10% predicted.
- A relative decline in FVC% predicted of at least 5% to less than 10% predicted with worsening of respiratory symptoms or increasing extent of fibrotic changes on high-resolution chest imaging.
- Worsening of respiratory symptoms as well as increasing extent of fibrotic changes on high-resolution chest imaging.

The committee understood that clinicians use multiple criteria to diagnose PF-ILD in practice.

### **A decline in FVC of at least 10% predicted defines progression and is associated with disease deterioration and mortality in PF-ILD**

3.3 The committee was aware that a relative (Cottin et al. 2019) or absolute (Goos et al. 2021; Wong et al. 2020) decline in FVC of at least 10% predicted is one of the criteria that defines disease progression in PF-ILD. The company used an absolute decline in FVC% predicted of 10% in its economic model to categorise disease progression and define health states (see section 3.10). The committee also understood that a decline in FVC, measured either as the change from baseline in millilitres or as a percentage of the predicted value, is commonly used in clinical studies as an end point to measure disease progression in ILD. Evidence, confirmed by clinical experts, suggests that a decline in FVC% predicted is

associated with disease deterioration and mortality in idiopathic pulmonary fibrosis (Cottin et al. 2019). Clinical experts explained that usually a lower FVC level is associated with a higher risk of death in PF-ILD. The committee was aware that the end point of the key trial was absolute FVC in millilitres, whereas the company modelled FVC% predicted. The committee was not presented with the algorithm chosen by the company to estimate FVC% predicted. The committee concluded that a relative or absolute decline of at least 10% predicted in FVC is an acceptable criterion for disease progression reflecting deterioration in lung function and is associated with higher risk of death in PF-ILD. The committee further concluded it would like to see how the company transformed trial data into modelled data.

## Company's positioning of nintedanib

### Company's positioning of nintedanib in the treatment pathway is appropriate and would reflect an add-on to standard care

3.4 Current treatment of PF-ILD in the NHS includes treatment for the underlying disease (see section 3.1). The conventional treatments used include, but are not limited to, systemic corticosteroids, mycophenolate mofetil, azathioprine, cyclophosphamide, methotrexate, rituximab, infliximab and best supportive care. The clinical experts highlighted that immunosuppressants are commonly used for underlying autoimmune diseases. The company explained that these treatments that are used as standard care do not have a marketing authorisation for PF-ILD. However, the committee was aware that [NICE's guide to the methods of technology appraisal](#) (2013), permits it to consider unlicensed treatments in appraisals if their use in clinical practice is established. The clinical and patient experts noted that treatments currently used for PF-ILD treat the inflammation rather than the fibrosis. They explained that there are no randomised placebo-controlled trials that suggest that these treatments delay progression of chronic fibrosis in PF-ILD. The committee heard from clinical experts that sometimes conventional treatments are stopped if

they do not slow the decline of pulmonary function, but may not be stopped because they are treating the underlying condition. The committee appreciated that if recommended, nintedanib would be an add-on therapy to conventional treatments. The committee understood it would be offered to people with ILD in whom the underlying disease has progressed despite receiving conventional treatments. The committee concluded that the company's positioning of nintedanib, as an add-on treatment to standard care in people with ILD whose disease has progressed despite receiving conventional treatments, is appropriate.

**The baseline characteristics of the INBUILD trial population, with the exception of concurrent treatments, are generalisable to the NHS**

3.5 INBUILD was a phase 3, randomised, multicentre trial comparing nintedanib (n=332) with placebo (n=331) in people with PF-ILD (see section 3.2 for diagnostic criteria). The trial consisted of 2 parts: PART A, in which all patients took nintedanib or placebo for an initial period of at least 52 weeks; and Part B, in which patients continued to have blinded, randomised treatment beyond week 52. The trial continued until all patients had completed Part A (reached their week-52 visit) and a benefit and risk profile of nintedanib over the 52 weeks had been assessed, or until a withdrawal. Patients had different treatment periods in Part B. The trial asked those who stopped the treatment to attend all visits as originally planned, including an end-of-treatment visit and a follow-up visit 4 weeks later. There were 2 database locks in the trial. Database lock 1 was beyond week 52, and database lock 2 (median follow up about 19 months, Brown et al. 2020) took place after all patients had completed the follow-up visit. Patients could receive open-label nintedanib in an extension study known as INBUILD-ON (n=436, completed) if the benefit and risk assessment of nintedanib was positive. The committee was aware that INBUILD included 22 patients from the UK. In both arms, patients had largely balanced baseline characteristics for sex, age, and decline in lung function measured before the trial. Clinical experts noted that the criteria for disease progression for entry into the trial, for example,

a relative decline in FVC of at least 10% predicted in the 24 months before screening, are in line with what would be seen in practice (see section 3.2). The committee agreed that the baseline characteristics of the INBUILD trial population, with the exception of concurrent treatments (section 3.6), were representative of NHS patients.

### **Concurrent treatments in the INBUILD trial do not reflect current NHS care**

3.6 The committee discussed that in both arms of the INBUILD trial, patients could not have immunosuppressants other than systemic corticosteroids at a maximum dose of 20 mg per day (see section 3.4) at randomisation and for the first 6 months (26 weeks) of the trial. The committee was aware that 68.6% of patients (70.1% in the placebo arm and 67.2% in the nintedanib arm) used corticosteroids over the 52-week period in INBUILD. Patients who had received other immunosuppressants could participate in the trial if they underwent a wash-out period 4 to 8 weeks before randomisation. After the first 6 months and for the rest of the trial, patients with worsening ILD or connective tissue disease could receive immunosuppressants. Approximately 16% of patients started immunosuppressants during the second 6 months of the initial 52-week period (21% in the placebo arm and 11% in the nintedanib arm). The committee interpreted this to show that fewer patients randomised to nintedanib than placebo needed immunosuppressants, but that a substantial proportion of participants needed the treatments that the protocol restricted earlier in the trial. The ERG noted that immunosuppressants are not restricted in clinical practice and that, because nintedanib is considered to be an add-on treatment (see section 3.4), placebo without conventional standard treatments does not reflect NHS clinical practice. Therefore, it is not an appropriate comparator. The clinical experts explained that nintedanib would be offered to reduce the dosage and use of corticosteroids, but the committee was not presented with any evidence for this. The committee concluded that the INBUILD trial does not represent NHS clinical practice because it restricted

immunosuppressant use, and if nintedanib was recommended, it would be added on to current treatments.

## Clinical effectiveness

### **Nintedanib is associated with a slower decline of lung function compared with placebo, but its long-term treatment effect is uncertain**

3.7 The primary end point of INBUILD (see section 3.5) was annual rate of decline in FVC (in millilitres per year) over 52 weeks adjusted for baseline FVC (in millilitres) and imaging pattern. The committee noted that this is different from FVC% predicted, which defines disease progression in the INBUILD inclusion criteria and as an outcome in the company's economic model. In the trial the company analysed the change in absolute FVC measured in millilitres using a random coefficient regression model with baseline FVC and imaging pattern as covariates. The slope of the FVC decline was calculated for every patient over 52 weeks and the average compared between the 2 arms. The analysis was done over 52 weeks and included people who had discontinued (see section 3.5). Results from adjusted annual rate of decline in FVC showed that:

- Over 52 weeks, the decline in lung function differed between the nintedanib (-80.8 ml/year) and placebo (-187.8 ml/year) arms (alpha level=0.05 [2-sided]) by 107.0 ml/year (95% confidence interval [CI] 65.4 to 148.5; p value <0.001). One of the clinical experts explained that 10% in FVC% predicted is an FVC of about 150 ml and this between-group difference of 107.0 ml/year in adjusted rate may represent a clinically meaningful change. The committee noted it was unclear whether a between-group difference of 107 ml/year in adjusted rate of decline in FVC over 52 weeks equals a 10% difference (relative or absolute) in FVC% predicted, which would indicate a clinically meaningful change in FVC (see section 3.3).
- After 52 weeks, the difference in adjusted annual rate of decline in FVC between nintedanib and placebo appeared to narrow until 84 weeks,

after which it plateaued. The company explained that this was because of the different treatment periods of patients in Part B of the trial and associated missing values of FVC, and that the analysis of the whole trial that the company used in its model, should be 'interpreted with caution'. The committee noted that the decrease of treatment effect suggested either a waning effect of nintedanib in the long term or a treatment effect of immunosuppressants, which more people had in the placebo arm than in the nintedanib arm (see section 3.6).

The committee was aware that what reflected a 'minimally clinically important difference' in the literature is based usually on FVC% predicted which is adjusted for age, sex, and height (see section 3.2) whereas the trial results were FVC adjusted for baseline FVC and imaging. Therefore, it concluded it was uncertain how these values were translated from the trial to the model. It concluded that the long-term relative effect of nintedanib may decrease and agreed that this should be reflected in the company's cost-effectiveness analyses.

### **Nintedanib's treatment effect on mortality is uncertain**

3.8 Secondary endpoints assessed in INBUILD included overall survival, change from baseline King's Brief Interstitial Lung Disease (K-BILD) questionnaire total score (a health-related quality-of-life measure) and time until first acute ILD exacerbation or death. Other endpoints included the EQ-5D questionnaire and measures of safety. Endpoints were collected at 52 weeks and at database lock 2 (median follow up: about 19 months, see section 3.5). Results showed that:

- There was no statistically significant difference between nintedanib (4.8%; n=16 out of 332) and placebo (5.1%; n=17 out of 331) arms in death at 52 weeks (hazard ratio [HR] 0.94; 95% CI 0.47 to 1.86; p=0.85). Evidence showed the same at database lock 2, with 10.8% (36 out of 332) and 13.6% (45 out of 331) of patients having died in the nintedanib and placebo arms, respectively (HR 0.78; 95% CI 0.50 to

1.21; p value not reported). Median overall survival was not reached because of the low number of events. The committee noted the wide confidence intervals and the additional 48 deaths between week 52 and database lock 2, which the clinical experts could not explain within the natural history of the disease.

- Absolute change from baseline in total score on K-BILD showed no statistically significant difference between nintedanib (adjusted mean: 0.55) and placebo (adjusted mean: -0.79) arms at 52 weeks (adjusted mean difference: 1.34; 95% CI -0.31 to 2.98; p=0.1115).
- Time to first acute ILD exacerbation or death also showed no statistically significant difference between nintedanib (7.8%, 26 out of 332) and placebo (9.7%, 32 out of 331) arms at 52 weeks (HR 0.80; 95% CI 0.48 to 1.34; p=0.3948). The difference was statistically different at database lock 2 with 13.9% (46 out of 332) and 19.6% (65 out of 331) of patients in nintedanib and placebo arms experiencing the event (HR 0.67; 95% CI 0.46 to 0.98).

The committee concluded that there is uncertainty about whether nintedanib was associated with a 'clinically meaningful change' in FVC% predicted compared with placebo. The committee was aware that the trial was not powered to detect a difference in mortality, but concluded that the trial did not show conclusively that nintedanib prolonged life.

## **Adverse events**

### **The adverse event profile of nintedanib is acceptable**

- 3.9 In INBUILD at 52 weeks, the nintedanib and placebo arms showed similar frequencies of any adverse event and of serious adverse events. Gastrointestinal discomfort, especially diarrhoea, was the most common adverse event (66.9% in the nintedanib arm compared with 23.9% in the placebo arm). People taking nintedanib were more likely to show

biochemical indication of hepatic injury and were more likely to reduce the dose of nintedanib because of gastrointestinal symptoms. Based on their observations, the clinical experts explained that 25% to 30% of people may not tolerate nintedanib longer term. They noted that taking nintedanib is not associated with an increased risk of infection unlike some immunosuppressants. The committee concluded that the safety profile of nintedanib was acceptable.

## **Economic model**

### **The model structure is acceptable for decision making but there are important uncertainties in its assumptions**

3.10 To model the natural history of PF-ILD beyond the end of the trial, the company assumed that idiopathic pulmonary fibrosis and PF-ILD have the same natural history of disease. The company therefore adopted the same Markov model structure it had used in TA379. In the model, patients accrue quality-adjusted life years by improving both quality and length of life. Nintedanib improves the quality of life because patients experience fewer exacerbations, and their lung function declines more slowly. In the model, patients on nintedanib also live longer compared with those on placebo. The company defined health states using lung function (10-point categorisations of FVC% predicted, distribution calculated from the INBUILD trial) and rates of exacerbation over a cycle length of 3 months in the model. Patients enter the model at different lung-function health states without exacerbation. They can then either remain in the same health state, or transition to a health state with the same lung function with the occurrence of exacerbation or with a 10-point lower FVC% predicted with or without exacerbation, or death. This was to reflect a clinically meaningful decline in FVC% predicted (see section 3.7) and the distribution of patients at randomisation in INBUILD.

3.11 The transition probabilities for exacerbation in the model were informed by data from the INBUILD trial's second database lock. The company

assumed that patients could not transition to a health state with better lung function. Also, after an exacerbation, patients could not transition to a health state without exacerbation for the rest of the time horizon.

Transition to death can either happen from any health state based on a survival analysis using the INBUILD data or, by reaching a level of FVC% predicted below 40%, at which the company assumed that lung function is unsustainable. The committee was aware that the model did not include absolute decline in FVC controlled for baseline FVC and imaging, which defined the primary end point of the trial. The ERG noted that the company modelled mortality based only on overall survival data from INBUILD. This means that mortality is independently modelled from lung function and that the same risk of death is applied to all health states, even for patients with the lowest level of lung function. The committee recalled that a lower level of lung function was strongly associated with a higher mortality rate (see section 3.3) and questioned the company on why it did not model a change in risk of death for lower levels of lung function. The company explained that this was to avoid double counting as the overall survival data includes all deaths. The company also modelled mortality independently from acute exacerbations, despite it stating that acute exacerbations are often fatal and a major cause of mortality in PF-ILD.

- 3.12 The ERG did not make any change to the company's modelling assumptions but noted that modelling mortality independently from lung-function decline and acute exacerbations can produce implausible results in relation to stopping treatment in the model. The committee understood that the way the company modelled mortality meant that modifying assumptions around the rate at which people stopped treatment with nintedanib or placebo (see section 3.23), exacerbation rates (see section 3.19), and decline in lung function (see section 3.20) would have minimal impact on the cost effectiveness. The committee was aware that mortality data from INBUILD trial were sparse and showed no difference in mortality between treatments (see section 3.8). Considering the trade-off

between scarcity in overall survival data and the potential confounding factors in the relationship between level of FVC% predicted, exacerbations, and mortality, the committee concluded that there are important uncertainties in the model structure and limitations in its implementation, which the committee accounted for in its decision making.

### **The company used the frequentist and Bayesian approaches to extrapolate overall survival**

The company used 2 approaches to extrapolate overall survival beyond the trial duration: a frequentist method and an exploratory Bayesian analysis.

#### **The frequentist approach involves fitting standard parametric distributions based on PF-ILD data**

3.13 The frequentist approach involved fitting standard parametric distributions independently to each arm of INBUILD based on PF-ILD data alone. The company considered the best fitting curves for the observed data based on statistical fits (lower Akaike information criterion [AIC]/Bayesian information criterion [BIC] scores) were log-logistic, Gompertz and Weibull curves. The committee appreciated that considerable uncertainty exists given the immaturity of the survival data with some 90% of the population alive after database lock 2 (see section 3.8).

#### **Using the Bayesian approach to model overall survival is reasonable**

3.14 The company aimed to improve the accuracy and precision of the extrapolated overall survival using a Bayesian approach. The company assumed PF-ILD and idiopathic pulmonary fibrosis have similar disease trajectories including death, based on evidence from Brown et al. (2020) and Simpson et al. (2020). The company used data from idiopathic pulmonary fibrosis trials, matched those idiopathic pulmonary fibrosis patients to PF-ILD patients in the INBUILD trial, and fitted the survival curves for idiopathic pulmonary fibrosis patients to inform the shape of the survival curves for PF-ILD in the model. This involved:

- Obtaining idiopathic pulmonary fibrosis data by combining several idiopathic pulmonary fibrosis trials with long-term follow up, including TOMORROW, a phase 2 study; INPULSIS I and II, 2 phase 3 randomised control trials (RCTs); and INPULSIS ON, a combined long-term extension of the 2 RCTs.
- Matching idiopathic pulmonary fibrosis patients (from the combined idiopathic pulmonary fibrosis trials) to PF-ILD patients (from INBUILD) on chosen baseline characteristics using a propensity score weighting method. They included age, sex, race, time since idiopathic pulmonary fibrosis or PF-ILD diagnosis, FVC% predicted at baseline and smoking status.
- Generating survival curves for the matched idiopathic pulmonary fibrosis patients. Parametric survival curves were fitted to the matched idiopathic pulmonary fibrosis data for both nintedanib and placebo arms. Models with the lowest AIC and BIC were selected.
- Generating an informative prior from idiopathic pulmonary fibrosis parametric survival curves and generating survival curves for PF-ILD. For the selected idiopathic pulmonary fibrosis parametric survival curves, the shape parameters were retained as the informative prior for nintedanib and placebo. Using these informative priors, parametric models were fitted to the INBUILD trial data. The company considered the 3 best fitting curves to be Weibull, log-logistic and gamma distributions.

The company validated those 3 best fitting curves by seeking clinicians' advice and comparing survival curves with external registry data. Based on this, the company considered the Bayesian approach, by using long-term idiopathic pulmonary fibrosis data and using a Weibull model, provided more robust estimates of long-term survival than the frequentist approach. It therefore adopted Weibull Bayesian curves for both nintedanib and placebo arms in its base case. The committee appreciated the company's approach and effort to use additional long-

term data from a related disease (idiopathic pulmonary fibrosis), but noted it was not without uncertainty. Given the scarcity of long-term data in PF-ILD, the committee concluded that the Bayesian approach itself was reasonable to model overall survival, but that uncertainties exist.

**In the absence of data, it is reasonable to consider that idiopathic pulmonary fibrosis and PF-ILD share similar disease trajectories including death**

3.15 The company's Bayesian approach assumed that PF-ILD and idiopathic pulmonary fibrosis have similar disease trajectories including death (see section 3.14). The committee was aware that PF-ILD and idiopathic pulmonary fibrosis may share similar features and natural history (see section 3.1) and questioned how valid it was to assume similar survival between the 2 populations. The company explained that they based this assumption on 2 sources of data. First, Brown et al. (2020), which is an observational study assessing the clinical course of PF-ILD and idiopathic pulmonary fibrosis, and includes people receiving placebo in the INBUILD trial (PF-ILD) and INPULSIS trials (idiopathic pulmonary fibrosis). The study, sponsored by the manufacturer, reported that PF-ILD and idiopathic pulmonary fibrosis had similar clinical course including death because there were no statistically significant difference in death between the 2 populations at 52 weeks. However, the committee was aware that the HR of 0.63 (95% CI 0.35 to 1.13) suggested the possibility of a lower rate of death for PF-ILD than for idiopathic pulmonary fibrosis. Given that deaths were sparse in people taking placebo in both the INBUILD trial (5.1%; 17 out of 331) and INPULSIS trials (7.8%; 33 out of 423), and the lack of reporting of what known or unknown confounding factors the authors adjusted for in Brown et al., the committee considered that the evidence should be interpreted with caution. Simpson et al. (2020), an observational study of PF-ILD across England, included 2,368 people with a new referral for ILD diagnosed with the same criteria for PF-ILD as used

in the INBUILD trial (see section 0). The study reports that people with PF-ILD and people with idiopathic pulmonary fibrosis have similar survival at 900 days (HR 1.06; 95% CI 0.84 to 1.35). Additionally, it showed that the shapes of the survival curves between PF-ILD and idiopathic pulmonary fibrosis were similar. The committee considered that the shape of the survival curve was more important than the actual HR when assuming similar disease trajectories, particularly given the approach taken by the company in its Bayesian analysis. The committee also recalled that idiopathic pulmonary fibrosis is a subset of PF-ILD (see section 3.1) and agreed it was reasonable to assume similar survival between people with PF-ILD and idiopathic pulmonary fibrosis. It concluded that the assumption that people with idiopathic pulmonary fibrosis and PF-ILD have similar disease trajectories, when controlling for confounders, on which the company based its Bayesian approach, is reasonable.

### **Modelling and validation of overall survival for the placebo arm is uncertain and its impact on the model results is not clear**

3.16 The committee discussed the company's validation of the Weibull Bayesian survival curves for both nintedanib and placebo arms. It noted that the company chose the Bayesian Weibull curves for both nintedanib and placebo based on statistical fits, clinician input and comparison with external idiopathic pulmonary fibrosis registry data. Clinicians consulted by the company noted that there is limited knowledge on nintedanib's long-term impact. For the placebo arm, the clinicians consulted by the company agreed that Weibull curves (both frequentist and Bayesian) were clinically plausible. The company also used several idiopathic pulmonary fibrosis registries to validate the survival curves fitted for the PF-ILD population in the Bayesian approach, specifically:

- Nintedanib arm: idiopathic pulmonary fibrosis registry data from EMPIRE and Greek idiopathic pulmonary fibrosis studies. The EMPIRE study (Vasakova et al. 2013) provides approximately 10 years of follow up in 637 people with idiopathic pulmonary fibrosis who were taking

nintedanib across Europe. The Greek registry (Antoniou et al. 2020) reports 5-year survival data in 244 people in Greece with idiopathic pulmonary fibrosis taking nintedanib.

- Placebo arm: data available from untreated people with idiopathic pulmonary fibrosis from the EMPIRE study, an Australian registry, a European registry including British patients, and a Finnish registry.

3.17 Considering the Bayesian approach itself reasonable (see section 3.14), the committee noted the following with regards to the company's validation of survival curves:

- Overall, there was a shorter follow up in placebo arms compared with nintedanib arms in the idiopathic pulmonary fibrosis trials that were used to inform the shape of survival curves for PF-ILD (see section 3.14). The placebo data from those idiopathic pulmonary fibrosis trials were not meaningfully more mature than those from the INBUILD trial for PF-ILD. Therefore, there may be greater uncertainties when extrapolating survival in the long term for the placebo arm compared with nintedanib arm.
- The company's Bayesian approach resulted in large differences in survivals between different extrapolated curves fitted for PF-ILD, and predicted large differences in life expectancy between nintedanib and placebo arms as well as within the placebo arm.
- For nintedanib, the committee appreciated that registries from countries where nintedanib is offered to people with severe idiopathic pulmonary fibrosis (that is, with FVC below 50% of predicted) may display worse survival than in the UK, where treatment should be offered only to people with less severe disease (that is, with FVC between 50% and 80% of predicted).
- For nintedanib, the ERG noted that the Weibull Bayesian curve selected by the company consistently overpredicts survival compared with the Greek registry and follows the survival from the EMPIRE study for the first year, but then overpredicts it. The ERG considered that the

Weibull frequentist model provided a better fit to the registry data and therefore included Weibull frequentist curves for both placebo and nintedanib in their base case.

- For placebo, the company noted that registries varied considerably with respect to survival. The committee noted that, nonetheless, the Weibull Bayesian curves dropped more quickly (had a higher death rate) than the registries' survival. This meant that the company may be underestimating the survival of patients who do not take nintedanib by using Weibull Bayesian curves. The committee recognised that this is despite the fact that patients in registries are generally sicker than patients in trials. The committee also noted that the European registry, which had baseline FVC measurement (mean FVC% predicted) comparable to the INBUILD trial population, showed the lowest death rate among all the registries presented. It included patients from the UK and the committee considered it the best source to validate the placebo arm survival estimates. The committee also noted that the European registry described people with idiopathic pulmonary fibrosis as having worse baseline characteristics and prognostic factors (a higher proportion of people who smoke,) than people in INBUILD. This suggested that the company's modelling of survival for people with PF-ILD on placebo (not on nintedanib) even when using the European registry, appears pessimistic.

Given the above, the committee was concerned that the modelling of survival in the placebo arm may not be appropriate. The committee concluded that it likely overpredicts deaths in the placebo arm. The committee considered that the impact of this on the modelling of the treatment effect of nintedanib compared with placebo is unknown, but would very likely worsen the estimates of cost effectiveness. The committee recalled the short follow up of the INBUILD trial, which does not provide robust evidence that nintedanib prolongs life, and the consideration of whether this effect would persist in the longer term

(see section 3.8). The committee concluded that it was uncertain whether nintedanib's treatment effect on survival would last, and that the way overall survival has been modelled for the placebo arm would likely underestimate the cost-effectiveness estimates for nintedanib, but that the extent of this is unknown.

**There are uncertainties in fitting independent parametric survival distributions to both arms and the observed and implied hazards should be explored**

3.18 The company fitted parametric distributions to the nintedanib and placebo arms independently to extrapolate overall survival beyond the trial data in both its frequentist and Bayesian approaches. The committee understood that when parametric distributions are fitted individually to each treatment arm, no assumptions about the proportionality of treatment effects are made. The committee noted that the company's chosen fits resulted in ever-increasing survival benefits for nintedanib compared with the placebo in the extrapolated portion of the survival curves. The committee considered that this is not supported by the short-term evidence on mortality from the INBUILD trial (see section 3.8) and recalled that the evidence suggested that nintedanib's treatment effect may decrease in the longer term (see section 3.7). The committee therefore considered that the company's approach of fitting parametric survival curves individually to the nintedanib and placebo arms may bias estimates of cost effectiveness in favour of nintedanib. The committee was not provided with evidence that the company had explored the proportionality of treatment effects in the observed data and had not been presented with information on the treatment effect over time implied by the company's chosen curves. It concluded that the company should explore the proportionality of hazards observed in the data and provide information on the treatment effect implied by the alternative survival modelling approaches considered.

### **The distribution used to model acute exacerbations is acceptable**

3.19 The company modelled time to first acute exacerbation using an exponential curve that generated a per-cycle (3-month cycle length) risk of exacerbation of 1.76% and 1.12% for people receiving placebo and nintedanib, respectively. The ERG noted that the exponential model overpredicted the risk of acute exacerbation after approximately 8 months and showed a sharp drop towards the end of follow up. This is likely to substantially increase the difference observed between nintedanib and placebo. The committee was aware that both the company and the ERG's varying risk of exacerbation in scenario analyses had little impact on the cost effectiveness. This was likely because of mortality not being directly linked to the occurrence of acute exacerbation in the model (see section 3.12), representing an important limitation of the model. The committee concluded that the curve used to model acute exacerbations was acceptable, but that the lack of a link between exacerbations and mortality was an important limitation of the model.

### **The modelling of loss in lung function is acceptable but the lack of link to mortality is an important limitation**

3.20 Patients started the model in different FVC% predicted categories (distribution informed by patients' baseline FVC% predicted values in the INBUILD trial; see section 3.10). The company used 2 different methods to calculate probabilities of losing lung function in the placebo arm (multivariate logistic regression based on data from INBUILD trial) and nintedanib arm (odds ratio sourced from INBUILD trial applied to baseline placebo risk). At clarification, the ERG asked the company to use the regression analysis for both arms, but this change had a minimal impact on the cost effectiveness. The ERG noted that this may be because 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 2 methods. It also noted that that both methods assumed a lifetime treatment effect for nintedanib. The

committee was aware that decline of lung function is not linked to mortality (see section 3.12) in the model, and that data from the INBUILD trial suggested that the treatment effect of nintedanib may decrease over time. The committee concluded that the method used to model the loss of lung function over time was acceptable but there were uncertainties over what algorithm the company used to estimate FVC% predicted and the treatment effect over the long term. It recognised that the lack of link between loss of lung function and mortality was an important limitation of the model.

### **The ERG's adjustment of utility values is appropriate**

3.21 EQ-5D was collected in INBUILD and the company used this data to estimate the utility values in the model for each FVC% predicted health state (see section 3.10). The regression analysis resulted in a higher value in people in the 80% to 89% FVC% predicted category than the ones in the 90% to 99% predicted category (0.7333 compared with 0.7287). Because the ERG considered this clinically implausible, they corrected it by applying a linear decline in utility from the 90% to 99% predicted and 70% to 79% predicted health states. This resulted in a utility value of 0.7265 for the 80% to 89% FVC% predicted health state in its base case cost-effectiveness analyses. The committee concluded it was satisfied with the ERG's utility adjustment.

### **The company's choice of utility decrement for recurrent exacerbations and gastrointestinal events is appropriate**

3.22 The company modelled utility decrements for recurrent exacerbations and adverse events. When people have an acute exacerbation, their utility drops by 0.167, which the company estimated from regression analysis using the EQ-5D data collected in INBUILD. The ERG identified 2 other estimates for acute exacerbations and explored the impact of those in scenario analyses. When people have a gastrointestinal adverse events, their utility drops by 0.034. The ERG based this on estimates from TA379, and the assumption that nintedanib has a similar safety profile regardless

of being used to treat PF-ILD or idiopathic pulmonary fibrosis. In TA379 the company had used a decrement of 0.068. But the company assumed that 0.034 (half of 0.068) was plausible for PF-ILD based on a phase 3 trial in recurrent non-small-cell lung cancer for nintedanib, which estimated a decrement for grade 3/4 diarrhoea of 0.042 (Boehringer Ingelheim: data on file, 2014). These were estimated from EQ-5D data from the INPULSIS trial in idiopathic pulmonary fibrosis. The ERG was unclear on the company's rationale behind the use of 0.034 rather than 0.068. However, it noted that adverse events had limited impact on the model results. The ERG identified 2 other estimates for gastrointestinal events and explored the impact of those in scenario analyses. The committee was broadly satisfied with the company's choice of utility decrements for recurrent exacerbations and gastrointestinal events but noted these had minimal impact on the cost effectiveness.

### **The way the company modelled time-to-stopping treatment is uncertain and may underestimate the cost of nintedanib**

3.23 The company modelled time-to-stopping treatment with nintedanib using an exponential curve. The company did this for consistency with TA379 and it resulted in an overall rate of 5.97% for stopping nintedanib per month. The company considered that the exponential model did not fit the INBUILD trial data. This is because the model underestimated stopping nintedanib in the first year, but from approximately 15 months onwards, the model appeared to overestimate stopping nintedanib. After validating the model with long-term data on safety and efficacy of nintedanib in people with idiopathic pulmonary fibrosis (Lancaster et al. 2019), the company considered that the exponential model may underestimate the true rate of stopping nintedanib, therefore overestimated its costs. However, the committee noted that the exponential curve dropped quickly and seemed to overestimate stopping of nintedanib compared with the INBUILD trial, and therefore underestimated its costs. In response to the technical engagement, the company also provided alternative models to assess the impact of using different curves to extrapolate time-to-stopping

treatment with nintedanib beyond the trial period. The Gompertz curve was closest to the Kaplan–Meier data from INBUILD over 3 years, but it unrealistically underestimated the stopping of nintedanib treatment over the long term. The ERG noted that, given the mean age of 65 years at baseline in the INBUILD trial, the Weibull model probably gave a more realistic extrapolation of time on nintedanib over the long term than other curves. This is because most patients would not stay on treatment for 35 years. The committee noted that the choice of distribution does not impact on the quality-adjusted life years and costs, other than the acquisition cost for nintedanib. This is because different health states in the model are not associated with different mortality rates (see section 3.11). The committee concluded that the way the company has modelled time-to-stopping treatment with nintedanib was uncertain and may have underestimated the cost of nintedanib in the model.

## End of life

### **The company's submission does not make a case for end of life care**

3.24 The committee considered the advice about life-extending treatments for people with a short life expectancy in [NICE's guide to the methods of technology appraisal](#). The company did not make a case for end of life care in its submission. The committee noted that the life expectancy of people who had placebo would normally be longer than 24 months.

### **The end-of-life criteria are not met**

3.25 The evidence suggested that the median overall survival was expected to be 2 to 5 years in people with PF-ILD not receiving anti-fibrotic therapy (Meltzer et al. 2008; Raghu et al. 2011). The committee concluded that the end-of-life criteria were not met for nintedanib in PF-ILD.

## Cost-effectiveness estimates

### The cost effectiveness of nintedanib for PF-ILD is unknown

3.26 Because of confidential commercial arrangements for nintedanib, the cost-effectiveness estimates cannot be reported here. The committee noted that none of the company's or ERG's analyses reflected assumptions that addressed the committee's concerns and, therefore, the true incremental cost-effectiveness ratio was unknown.

### Uncertainties should be explored

3.27 The committee noted that the company should explore the following:

- The impact of restricting concurrent NHS treatments on the treatment effect of nintedanib in the INBUILD trial (see section 3.6).
- The algorithm used to adjust trial results (FVC in millilitres per year adjusted for baseline FVC value in millilitres and imaging pattern) to FVC% predicted; and whether the trial result reflects a clinically meaningful difference (see sections 3.4 and 3.7).
- Incorporating a decrease in nintedanib's treatment effect in the long term in the company's cost-effectiveness analyses (see section 3.7).
- The uncertainties in the modelling and validation of the overall survival in the placebo arm (see sections 3.16 and 3.17).
- The uncertainties in fitting individual parametric distributions to the nintedanib and placebo arms; to explore the proportionality of hazards observed in the data and provide information on the treatment effect implied by the alternative survival modelling approaches considered (see section 3.18).
- The limitations in modelling mortality independently from exacerbations (see section 3.19) as well as decline in lung function (see section 3.20).
- The uncertainties in the modelling of stopping nintedanib treatment (see section 3.22).

## **Nintedanib does not meet NICE's criteria for an innovative treatment**

3.28 The company considered nintedanib to be innovative because it is the only treatment licensed for people with PF-ILD. In addition, the company stated that nintedanib is the first pharmacological treatment to show clinical evidence of slowing disease progression in people with PF-ILD. However, given the shortcomings of the company's modelling, the committee could not determine whether this reflects a 'step change' in treatment.

## **Equality issues**

### **There are no equality issues within the committee's remit**

3.29 The patient experts considered there was inequality because people with idiopathic pulmonary fibrosis can access nintedanib while people with PF-ILD currently cannot. The committee noted that equality issues could only be assessed within the marketing authorisation defined in this appraisal. The patient experts further explained that people with PF-ILD are generally younger than people with idiopathic pulmonary fibrosis and more likely to be of South Asian and African-Caribbean family origin, which the committee noted was not an equality issue itself in this topic. The committee concluded that there were no equality issues within the committee's remit.

## **Conclusion**

### **The committee cannot assess nintedanib's long-term effectiveness and value for money and therefore it is not recommended**

3.30 The committee acknowledged that PF-ILD is a debilitating condition associated with morbidity and mortality. It noted that if recommended, nintedanib would be an add-on to the current standard care. It noted that the clinical evidence suggested that nintedanib may slow the decline of lung function, but there were uncertainties about its effectiveness and

around the assumptions and approaches in the model. The committee concluded that:

- The impact of restricting concurrent NHS treatments on the treatment effect of nintedanib is unclear.
- It is unclear whether the primary end point measured by FVC in millilitres per year over 52 weeks reflects a clinically meaningful change as measured by FVC% predicted.
- Nintedanib's treatment effect may decrease in the long term, but this was not incorporated in the company's economic analyses.
- There are uncertainties in the company's modelling and validation for overall survival in the placebo arm. It likely overpredicts deaths in the placebo arm. The impact of this on the modelling of the treatment effect of nintedanib compared with placebo is unknown but would very likely worsen the estimates of cost effectiveness.
- There are uncertainties in fitting individual parametric distributions to the nintedanib and placebo arms. The modelling resulted in ever-increasing survival benefits for nintedanib compared with placebo in the extrapolated periods. However, the evidence from the INBUILD trial suggested that nintedanib may not prolong life compared with placebo and its long-term effect uncertain.
- There are uncertainties in company's modelling of exacerbations and decline in function because of their lack of a link with mortality in the model.
- The modelling of stopping treatment was uncertain and may have underestimated the costs of nintedanib.

Taking these and other uncertainties into account, the committee concluded that it had not been presented with the evidence it needed to assess nintedanib's long-term effectiveness and value for money. Therefore, nintedanib is not recommended as an option for treating PF-ILD in the NHS.

## 4 Proposed date for review of guidance

- 4.1 NICE proposes that the guidance on this technology is considered for review by the guidance executive 3 years after publication of the guidance. NICE welcomes comment on this proposed date. The guidance executive will decide whether the technology should be reviewed based on information gathered by NICE, and in consultation with consultees and commentators.

Amanda Adler  
Chair, appraisal committee B  
July 2021

## 5 Appraisal committee members and NICE project team

### Appraisal committee members

The 4 technology appraisal committees are standing advisory committees of NICE. This topic was considered by [committee B](#).

Committee members are asked to declare any interests in the technology to be appraised. If it is considered there is a conflict of interest, the member is excluded from participating further in that appraisal.

The [minutes of each appraisal committee meeting](#), which include the names of the members who attended and their declarations of interests, are posted on the NICE website.

### NICE project team

Each technology appraisal is assigned to a team consisting of 1 or more health technology analysts (who act as technical leads for the appraisal), a technical adviser and a project manager.

**Aminata Thiam**

Technical lead

**Yelan Guo**

Technical adviser

**Joanne Ekeledo**

Project manager

ISBN: [to be added at publication]