

**NATIONAL INSTITUTE FOR HEALTH AND CARE  
EXCELLENCE**

**Final appraisal document**

**Nintedanib for treating progressive fibrosing  
interstitial lung diseases**

**1 Recommendations**

- 1.1 Nintedanib is recommended, within its marketing authorisation, as an option for treating chronic progressive fibrosing interstitial lung diseases (PF-ILD) in adults.

**Why the committee made these recommendations**

Current treatment for progressive fibrosing interstitial lung diseases (PF-ILD) often starts with immunosuppressants, which may or may not be continued when nintedanib is offered.

The clinical trial evidence suggests that nintedanib slows the decline of lung function compared with placebo. But, there are uncertainties in the evidence: it is unclear if nintedanib helps people to live longer, and the trial reflects how nintedanib would be used in the NHS in some, but not all, people with PF-ILD.

Because follow-up was short in the trial for nintedanib in FP-ILD, the economic model uses longer follow up data from nintedanib trials in idiopathic pulmonary fibrosis, a related condition that progresses in a similar way. This allows better modelling of nintedanib's long-term effect on life expectancy. The cost-effectiveness estimates are likely to be within what NICE considers an acceptable use of NHS resources. So, nintedanib is 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 (simple patient access scheme). This makes nintedanib available to the NHS with a discount. 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](#) 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. Idiopathic pulmonary fibrosis is one of the most common types of ILD and NICE has recommended nintedanib (see [NICE's](#)

[technology appraisal guidance on nintedanib for treating idiopathic pulmonary fibrosis](#), from now on referred to as TA379) and pirfenidone for its treatment (see [NICE's technology appraisal guidance on pirfenidone for treating idiopathic pulmonary fibrosis](#), from now on referred to as TA504). Some ILDs progress and become fibrotic and are referred to as progressive fibrosing ILD (PF-ILD). This appraisal addresses PF-ILD excluding idiopathic pulmonary fibrosis and covers systemic sclerosis-PF-ILD, for which, it should be noted, nintedanib has a separate marketing authorisation. Other types of PF-ILD include, but are not limited to, idiopathic interstitial pneumonias, autoimmune ILDs, and hypersensitivity pneumonitis. People with PF-ILD often have underlying systemic conditions such as rheumatoid arthritis or sarcoidosis. PF-ILD is characterised by a gradual decline in lung function, breathlessness, worsening physical performance and quality of life, poor response to immunomodulatory therapies and mortality (Cottin et al. 2019). One patient expert explained that he understood that most people with cancer live longer than people with PF-ILD and although the committee was not presented with these data, it appreciated that PF-ILD increased the risk of dying. The patient expert explained that symptoms can evolve quickly and, within a short time, breathlessness limits physical activity. Ultimately, some people become housebound and depend on supplementary oxygen. 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

### The key trial uses multiple criteria to diagnose PF-ILD, which reflects NHS practice

- 3.2 Clinical experts explained that in the NHS, people with PF-ILD are seen by different medical specialities, notably respiratory and rheumatology. Clinicians identify disease progression and fibrosis using multiple criteria. These include a relative decline of at least 10% on spirometry in forced vital capacity percentage (FVC%) predicted compared with pre-screening,

based on algorithms adjusting for age, sex and height; worsening of fibrosis on CT scan; and respiratory symptoms. Clinical experts explained that these criteria are comparable to the diagnostic criteria of PF-ILD in INBUILD, the company's main trial (see section 3.5). This trial included people who met 1 or more criteria despite treatment, including:

- A relative decline in FVC% predicted of at least 10% predicted compared with pre-screening in the past 24 months.
- A relative decline in FVC% predicted of at least 5% predicted, but less than 10% predicted, with worsening respiratory symptoms or increasing fibrotic changes on high-resolution chest imaging compared with pre-screening in the past 24 months.
- Worsening respiratory symptoms and increasing fibrotic changes on high-resolution chest imaging in the past 24 months.

The company estimates that there are nearly 900 people living with PF-ILD in the UK. The committee understood that clinicians use multiple criteria to diagnose PF-ILD in practice which are comparable to the key trial.

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

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 from baseline is one of the criteria that defines disease progression in PF-ILD. The committee recognised that the company reported the trial results in difference between treatment groups in absolute FVC adjusted for baseline FVC and imaging pattern, whereas in its economic model, the company used a 10% absolute decline in FVC% predicted to categorise disease progression and health states (see section 3.12). In its first meeting, the committee requested details of this transformation. At consultation, the company presented the equation it used to estimate FVC% predicted from absolute FVC in millilitres

accounting for ethnicity, age, and height (Quanjer et al. 2012). Evidence suggests that a decline in FVC% predicted is associated with worsening disease and mortality in idiopathic pulmonary fibrosis (Cottin et al. 2019) and, according to the clinical experts, also in PF-ILD. The committee concluded that a relative or absolute decline of at least 10% predicted in FVC is an acceptable criterion for disease progression and is associated with higher risk of death in PF-ILD.

## The company's positioning of nintedanib

### The company's positioning of nintedanib as an add-on treatment that will be stopped if disease progression is not slowed is appropriate

3.4 Current treatment of PF-ILD in the NHS includes treatments for the underlying disease (see section 3.1) including, but not limited to, systemic corticosteroids, mycophenolate mofetil, azathioprine, cyclophosphamide, methotrexate, rituximab, infliximab and best supportive care. The company explained that these treatments do not have marketing authorisations for PF-ILD. However, the committee was aware that [NICE's guide to the methods of technology appraisal](#), permits it to consider unlicensed treatments as part of standard care and if relevant, as comparators if their use in clinical practice is established. The clinical and patient experts noted that treatments for PF-ILD treat the inflammation rather than the fibrosis, and that immunosuppressants are used for the underlying autoimmune diseases. They explained that there are no trials that suggest that these treatments delay progression of chronic fibrosis in PF-ILD. Clinical experts explained that while they stop treatments that have not slowed the decline of pulmonary function, they would not stop drugs that treat the underlying condition. The committee understood that clinical experts would stop nintedanib if disease progression was not sufficiently slowed. The committee concluded that, if recommended, nintedanib would be an add-on treatment rather than a direct comparator to conventional treatments that may or may not be continued when adding nintedanib. The committee appreciated that the company did not model a

formal stopping rule because clinicians would stop treatment with nintedanib if they deemed that it was not effective. The committee concluded that the company's positioning of nintedanib, as an add-on treatment to standard care when ILD has progressed despite conventional treatments, is appropriate. The committee also agreed that nintedanib would be stopped if disease progression was not markedly slowed.

### **The baseline characteristics of the INBUILD population 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. Clinical experts noted that the criteria for disease progression for entry into INBUILD (see section 3.2) are in line with NHS practice. INBUILD included 22 participants from the UK. The trial consisted of 2 parts: Part A, in which patients took nintedanib or placebo for 52 weeks; and Part B, in which patients continued blinded, randomised treatment beyond week 52. The trial continued until all patients reached their week-52 visit or withdrew. Patients had different durations of treatment in Part B. Those who stopped study treatments were asked to attend all visits as originally planned. The company performed 2 database locks: database lock 1 occurred after the last participant had completed the week-52 visit, and database lock 2 (median follow up about 19 months; Brown et al. 2020) after all patients had completed the follow-up visit, about 3 months after finishing Part A. The company chose database lock 2 to inform the economic model. An extension study, INBUILD-ON (n=436, completed), offered open-label nintedanib to participants who the investigator deemed would benefit. The clinical experts noted that age, sex, and decline in lung function of INBUILD participants reflected NHS patients. The committee concluded that these baseline characteristics of the INBUILD population were representative of NHS patients.

## **Concurrent treatments in INBUILD reflect current NHS care for some, but not all, patients**

3.6 The committee was aware that people in the INBUILD trial could not have immunosuppressants other than up to 20 mg per day of oral corticosteroids (see section 3.4) at randomisation and for the first 6 months or unless they underwent a wash-out period 4 to 8 weeks before randomisation. After the first 6 months and for the rest of the 52-week period, patients with worsening ILD or connective tissue disease could have immunosuppressants. During this time, approximately 16% of patients started immunosuppressants (21% in the placebo arm and 11% in the nintedanib arm). The committee interpreted this as showing that some participants needed treatments that the protocol restricted earlier in the trial. The committee was aware that the NICE scope listed established clinical management of PF-ILD as azathioprine, cyclophosphamide, mycophenolate, corticosteroids, infliximab, rituximab, and best supportive care. The ERG noted that immunosuppressants are not restricted in clinical practice and that the trial population therefore does not reflect NHS clinical practice. The committee concluded that the INBUILD trial, which restricted concurrent medications, reflects NHS clinical practice in some, but not all people, with PF-ILD.

## **Placebo is an appropriate comparator for NHS clinical practice**

3.7 The committee appreciated that the INBUILD trial was a placebo-controlled trial. The committee was aware that the NICE scope listed the comparators as established clinical management without nintedanib. The committee considered nintedanib to be an add-on treatment (see section 3.4) and concluded that placebo was the appropriate comparator for the decision problem and NHS clinical practice.

## Clinical effectiveness

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

3.8 The primary endpoint of INBUILD (see section 3.5) was annual rate of decline in FVC (in millilitres per year) over 52 weeks analysed as the change in absolute FVC in millilitres using a random coefficient regression model, with baseline FVC and imaging pattern as covariates. The analysis included people who had discontinued treatment (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;  $p < 0.001$ ). The corresponding results on decline in FVC% predicted showed that nintedanib was associated with a slower decline in FVC% predicted from baseline (-2.6%) compared with placebo (-5.9%) at week 52 (mean difference: 3.2%; 95% CI 2.09 to 4.40). At consultation, the company reported that the minimum clinically meaningful difference for decline in FVC% predicted could be 2% to 6% in people with idiopathic pulmonary fibrosis according to Bois et al. (2011). Also at consultation, stakeholders explained that a difference of 107 ml in decline of FVC is significant in clinical practice.
- After 52 weeks, the difference in adjusted annual rate of decline in FVC between nintedanib and placebo narrowed. The committee noted that the data suggests a waning effect of nintedanib in the long term, or effect modification by concurrent medications added after the first 6 months of the trial (section 3.6). The company disputed this and explained that the results reflected different durations of treatment that patients had in Part B of the trial and missing FVC values. The company considered that the analysis that incorporated data from Part B that it used in its model, should be 'interpreted with caution' and had 'methodological limitations'. In the company's view, this included the



possibility of ‘healthy survivor bias’ because patients in the nintedanib arm stopped treatment because of adverse events while those in the placebo arm stopped treatment because of lack of effect. However, the ERG and committee appreciated that the company did not present data to support this. The committee also appreciated the limitations of INBUILD in measuring the effect of long-term treatment but recalled that the company chose the timepoint of database lock 2 to inform its economic model (see section 3.5). The committee concluded that there was uncertainty in nintedanib’s long-term treatment effect and that it would account for this in its decision making.

### **Effect modification by concurrent treatment remains uncertain**

3.9 The committee questioned whether the treatment effect of nintedanib without concurrent conventional treatments, as in INBUILD, would be the same as with them. During the committee’s second meeting, the company presented a post-hoc analysis of the treatment effect of nintedanib on adjusted annual rate of decline in FVC in the whole INBUILD population compared with the subgroup of patients who did not receive concurrent conventional treatments . The committee concluded that comparing a subgroup with the whole population (including the subgroup) was not a statistically meaningful comparison and that the subgroup used did not reflect NHS practice. The company presented another post-hoc analysis from INBUILD suggesting that there was no difference in the treatment effect of nintedanib between patients receiving or not receiving glucocorticoids at baseline (interaction  $p=0.18$ ). The committee was aware that 70.1% in the placebo arm and 67.2% in the nintedanib arm used prednisolone over the 52-week period in INBUILD. The committee concluded that it did not see evidence of an interaction, but could not discount the possibility that effect modification exists.

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

3.10 Secondary endpoints assessed in INBUILD included overall survival, change from baseline in the score for the King’s Brief Interstitial Lung

Disease (K-BILD) questionnaire (a health-related quality-of-life measure) and time until first acute exacerbation of ILD or death. Other endpoints included the EQ-5D questionnaire and safety. Endpoints were collected at 52 weeks and at database lock 2 (see section 3.5). Results showed that:

- The hazard ratio for death for people randomised to nintedanib (4.8%; n=16 of 332) compared with people randomised to placebo (5.1%; 17 of 331) at 52 weeks was 0.94 (95% CI 0.47 to 1.86; p=0.85). The hazard ratio for death for nintedanib (10.8%; 36 of 332) compared with placebo (13.6%; 45 of 331) at database lock 2 was 0.78 (95% CI 0.50 to 1.21; p value not reported). Median overall survival was not reached because of a low number of events. The committee noted the wide confidence intervals.
- Absolute change from baseline in total score on K-BILD showed no statistically significant difference at a 5% alpha 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.
- Results for time to first acute ILD exacerbation or death for nintedanib (7.8%, 26 of 332) and placebo (9.7%, 32 of 331) arms up to 52 weeks were HR 0.80; 95% CI 0.48 to 1.34; p=0.3948. At database lock 2, 13.9% (46 of 332) and 19.6% (65 of 331) of patients in nintedanib and placebo arms experienced an event (HR 0.67; 95% CI 0.46 to 0.98). The committee noted that this composite endpoint did not specifically inform the risk of exacerbation associated with nintedanib, which the company included in its model. It was aware that occurrence of exacerbation can increase the risk of death in PF-ILD (see section 3.13) and agreed that it would have been helpful to have data on this.

The committee was aware that the trial was not powered to detect a difference in mortality. It concluded that the data from INBUILD did not show conclusively that nintedanib prolongs life.

## Adverse events

### The adverse event profile of nintedanib is acceptable

3.11 In INBUILD at 52 weeks, the nintedanib and placebo arms showed similar frequencies of any adverse event and of serious adverse events based on on-treatment analyses. 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 have raised biochemical serum markers 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. They noted that nintedanib is not associated with an increased risk of infection unlike 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

3.12 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 a similar natural history of disease. The company adopted the same Markov model structure it had used in TA379. In the model, patients accrue quality-adjusted life years (QALYs) by improving both quality and length of life. The company assumed in its model that nintedanib improves the quality of life because patients experience fewer exacerbations, and their lung function declines more slowly. The company defined health states using lung function (10%-point categorisations of FVC% predicted) and rates of exacerbation over a cycle length of 3 months in the model. Patients enter the model at different lung-function health states assuming they have not yet had an exacerbation. They can then remain in the same health state, have an exacerbation and transition to a health state with the

same lung function or move to health state with a lower lung function, or not have an exacerbation and transition to health states characterised by a lower FVC% predicted, or die. The company defined the health states to reflect a clinically meaningful decline in FVC% predicted (see section 3.8). The health state in which a patient starts reflects the distribution of patients at randomisation in INBUILD.

- 3.13 The company used the data from database lock 2 of INBUILD to inform the model parameters including overall survival, time to first acute exacerbation, loss of lung function, stopping treatment, utility values and costs. The company estimated transition probabilities of mortality on parametric extrapolations (Bayesian approach) of overall survival data from INBUILD and applied irrespective of health states or model events with the exception of a FVC% predicted below 40% (see sections 3.15 and 3.16). The company fitted standard parametric models to extrapolate the risk of time to first exacerbation (see section 3.24), and calculated the probabilities of losing lung function in the 2 treatment arms using 2 different methods (see section 3.25). The committee was aware that the modelled FVC% predicted was transformed from the absolute decline in FVC controlled for baseline FVC and imaging, the primary endpoint of INBUILD (see section 3.3). The company assumed that patients could not transition to a health state with better lung function. 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, but does not differ by health state; it is 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. Although there were 2 routes to death in the model, 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. 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 as a function of FVC percent predicted. 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 may be fatal and a cause of mortality in PF-ILD.

- 3.14 The ERG did not change any of the company's modelling assumptions. It 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 stop treatment with nintedanib or placebo (see section 3.28), exacerbation rates (see section 3.24), and decline in lung function (see section 3.25) have minimal impact on cost effectiveness. At consultation, the company considered changing the structure of the model to include a link between mortality and exacerbations and decline in lung function. However, the company stated that the adapted model produced increased and unrealistic life years for both placebo and nintedanib compared with the current model. It considered this was because of the uncertainty with separate risks of death for each modelled health state. The committee was aware that mortality data from INBUILD showed no clear difference in mortality between treatments (see section 3.10). The committee concluded that there are important uncertainties in the model structure and limitations when implementing it.

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

The company used 2 approaches to extrapolate overall survival for people with PF-ILD beyond the trial duration: a frequentist method and a Bayesian analysis.

**The frequentist approach involves fitting standard parametric distributions based on Kaplan–Meier survival curves from the PF-ILD INBUILD data**

3.15 The frequentist approach involved fitting standard parametric distributions independently to each arm of INBUILD. The study included people with PF-ILD only. The company considered that the best fitting curves for the observed data had the best statistical fits (lower Akaike information criterion [AIC]/Bayesian information criterion [BIC] scores). These 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 3.10).

**Using the Bayesian approach to extrapolate overall survival is reasonable**

3.16 The company extrapolated overall survival using a Bayesian approach by combining short-term data on PF-ILD from INBUILD (median 12 months) with longer-term data from clinical trials of idiopathic pulmonary fibrosis (up to median 56.3 months for nintedanib arm, but a median of only up to 13.1 months for placebo). 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). In general, the company used data from trials of idiopathic pulmonary fibrosis and matched participants from these trials to participants with PF-ILD in INBUILD based on characteristics that might be associated with survival. The company then fitted survival curves to matched patients with idiopathic pulmonary fibrosis to inform the shape parameter of the survival curves for PF-ILD in the model. This specifically involved:

- Obtaining data on idiopathic pulmonary fibrosis by combining several trials of idiopathic pulmonary fibrosis, including: TOMORROW, a phase 2 trial; INPULSIS I and II, 2 phase 3 randomised control trials (RCTs); and INPULSIS ON, a combined long-term extension of the 2 RCTs.
- Matching patients with idiopathic pulmonary fibrosis (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, ethnicity, duration of disease, FVC% predicted at baseline and smoking.

- Generating survival curves for the matched idiopathic pulmonary fibrosis patients. The company then fitted parametric survival curves to the matched idiopathic pulmonary fibrosis data for both nintedanib and placebo arms and selected models with the lowest AIC and BIC.
- Generating an informative prior from parametric survival curves for idiopathic pulmonary fibrosis and generating survival curves for PF-ILD. For the selected parametric survival curves for idiopathic pulmonary fibrosis, the company retained the shape parameters as the informative prior for nintedanib and placebo. Using these informative priors, the company fitted parametric models to the INBUILD trial data. The company considered the 3 best fitting curves to be Weibull, log-logistic and gamma distributions.
- Validating the 3 best fitting curves by seeking clinicians' advice and comparing survival curves with external registry data on idiopathic pulmonary fibrosis.

Based on this, the company considered that the Bayesian approach, by using data from clinical trials of idiopathic pulmonary fibrosis and a Weibull model, better estimated long-term survival than the frequentist approach. The company adopted Weibull Bayesian curves for both nintedanib and placebo arms in its base case. The committee appreciated the company's approach using longer-term trial data from a related disease, but noted it was not without uncertainty. Given the absence 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.

### **It is reasonable to consider that idiopathic pulmonary fibrosis and PF-ILD share similar disease trajectories**

3.17 In using a Bayesian approach, the company assumed that PF-ILD and idiopathic pulmonary fibrosis have similar disease trajectories for death (see section 3.16). The committee was aware that PF-ILD and idiopathic pulmonary fibrosis may share features and natural history but questioned how valid it was to assume similar survival. The company explained that it based this assumption on 2 sources. The first, Brown et al. (2020), an observational study assessing the clinical course of PF-ILD and idiopathic pulmonary fibrosis, includes people randomised to placebo in the INBUILD trial (PF-ILD,) and INPULSIS trials (idiopathic pulmonary fibrosis). The study, sponsored by the manufacturer, reported death rates at 52 weeks for PF-ILD (5.1%, 17 of 331) and idiopathic pulmonary fibrosis (7.8%, 33 of 423). 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. The committee was aware that deaths were sparse, and that Brown et al. did not report what confounding factors were considered, and the possibility of residual confounding. Simpson et al. (2020), an observational study of PF-ILD across the UK, included 2,368 people with a new referral for ILD and 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 reported crude HR when assuming similar disease trajectories. The committee concluded it was reasonable to assume a similar shape parameter for survival between PF-ILD and idiopathic pulmonary fibrosis.



## Using registries of idiopathic pulmonary fibrosis to validate the survival extrapolations is a reasonable approach

3.18 The committee discussed the company's attempt to validate its choice of the Weibull overall survival curves based on Bayesian analyses for both nintedanib and placebo using clinical input and registry data for idiopathic pulmonary fibrosis. Clinicians consulted by the company noted that they had limited knowledge on long-term survival for people with PF-ILD on nintedanib, but, for placebo, Weibull curves (based either on the frequentist or Bayesian approaches) were plausible. The company validated its choice using the following studies and registries for idiopathic pulmonary fibrosis:

- Nintedanib: registry data from the EMPIRE study (Vasakova et al. 2013) and a Greek registry (Antonίου et al. 2020) that provided longer-term data on the effect of nintedanib. The EMPIRE study (n=637) provides approximately 10 years of follow up in people across Europe (excluding the UK) with idiopathic pulmonary fibrosis. The Greek registry (n=244) reports 5-year survival data in 244 people in Greece with idiopathic pulmonary fibrosis taking nintedanib (see section 3.19).
- Standard care (placebo): Kaplan–Meier data from the treatment arms without anti-fibrotic drugs (nintedanib or pirfenidone) in the EMPIRE study, an Australian registry (Helen et al. 2016), a European registry (eurIPFreg, Guenther et al. 2018) including British patients, and a Finnish registry (Kaunisto et al. 2018, see section 3.20) were compared with the extrapolations fitted for people with idiopathic pulmonary fibrosis not treated with nintedanib. The Australian registry provides up to 4.5 years of follow up (median 2 years) for 647 people with idiopathic pulmonary fibrosis (mean age 70.9 years, 71.7% were smokers or had history of smoking). The European registry provides data on 525 people with idiopathic pulmonary fibrosis across Europe (no follow-up duration reported, mean age 65.2 years, 69.4% were smoker or had history of smoking). The Finnish registry provides data on up to 4.6

years of follow up on 453 people with idiopathic pulmonary fibrosis across Finland (mean age 73 years, 54% were smoker or had history of smoking).

The company explained that nintedanib and pirfenidone had similar treatment effect, as suggested by the network meta-analysis carried out in TA504. During the first committee meeting, the committee noted that extrapolating overall survival posed uncertainties, especially for the placebo arm; and survival for people not treated with an antifibrotic varied across registries (see section 3.20). However, given the scarcity in long-term PF-ILD data, the committee concluded that using registries for idiopathic pulmonary fibrosis to validate potential extrapolation curves generated by the Bayesian approach is reasonable.

**For the nintedanib arm, it is appropriate to use the Weibull distribution based on Bayesian analyses to model overall survival**

3.19 The committee discussed validating the extrapolations fitted for nintedanib for overall survival. It understood that the company used EMPIRE and Greek registries of idiopathic pulmonary fibrosis to select and validate the Weibull curves based on Bayesian analyses it chose in its base case (see section 3.16). The company considered that its model overpredicts survival compared with the Greek registry and the EMPIRE study after the first 2 years. The committee appreciated that the Greek registry and EMPIRE study are from countries where nintedanib may be offered to people with more severe idiopathic pulmonary fibrosis than in the UK, whereas NICE recommends treatment only to people with an FVC between 50% and 80% predicted. The committee was also aware that neither registry matched the Weibull curves well. The committee understood that these registries, although imperfect, were the only observational data available for nintedanib (see section 3.18). The committee concluded that, despite uncertainties, it was reasonable to use Weibull curves based on Bayesian analyses to model overall survival in the nintedanib arm.

**For the placebo arm, it is appropriate to use the log-logistic distribution based on the Bayesian analyses to model overall survival**

3.20 The committee discussed validating extrapolations fitted for the placebo arm based on a Bayesian approach. During its first meeting, the committee noted that placebo arms had shorter follow up than nintedanib arms (up to median 13.1 months compared with up to median 56.3 months) in the trials of idiopathic pulmonary fibrosis that the company used to inform the shape of survival curves for PF-ILD (see section 3.16). The committee appreciated that mortality data for patients randomised to placebo in the trials on idiopathic pulmonary fibrosis was not meaningfully more mature than data from INBUILD. Therefore, the committee concluded that there is more uncertainty when extrapolating survival for the placebo arm compared with the nintedanib arm. The committee noted that the Weibull curve from Bayesian analyses fitted to the placebo arm of the model had a higher death rate than the registries of idiopathic pulmonary fibrosis. The committee interpreted this to mean that by using Weibull curves (based on Bayesian analyses) the company may underestimate the survival of patients who do not take nintedanib, despite patients in registries being generally sicker than patients in trials. When comparing the registries, the company noted the survival of people not treated with antifibrotic drugs varied considerably. It considered that the Weibull curve from Bayesian analyses aligned closely with the Australian idiopathic pulmonary fibrosis registry for people not treated with antifibrotics. The committee noted that the European registry, which had a baseline mean FVC% predicted comparable to INBUILD, showed the lowest death rate among all the registries. The committee considered at its first meeting that the European registry which included patients from the UK may be the appropriate source to validate placebo arm survival estimates. But they also noted that people in the European registry had 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 who do not

take antifibrotics appears pessimistic even when using the European registry. Therefore, the committee was concerned that the modelling of survival in the placebo arm may overpredict deaths in the placebo arm.

3.21 At consultation in response to the committee's considerations on modelling and validating survival of people not treated with antifibrotics, the company presented survival curves other than a Weibull based on Bayesian analyses. Because survival in the European registry for patients treated and not treated with antifibrotics was higher than other registries or trials in idiopathic pulmonary fibrosis, the company fitted other curves for placebo and nintedanib arms aligning to the European registry. The company explained that the European registry might have included patients "without true idiopathic pulmonary fibrosis", which could explain the high survival. The company considered that the lognormal and exponential curves based on frequentist analyses matched the INBUILD placebo arm when using the European registry to validate. In addition, the company presented alternative survival curves for placebo to match the people not treated with antifibrotics from the Australian registry. This was the registry the company considered most appropriate for validation based on reports from clinical experts of the similarities in Australian and UK clinical practice noted during the first committee meeting. The company considered that both the Bayesian gamma and Bayesian log-logistic curves matched the Australian registry data for placebo. When validating against the Australian data, the company considered it most appropriate to continue to use the Bayesian Weibull curve for nintedanib.

3.22 The ERG reproduced the company's model, fitting alternative curves for placebo to match the European and Australian registries. Using Weibull (based on Bayesian analyses) for the nintedanib arm (as considered reasonable by the committee, see section 3.18) and alternative curves for the placebo arm, the ERG calculated hazard ratios for overall survival for nintedanib compared with placebo over time. The committee noted that, when using Weibull (Bayesian) analyses for nintedanib and the log-logistic

(Bayesian) for placebo, the extrapolated hazard ratio initially declines to about 0.45 (increasing treatment effect) at about 2.5 years. The committee understood that this drop in the hazard ratio was driven by INBUILD trial data; the committee recalled that the hazard ratio for overall survival reported by INBUILD was 0.78 at database lock 2 (see section 3.10). However, it considered the 0.78 figure not meaningful because it relied on short term data and non-proportional hazards (see section 3.23) and agree that the decline of hazard ratio was representative of what happens in INBUILD in the short term. The committee noted that, after the initial decline of the hazard ratios in the plots produced by the ERG, the model then predicts a slowing of that decline before the hazard ratio increases (treatment effect reduces) after about 3 years. For these reasons, and because the Australian registry was deemed to represent NHS practice, the committee agreed that the log-logistic curve (Bayesian) seemed reasonable to extrapolate overall survival in the placebo arm. The committee noted biases that may be introduced when using data from registries for idiopathic pulmonary fibrosis to validate extrapolations fitted for PF-ILD, but it considered the uncertainty acceptable. The committee concluded that the log-logistic curve based on Bayesian analyses was appropriate to model overall survival in the placebo arm.

### **Fitting independent parametric survival distributions to both arms is reasonable although uncertainties remain**

3.23 The company fitted parametric distributions to the nintedanib and placebo arms independently to extrapolate overall survival beyond the trial data in its frequentist and Bayesian approaches (sections 3.18 to 3.22). The committee understood that when fitting parametric distributions individually to each treatment arm, there is no need to assume proportionality of treatment effects over time. During its first meeting, the committee noted that the company's chosen fits resulted in ever-increasing survival benefits for nintedanib compared with placebo in the extrapolated portion of the survival curves and that the mortality data from INBUILD did not support this (see section 3.10). The committee

recommended that the company explore the proportionality of hazards in the data. At consultation, the company confirmed that the proportional hazards assumption was not met for the time to discontinuation outcome. The company explained during the second committee meeting that it had not calculated hazard ratio over time assuming non-proportional hazards. Instead, the ERG calculated the hazard ratios (see section 3.22). The committee was aware that there were uncertainties in fitting parametric distributions independently to each arm, but, because of INBUILD's short follow up, there would also be uncertainties around whether the proportional hazards assumption held, and how long it would hold for, if using a joint model to extrapolate overall survival. Considering the evidence, the committee concluded that fitting independent parametric survival curves to the 2 arms was reasonable although uncertainties remain, and it would take this into account in its decision making.

### **The modelling of exacerbations is acceptable**

3.24 The company fitted standard parametric models to extrapolate time to exacerbation (see section 3.13). It selected the exponential curve that generated risks per 3-month cycle and fixed risk of exacerbation of 1.76% and 1.12% for people receiving placebo and nintedanib, respectively. The ERG noted that the exponential model overpredicts the risk of exacerbation after approximately 8 months for both arms and sharply drops towards the end of follow up. The ERG considered this likely to increase the difference observed between nintedanib and placebo. The committee was aware that in both the company and the ERG's scenario analyses, varying the risk of exacerbation had little impact on cost effectiveness because the model does not link exacerbation to mortality (see section 3.13). During the committee's second meeting it heard from the company that exacerbations, a safety endpoint, were rare in INBUILD. The committee concluded that the extrapolation curve used to model exacerbations was acceptable although the lack of a link between exacerbations and mortality was an uncertainty of the model.

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

3.25 Simulated patients start the model in different categories of FVC% predicted, with the distribution informed by patients' baseline values in INBUILD (see section 3.13). The company used different methods to calculate transition probabilities in the placebo and nintedanib arms. For the placebo arm, the company modelled the baseline risk of decline in lung function per cycle using data from INBUILD and a multivariate mixed effects logistic regression model. Predictors of decline in lung function assessed in the model for the placebo arm included age, sex, ethnicity, underlying ILD diagnosis, FVC% predicted, and exacerbation during the analysed 3-month cycle. The company then applied an odds ratio reflecting a treatment effect to this baseline placebo risk, which it assumed was constant over time, to estimate the risk of loss of lung function in the nintedanib arm. The company estimated this treatment effect odds ratio using data from INBUILD using logistic regression, in which nintedanib treatment was the only predictor. Before the committee's first meeting, the ERG suggested that the company use a full regression model to estimate the probability of progression in the placebo arm instead of a separate model. The ERG noted that the 2 different methods produced different probabilities of loss in lung function but minimally changed cost effectiveness. This is probably because, while the absolute values differ substantially, the relative differences in probability of lung-function decline between pre- and post-exacerbation and between nintedanib and placebo do not differ substantially between the 2 methods. The ERG noted 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 3.13) 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 the lack of

link between loss of lung function and mortality was an important limitation that it would consider.

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

3.26 EQ-5D was collected in INBUILD and the company used the data to estimate the utility values in the model for each FVC% predicted health state (see section 3.12). The regression analysis to estimate the utility values resulted in a higher value in people in the 80% to 89% FVC% predicted category than in the 90% to 99% predicted category (0.7333 compared with 0.7287). Because the ERG considered this implausible, it applied a linear decline in utility from the 90% to 99% predicted and 70% to 79% predicted health states resulting 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 adjustment to utility.

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

3.27 The company modelled utility decrements for recurrent exacerbations and adverse events. When people have an exacerbation, utility drops by 0.167, which the company estimated from regression analysis using EQ-5D data collected in INBUILD. The company estimated that when people experience gastrointestinal adverse events, utility drops by 0.034. In TA379 (idiopathic pulmonary fibrosis) the company used a decrement of 0.068. But for this appraisal, the company assumed that 0.034 (half of 0.068) was plausible for PF-ILD. The company based this on a phase 3 trial in recurrent non-small-cell lung cancer using nintedanib, which estimated a decrement for grade 3/4 diarrhoea of 0.042 (Boehringer Ingelheim: data on file, 2014). The ERG was unclear on the company's rationale behind using 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, noting 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**

3.28 The cost of treatment reflects the price of nintedanib, which the company offered to the NHS at a confidential discount, the dose, and the duration of treatment. The company modelled time to stopping treatment with nintedanib using an exponential curve, consistent with TA379. This resulted in a rate for stopping nintedanib of 5.97% per month. The company considered that the exponential model did not fit the INBUILD data because the model underestimated stopping nintedanib in the first year, while from about 15 months onwards, the model appeared to overestimate stopping nintedanib. After validating the model with long-term data on the 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 and overestimate its costs. Yet the committee noted that the exponential curve dropped quickly and seemed to overestimate stopping of nintedanib compared with the INBUILD trial, and therefore underestimate its costs. In response to the technical engagement, the company provided alternative models to extrapolate time to stopping treatment with nintedanib beyond the trial period. The company considered that the Gompertz curve was closest to the Kaplan–Meier data on time to stopping treatment from INBUILD over 3 years, but it underestimated the rate of stopping 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 gave a more realistic extrapolation of time on nintedanib than other curves. The committee noted that the choice of distribution does not impact on the QALYs. During the second committee meeting, the committee discussed whether applying a stopping rule could help resolve the uncertainty and reflect TA379. It appreciated

that stopping rules generally improve cost effectiveness by minimising continued treatment and costs in people for whom a medicine is not effective. However, it had heard from clinicians that they would stop nintedanib in people whose disease did not respond to treatment (section 3.4). 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, and it would account for this in its decision making.

## End of life

### The company does not make a case for end of life criteria

3.29 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 criteria in its submission. The committee noted that people in INBUILD randomised to placebo would normally live longer than 24 months. Further evidence suggested that the median overall survival is 2 to 5 years in people with PF-ILD not receiving antifibrotic therapy (Meltzer et al. 2008; Raghu et al. 2011). The committee concluded that the criteria were not met for nintedanib for PF-ILD.

### Nintedanib is not innovative

3.30 The committee discussed innovation defined by NICE as a step-change in treatment and benefits not captured in the company's model. In discussing whether nintedanib reflects a 'step-change' in treatment for PF-ILD, the committee noted that the company considered nintedanib to be innovative because it is the only treatment licensed for people with PF-ILD. Stakeholders stated that nintedanib is the first treatment to show evidence of slowing disease progression in PF-ILD. However, the committee did not hear that there were any additional gains in health-related quality of life not already captured in the modelling. The committee deemed that nintedanib was not 'innovative'.

## There are no equality issues

3.31 Patient experts considered that inequalities exist because people with idiopathic pulmonary fibrosis can access nintedanib, while people with PF-ILD cannot. They noted that compared with people with idiopathic pulmonary fibrosis, people with PF-ILD are generally younger and more likely to be of South Asian or African-Caribbean family origin. The committee concluded that these did not constitute equality issues.

## Cost-effectiveness estimates

### Nintedanib is likely to be cost effective for PF-ILD

3.32 Because of confidential commercial arrangements for nintedanib, the cost-effectiveness estimates cannot be reported. For its second meeting, the company provided the committee with an alternative overall survival curve (log-logistic based on Bayesian analyses) for people who do not take nintedanib. That curve provided a reasonable match to the Australian registry and addressed other uncertainties including the impact of the trial not reflecting NHS treatments. The committee recognised that the company chose not to explore several uncertainties during consultation. These included incorporating a decrease in nintedanib's treatment effect in the cost-effectiveness analyses, addressing limitations in the model structure, and addressing uncertainties in modelling stopping nintedanib. Considering the evidence and uncertainties, and recognising the absence of any licensed treatment for PF-ILD, the committee noted that the incremental cost-effectiveness ratio that most likely reflected its preferred assumptions was within £20,000 to £30,000 per quality adjusted life year gained. It concluded that nintedanib is a cost-effective use of NHS resources for treating PF-ILD.

## 4 Implementation

4.1 [Section 7 of the National Institute for Health and Care Excellence \(Constitution and Functions\) and the Health and Social Care Information](#)

[Centre \(Functions\) Regulations 2013](#) requires clinical commissioning groups, NHS England and, with respect to their public health functions, local authorities to comply with the recommendations in this appraisal within 3 months of its date of publication.

- 4.2 The Welsh ministers have issued directions to the NHS in Wales on implementing NICE technology appraisal guidance. When a NICE technology appraisal recommends the use of a drug or treatment, or other technology, the NHS in Wales must usually provide funding and resources for it within 2 months of the first publication of the final appraisal document.
- 4.3 When NICE recommends a treatment ‘as an option’, the NHS must make sure it is available within the period set out above. This means that, if a patient has progressed fibrosing interstitial lung disease and the doctor responsible for their care thinks that nintedanib is the right treatment, it should be available for use, in line with NICE’s recommendations

## **5 Review of guidance**

- 5.1 The guidance on this technology will be considered for review 3 years after publication. The guidance executive will decide whether the technology should be reviewed based on information gathered by NICE, and in consultation with consultees and commentators. 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.
- 5.2 The guidance on this technology will be considered for review 3 years after publication. 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

October 2021

## **6 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]