Common Drug Review
Clinical Review Report

October 2015

Drug	nintedanib (Ofev)
Indication	For the treatment of idiopathic pulmonary fibrosis (IPF).
Listing request	For adult patients who have a diagnosis of IPF confirmed by a respirologist and a high-resolution computed tomography scan within the previous 24 months with a forced vital capacity ≥ 50% of predicted normal.
Dosage form(s)	100 mg and 150 mg capsules
NOC date	June 25, 2015
Manufacturer	Boehringer Ingelheim Canada Ltd.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

TABLE OF CONTENTS

Tables

Figures

ABBREVIATIONS

6MWT	six-minute walk test
AE	adverse event
ALAT	Latin American Thoracic Association
ATS	American Thoracic Society
CASA-Q	Cough and Sputum Assessment Questionnaire
CDR	CADTH Common Drug Review
CI	confidence interval
CrI	credible interval
DIC	deviance information criterion
DL_{CO}	diffusion capacity of lung for carbon monoxide
ERS	European Respiratory Society
EQ-5D	EuroQol 5-Dimensions Questionnaire
FEV_1	forced expiratory volume in one second
FVC	forced vital capacity
HRCT	high-resolution computed tomography
HRQoL	health-related quality of life
HTA	health technology assessment
IPF	idiopathic pulmonary fibrosis
JRS	Japanese Respiratory Society
MCID	minimal clinically important difference
NICE	National Institute for Health and Care Excellence
PGI-C	Patient Global Impression of Change questionnaire
SAE	serious adverse event
SGRQ	St. George's Respiratory Questionnaire
SOBQ	Shortness of Breath Questionnaire
SpO_2	oxygen saturation of peripheral blood
UIP	usual interstitial pneumonia
ULN	upper limit of normal
VAS	visual analogue scale

EXECUTIVE SUMMARY

Introduction

Idiopathic pulmonary fibrosis (IPF) is a rare, chronic, progressive, fibrosing, interstitial pneumonia.[1] The exact cause is unknown, but it is believed to be the result of repetitive alveolar epithelial cell injury and disregulated repair, resulting in excessive deposits of extracellular matrix proteins in the interstitial space.[2] IPF is a fatal lung disease and there is no known cure. The median survival from time of diagnosis of IPF is reported to be three to five years.[3] Disease progression is manifested by increasing respiratory symptoms, worsening pulmonary function test results, progressive fibrosis on high-resolution computed tomography, acute respiratory decline, and ultimately death.[1]

Nintedanib is an intracellular inhibitor of multiple tyrosine kinases, including fibroblast growth factor receptors 1, 2, and 3, platelet-derived growth factors alpha and beta, and vascular endothelial growth factor receptors 1, 2, and 3.[4] By blocking substrate binding, nintedanib inhibits a number of downstream signalling cascades that interfere with fibroblast proliferation, migration, and differentiation and the secretion of extracellular matrix proteins.[4] Ofev (nintedanib) is available as 100 mg and 150 mg oral capsules, and the recommended dose is 150 mg twice daily, administered approximately 12 hours apart with food.[5] The management of adverse events (AEs) associated with nintedanib may require dose reduction to 100 mg twice daily or temporary treatment interruption until the specific AE has resolved to a level that allows continuation of therapy.[5] Before nintedanib, pirfenidone (Esbriet 267 mg oral capsules) was the only pharmacologic therapy approved in Canada specifically for the treatment of IPF.

Indication under review
Ofev (nintedanib) is indicated for the treatment of IPF.
Listing criteria requested by sponsor
For adult patients who have a diagnosis of IPF confirmed by a respirologist and a high resolution computed tomography (HRCT) scan within the previous 24 months with a forced vital capacity (FVC) ≥ 50% of predicted normal.

FVC = forced vital capacity; HRCT = high resolution computed tomography; IPF = idiopathic pulmonary fibrosis.

The objective of this review is to evaluate the beneficial and harmful effects of nintedanib 100 mg and 150 mg capsules for the treatment of IPF in adult patients.

Results and Interpretation

Included studies

Two replicative, placebo-controlled, double-blind, phase 3, randomized controlled trials of 52 weeks' duration met the selection criteria for inclusion in this review: INPULSIS-1 (N = 515) and INPULSIS-2 (N = 551). The trials enrolled patients who were at least 40 years of age with a diagnosis of IPF within the past five years according to the most recent American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association guidelines,[1] with baseline forced vital capacity (FVC) ≥ 50% predicted and carbon monoxide diffusion capacity of 30% to 79% predicted. All efficacy and safety end points were analyzed using the treated set, which included all randomized patients who were documented to have taken at least one dose of study drug. The primary efficacy end point was the adjusted rate of decline in FVC over 52 weeks. The two key secondary end

points were the change from baseline to 52 weeks in the St. George's Respiratory Questionnaire (SGRQ) total score and the time to first investigator-reported acute IPF exacerbation. Key limitations are the inconsistent results observed for these key secondary end points in the included trials and the lack of data for a number of outcomes pre-specified in the clinical review protocol of importance to patients (e.g., fatigue, mental health/psychological well-being, functional capacity/productivity, and caregiver burden). Various baseline patient and disease characteristics may also limit the generalizability of the results to Canadian patients with IPF, and there was no adjustment for multiplicity for the outcomes outside the pre-specified hierarchy.

Efficacy

Key efficacy outcomes identified in the review protocol included mortality (all-cause and IPF-related), lung function (FVC decline, FVC responders), quality of life, and health care resource utilization.

The proportions of patients who died from any cause over the 52-week trial period were 4.2% and 6.7% (nintedanib) and 6.4% and 9.1% (placebo) in the INPULSIS-1 and INPULSIS-2 trials, respectively. In a pre-specified pooled analysis of both trials, 5.5% of nintedanib-treated patients compared with 7.8% of placebo-treated patients died due to any cause, corresponding with a non-statistically significant hazard ratio of 0.70 (95% confidence interval [CI], 0.43 to 1.12). Similarly, there were no statistically significant between-group differences in death due to a respiratory cause (adjudicated) or time to on-treatment death, or other composite measures such as time to death or lung transplant, or time to death or lung transplant or qualifying for a lung transplant, over the 52 weeks of the trials. The INPULSIS trials were not powered to demonstrate a statistically significant reduction in all-cause mortality; however, a numerically smaller proportion of nintedanib-treated patients died compared with placebo-treated patients, with a corresponding hazard ratio less than 1.

The primary efficacy end point in the INPULSIS trials was the adjusted rate of decline in FVC over 52 weeks, which was statistically significantly lower in the nintedanib groups in both trials. In INPULSIS-1, the rate of FVC decline was −114.65 mL per year with nintedanib compared with −239.91 mL per year with placebo, a between-group difference of 125.26 mL per year (95% CI, 77.68 to 172.84). In INPULSIS-2, the results were −113.59 mL per year (nintedanib) compared with −207.32 mL per year (placebo), a difference of 93.73 mL per year (95% CI, 44.78 to 142.68). The results of the pre-specified pooled analysis were consistent with those of the individual trials (i.e., statistically significant between-group difference of 109.94 mL per year [95% CI, 75.85 to 144.03]). The clinical expert consulted for this review advised that a difference in FVC decline between groups of 100 mL was clinically meaningful. In their reviews of pirfenidone and nintedanib, the US FDA took the position that because IPF causes a progressive decline in pulmonary function in a restrictive or scarring pattern, it is logical to monitor for change in a surrogate lung function parameter such as FVC that reflects such changes.[6] Due to the relationship between FVC and mortality trends observed in the trials for nintedanib and pirfenidone, the US FDA concurred that FVC was a clinically relevant efficacy measure in IPF.[6] As detailed in Appendix 5: VALIDITY OF OUTCOME MEASURES, change in per cent predicted FVC of ≥ 10% is predictive of mortality in IPF.[1,7-10]

The effect of nintedanib on FVC was shown to be robust, as other secondary outcomes based on FVC demonstrated consistent findings. The absolute change from baseline in FVC and in FVC per cent predicted over 52 weeks was statistically significantly lower in the nintedanib-treated groups in both INPULSIS trials compared with the placebo-treated groups. A responder analysis based on no absolute decline in the per cent predicted FVC > 5% was also statistically significant in favour of nintedanib. In both INPULSIS trials, a greater proportion of nintedanib-treated patients (52.75% and 53.19%) compared

with placebo-treated patients (38.24% and 39.27%) had an FVC response at the 5% threshold. When a threshold of no absolute decline in per cent predicted FVC > 10% was used, a statistically significant difference was observed only in INPULSIS-1 (i.e., 70.55% patients in the nintedanib group versus 56.86% patients in the placebo group had an FVC response at the 10% threshold). In INPULSIS-2, the difference between the nintedanib (69.60%) and placebo (63.93%) groups was not statistically significant. The minimal clinically important difference (MCID) for per cent predicted FVC is reported to be between 2% and 6%.[7] According to the clinical expert consulted for this review, a 5% decline in FVC may be clinically significant in clinical practice, with a 10% decline being clearly clinically significant.[8] It did not appear, however, that the reduced decline in FVC was associated with improvements in health-related quality of life (HRQoL) or in symptoms as measured by patient-reported outcomes.

HRQoL (as measured by the SGRQ total score) was a key secondary outcome in the INPULSIS trials for which the trials yielded discordant results. The SGRQ is a self-administered instrument that assesses HRQoL, with higher scores indicating worse HRQoL. In INPULSIS-1, there was no statistically significant between-group difference in the adjusted mean change in the total SGRQ score from baseline to week 52 (i.e., an increase of 4.34 points in the nintedanib group and 4.39 points in the placebo group for a difference of −0.05 points [95% CI, −2.50 to 2.40]). In contrast, in INPULSIS-2, there was a statistically significant smaller increase in the total SGRQ score at week 52 (consistent with less deterioration in HRQoL) in the nintedanib group than in the placebo group (increase of 2.80 points versus 5.48 points; difference −2.69 points [95% CI, −4.95 to −0.43]). In the pre-specified pooled analysis, no statistically significant difference between groups in total SGRQ score was shown (difference −1.43 points [95% CI, −3.09 to 0.23]). Additional analyses based on the SGRQ, including proportion of SGRQ responders (defined as ≤ –4 points change from baseline in SGRQ total score at 52 weeks), change from baseline in SGRQ domain scores, and the IPF-specific SGRQ total score, were consistent with the changes in the total SGRQ score in each trial. While the MCID for the SGRQ total score in patients with chronic obstructive pulmonary disease is 4 points,[11,12] the MCID for SGRQ total score in patients with IPF is reported to be between 6 and 13 points.[13] Therefore, all of the between-group differences in this outcome at 52 weeks do not appear to be clinically significant. Another measure of HRQoL used in the INPULSIS trials was the EuroQol 5-Dimensions Questionnaire (EQ-5D) visual analogue scale (VAS) score. An MCID for the EQ-5D has not been established in IPF patients. Various methodological issues with the EQ-5D results reported in the INPULSIS trials (e.g., missing data, failure to calculate utility or index scores for the descriptive system of the EQ-5D) put the validity of the results into question. Nonetheless, the change in EQ-5D VAS score from baseline to 52 weeks was less (suggesting less deterioration of HRQoL) with nintedanib (–2.46 and –2.52) than with placebo (–5.88 and –5.60) in both INPULSIS-1 and INPULSIS-2, respectively, although no statistical comparisons between groups were made.

No data were reported for the key efficacy outcome of health care resource utilization identified in the clinical review protocol.

The time to first investigator-reported acute IPF exacerbation was the other key secondary outcome in the INPULSIS trials. In INPULSIS-1, the proportion of patients with at least one investigator-reported acute exacerbation was similar in the nintedanib (6.1%) and placebo (5.4%) groups and there was no statistically significant between-group difference in the time to the first investigator-reported acute IPF exacerbation (hazard ratio 1.15 [95% CI, 0.54 to 2.42]). In contrast, in INPULSIS-2, the proportion of patients with at least one investigator-reported acute exacerbation was lower in the nintedanib group (3.6%) compared with the placebo group (9.6%) and there was a statistically significant increase in the time to the first acute exacerbation in favour of nintedanib (hazard ratio 0.38 [95% CI, 0.19 to 0.77]). In the pre-specified pooled analysis of investigator-reported events, the proportion of patients with events

was 4.9% (nintedanib) and 7.6% (placebo); however, the difference between groups was not statistically significant (hazard ratio 0.64 [95% CI, 0.39 to 1.05]). Other analyses based on investigator-reported acute IPF exacerbations (e.g., incidence of patients with at least one acute exacerbation over 52 weeks or the risk ratio for the incidence of acute exacerbation rate per 100 patient-years) followed a pattern of statistical significance similar to that of the time to first investigator-reported acute IPF exacerbation in both trials. The heterogeneity in findings may be due to the relatively rare occurrence of exacerbations in IPF patients in clinical trials and the difficulty in assessing and categorizing the events.[14] Another pre-specified analysis of pooled data for the time to the first adjudicated acute exacerbation (confirmed or suspected events as adjudicated by an independent adjudication committee) was conducted, and in this analysis, nintedanib was shown to be associated with a statistically significant increase in the time to first adjudicated acute IPF exacerbation (hazard ratio 0.32 [95% CI, 0.16 to 0.65]). During the adjudication process, a large number of events were deemed either to have insufficient data for adjudication or to not be true acute exacerbations. Thus, it is possible that the reason for the heterogeneous findings are due to a treatment-effect dilution resulting from the random addition of non-events that resulted in a dilution of the drug effect in the investigator-reported analysis.[15] The clinical expert consulted for this review also advised that the investigation of exacerbations may have been very centre-specific, and this may, in part, also explain the discordance.

There were limited data from the INPULSIS trials for the outcome of lung transplantation, as the proportion of patients who received a lung transplant was very low across all treatment groups and the proportion of patients qualifying for a lung transplant was similar across groups. Although lung transplantation might be a clinically meaningful end point, it is not one that is necessarily related to the natural history of IPF due to disease-independent factors such as donor availability, age, comorbidity, social support, insurance status, and differences in allocation regimens across centres.[16] There were no statistically significant differences demonstrated between treatment groups in any of the patient-reported outcomes used to measure symptoms in either INPULSIS-1 or INPULSIS-2. This included the Shortness of Breath Questionnaire (SOBQ) total score, proportion of Patient Global Impression of Change questionnaire responders, and the Cough and Sputum Assessment questionnaire cough score (symptoms and impact domains). Although patient-reported outcomes can be clinically meaningful endpoints, it appears that the none of the patient-reported outcomes used in the INPULSIS trials were validated for use in IPF (there is limited evidence for a weak correlation between SOBQ and changes in FVC).[17] Therefore it is unknown if the patient-reported outcomes can accurately measure disease symptoms (e.g., dyspnea, cough) or broader constructs such as health status in patients with IPF or if they were sensitive enough to detect treatment effects.[16] There also was a lack of measurement of other symptoms associated with IPF that were identified as being important to patients (e.g., fatigue, loss of energy, reduced physical activity, functional status, mental well-being, and caregiver burden) (APPENDIX 1: PATIENT INPUT SUMMARY).

Patient adherence with study medication was high across all treatment groups (mean adherence was > 96% for both the nintedanib and placebo treatment groups in both trials). No data were reported for many of the other efficacy outcomes identified in the review protocol (i.e., six-minute walk test, progression-free survival, mental health/psychological well-being, and functional capacity/productivity).

The results of two indirect comparisons of nintedanib, pirfenidone, and N-acetylcysteine are summarized and critically appraised in Appendix 6: SUMMARY OF INDIRECT COMPARISONS. In the absence of a head-to-head randomized controlled trial of nintedanib and pirfenidone, it was anticipated that the indirect comparisons would provide valuable comparative information, but due to differences in study populations (particularly in baseline FVC and potential IPF severity) between the nintedanib and

pirfenidone trials, the underlying assumption that the placebo groups in the trials are similar is probably invalid. In addition, the sparse networks and the variability in study durations, and therefore in the outcome assessments, were other key limitations of the analyses. Therefore, given these limitations and the corresponding high degree of uncertainty in the comparative results, no meaningful conclusions can be drawn from these analyses regarding the relative efficacy and safety of nintedanib and pirfenidone.

Harms

Most patients in both treatment groups in the INPULSIS trials experienced AEs (94.5% to 96.4% with nintedanib compared with 88.7% to 90.4% with placebo). The most frequent treatment-emergent AE in the nintedanib groups was diarrhea (61.5% and 63.2%) compared with the placebo groups (18.6% and 18.3%). More nintedanib-treated patients also received concomitant antidiarrheal therapy compared with placebo-treated patients. In addition, other gastrointestinal AEs (e.g., nausea, decreased appetite, vomiting, and weight loss) occurred more frequently with nintedanib. The proportions of patients with serious adverse events were similar between groups, and the most frequent SAE in either treatment group was IPF. Cardiac disorder AEs or serious AEs and ischemic heart disease or serious ischemic heart disease occurred in similar proportions of patients in both the nintedanib and placebo groups. Despite this, there appeared to be an imbalance in myocardial infarctions, as almost three times as many nintedanib-treated patients compared with placebo-treated patients experienced myocardial infarctions. Of these, two infarction events in the nintedanib groups and one event in the placebo groups were fatal. The clinical significance of this finding is unknown, and further observation in larger cohorts is needed.[14] There did not appear to be any clinically significant differences between groups in coagulation AEs. In both INPULSIS trials, nintedanib was associated with higher proportions of patients with elevations in alanine aminotransferase or aspartate aminotransferase, or both, that were three or more times the upper limit of normal and total bilirubin and alkaline phosphatase liver enzymes that were 1.5 or more times the upper limit of normal compared with placebo. The proportion of patients who withdrew due to AEs was higher in the nintedanib treatment group, with the most common reason for withdrawal being gastrointestinal disorders.

Conclusions

Two replicative, placebo-controlled, double-blind, phase 3, randomized controlled trials met the selection criteria for inclusion in the systematic review (INPULSIS-1 and INPULSIS-2). Data from these trials provide evidence that, compared with placebo, nintedanib reduced the rate of decline in FVC in patients with IPF over the 52-week duration of the trials. The magnitude of the reduction was sufficiently large to be clinically meaningful. There were no statistically significant differences between nintedanib and placebo in mortality (all-cause or due to adjudicated respiratory causes) in the individual trials or in the pre-specified pooled analyses of survival data. The effect of nintedanib on HRQoL (measured as the change from baseline to week 52 in the SGRQ total score) and on the time to the first investigator-reported acute IPF exacerbation was discordant between the trials. Statistically significant differences in favour of nintedanib were observed for both outcomes in INPULSIS-2 but not in INPULSIS-1. Pre-specified pooled analyses also did not demonstrate statistically significant differences between groups for either of these outcomes. There were no statistically significant between-group differences in the patient-reported outcomes that measured patient's symptoms in either trial. The most frequent AEs associated with nintedanib were gastrointestinal in nature, with most nintedanib-treated patients experiencing diarrhea of mild to moderate intensity. Dose reduction or temporarily interrupting therapy appeared to be effective at managing the AE profile of nintedanib.

TABLE 1: SUMMARY OF RESULTS

	INPULSIS-1		INPULSIS-2	
Treatment Group (Treated Set)	**NTB 150 mg N = 309**	**PL N = 204**	**NTB 150 mg N = 329**	**PL N = 219**
Mortality				
Deaths due to any cause, n (%)	13 (4.2)	13 (6.4)	22 (6.7)	20 (9.1)
HR (95% CI); *P* value	0.63 (0.29 to 1.36); 0.2880		0.74 (0.40 to 1.35); 0.2995	
Deaths due to respiratory cause, n (%)	10 (3.2)	10 (4.9)	14 (4.3)	11 (5.0)
HR (95% CI); *P* value	0.61 (0.25 to 1.47); 0.3515		0.86 (0.39 to 1.90); 0.6654	
Lung Function				
Rate of decline in FVC, adj. rate (SE)	–114.65 (15.33)	–239.91 (18.71)	–113.59 (15.73)	–207.32 (19.31)
Adj. rate diff.[a] (95% CI); *P* value	125.26 (77.68 to 172.84); < 0.0001		93.73 (44.78 to 142.68); 0.0002	
Change from BL in FVC, mean (SD)				
BL	2,756.8 (735.12)	2,844.5 (820.11)	2,672.8 (775.96)	2,512.5 (821.44)
Change from BL	–90.93 (242.70)	–201.81 (305.87)	–86.94 (283.40)	–204.04 (280.54)
Adj. mean diff.[a] (95% CI); *P* value	109.93 (71.27 to 148.59); < 0.0001		109.77 (70.92 to 148.62); < 0.0001	
Change from BL in FVC (% pred.), mean (SD)				
BL	79.47 (17.03)	80.53 (17.34)	79.99 (18.08)	78.09 (18.97)
Change from BL	–2.65 (7.06)	–5.90 (8.75)	–2.78 (8.68)	–6.07 (8.03)
Adj. mean diff.[a] (95% CI); *P* value	3.22 (2.11 to 4.33); < 0.0001		3.06 (1.87 to 4.25); < 0.0001	
FVC responders ≤ 5% decline in FVC, n (%)	163 (52.75)	78 (38.24)	175 (53.19)	86 (39.27)
OR (95% CI); *P* value	1.85 (1.28 to 2.66); 0.0010		1.79 (1.26 to 2.55); 0.0011	
FVC responders ≤ 10% decline in FVC, n (%)	218 (70.55)	116 (56.86)	229 (69.60)	140 (63.93)
OR (95% CI); *P* value	1.914 (1.32 to 2.79); 0.0007		1.286 (0.89 to 1.86); 0.1833	
Quality of Life				
SGRQ total score, mean (SD)				
BL	39.55 (17.63)	39.79 (18.48)	39.46 (20.47)	39.39 (18.65)
Change from BL to week 52	3.91 (16.0)	3.73 (15.35)	2.16 (15.20)	5.34 (16.10)
Adj. mean diff.[a] (95% CI); *P* value	–0.05 (–2.50 to 2.40); 0.9657		–2.69 (–4.95 to –0.43); 0.0197	
EQ-5D VAS score, mean (SD)				
BL	66.71 (17.42)	68.02 (16.34)	69.77 (18.84)	67.75 (16.47)
Change from BL to week 52	–2.46 (18.92)	–5.88 (19.17)	–2.52 (16.95)	–5.60 (17.67)

	INPULSIS-1		INPULSIS-2	
Treatment Group (Treated Set)	**NTB 150 mg N = 309**	**PL N = 204**	**NTB 150 mg N = 329**	**PL N = 219**
Acute IPF Exacerbations				
Pts with ≥ 1 exacerbation,[b] n (%)	19 (6.1)	11 (5.4)	12 (3.6)	21 (9.6)
HR (95% CI); *P* value	1.15 (0.54 to 2.42); 0.6728		0.38 (0.19 to 0.77); 0.0050	
Lung Transplantation				
Lung transplantation, n (%)	4 (1.3)	1 (0.5)	0 (0)	2 (0.9)
Symptoms				
SOBQ total score, adj. mean[a] (SE)	6.73 (1.11)	7.61 (1.38)	6.69 (1.07)	9.07 (1.30)
Adj. mean diff.[a] (95% CI); *P* value	–0.88 (–4.35 to 2.60); 0.6203		–2.38 (–5.68 to 0.93); 0.1587	
CASA-Q score (symptoms), adj. mean[a] (SE)	–0.76 (1.14)	–0.52 (1.40)	–0.33 (1.09)	–2.38 (1.33)
Adj. mean diff. (95% CI); *P* value	–0.24 (–3.78 to 3.30); 0.8942		2.05 (–1.31 to 5.41); 0.2326	
Harms				
Pts with ≥ 1 AE, n (%)	298 (96.4)	181 (88.7)	311 (94.5)	198 (90.4)
Pts with ≥ 1 SAE, n (%)	96 (31.1)	55 (27.0)	98 (29.8)	72 (32.9)
Pts with ≥ 1 WDAE, n (%)	65 (21.0)	22 (10.8)	58 (17.6)	33 (15.1)
Notable Harms				
Pts with any diarrhea, n (%)	190 (61.5)	38 (18.6)	208 (63.2)	40 (18.3)
Pts with any cardiac disorder AE, n (%)	30 (9.7)	19 (9.3)	34 (10.3)	26 (11.9)
Pts with MI, n (%)	5 (1.6)	1 (0.50)	5 (1.5)	1 (0.5)
Pts with ALT or AST ≥ 3 × ULN, n (%)	15 (4.9)	1 (0.5)	17 (5.2)	2 (0.9)
Pts with bilirubin total ≥ 1.5 × ULN, n (%)	5 (1.6)	1 (0.5)	19 (3.0)	2 (0.9)

% pred. = percent predicted; adj. = adjusted; AE = adverse event; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BL = baseline; CASA-Q = Cough and Sputum Assessment Questionnaire; CI = confidence interval; diff. = difference; EQ-5D = EuroQol 5-Dimensions Questionnaire; FVC = forced vital capacity; HR = hazard ratio; IPF = idiopathic pulmonary fibrosis; MI = myocardial infarction; NTB = nintedanib; OR = odds ratio; PL = placebo; Pts = patients; SAE = serious adverse event; SD = standard deviation; SE = standard error; SGRQ = St. George's Respiratory Questionnaire; SOBQ = Shortness of Breath Questionnaire; ULN = upper limit of normal; VAS = visual analogue scale; WDAE = withdrawal due to adverse event.

[a] In general included treatment and visit as fixed effects; baseline value, sex, age, and height as covariates; and treatment-by-visit and baseline-score-by-visit as interaction terms. Patient effect was assumed to be random. Please see detailed outcomes tables in Appendix 4 for specific details on each outcome.

[b] Investigator-reported events.

Note: All changes from BL are to week 52.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

1. INTRODUCTION

1.1 Disease Prevalence and Incidence

Idiopathic pulmonary fibrosis (IPF) is a rare, chronic, progressive, fibrosing, interstitial pneumonia associated with a histopathologic or radiologic pattern of usual interstitial pneumonia (UIP).[1] The exact cause is unknown, but it is believed to result from repetitive alveolar epithelial cell injury and disregulated repair in which there is uncontrolled proliferation of lung fibroblasts and differentiation of fibroblasts into myofibroblasts, which excessively deposits extracellular matrix proteins in the interstitial space.[2] The natural history of IPF is variable and is associated with an unpredictable, progressive decline;[1] IPF is often fatal and there is no known cure. The median survival from time of diagnosis is reported to be three to five years, although it is difficult to predict the rate of IPF disease progression in individual patients.[3] Disease progression is manifested by increasing respiratory symptoms, worsening pulmonary function test results, progressive fibrosis on high-resolution computed tomography (HRCT), acute respiratory decline, and ultimately death.[1]

The diagnosis of IPF requires exclusion of other known causes of interstitial lung disease, the presence of a UIP pattern on HRCT in patients not subjected to surgical lung biopsy, or specific combinations of HRCT and surgical lung biopsy patterns.[1] IPF does not have specific clinical symptoms, although it should be considered in all adult patients with unexplained chronic exertional dyspnea and it commonly presents with cough, bibasilar inspiratory crackles, and finger clubbing.[1] It typically is diagnosed in the sixth and seventh decades, is more common in men, and is associated with cigarette smoking, although environmental exposures, infectious agents, gastroesophageal reflux disease, and genetic factors have also been implicated.[3]

The true incidence and prevalence of IPF is unknown; however, based on a review of the published evidence on the global epidemiology of IPF, prevalence estimates in the US ranged from 14 to 27.9 cases per 100,000 population using narrow case-finding criteria and 42.7 to 63 per 100,000 population using broader criteria.[20] In Europe, IPF prevalence ranged from 1.25 to 23.4 cases per 100,000 population.[20] The annual incidence of IPF in the US was estimated to be 6.8 to 8.8 per 100,000 population (narrow criteria) and 16.3 to 17.4 per 100,000 population (broad criteria), whereas in Europe, the annual incidence ranged between 0.22 and 7.4 per 100,000 population.[20] A recent analysis of administrative claims data from US Medicare beneficiaries aged 65 years and older from 2000 to 2011 reported the annual incidence and prevalence of IPF to be substantially higher than previously reported.[21] While the annual incidence of IPF was relatively stable (estimated at 93.7 cases per 100,000 person-years), the annual cumulative prevalence increased from 202.1 cases per 100,000 people in 2001 to 494.5 cases per 100,000 people in 2011.[21] The observed variability in IPF incidence and prevalence may be explained by differences in diagnostic criteria, case definition, study population, and study designs.[20]

1.2 Standards of Therapy

The most recent (2015) American Thoracic Society (ATS)/European Respiratory Society (ERS)/Japanese Respiratory Society (JRS)/Latin American Thoracic Association (ALAT) clinical practice guidelines for the treatment of IPF conditionally recommends the use of nintedanib and pirfenidone for the treatment of IPF due to moderate confidence in the effect estimates.[22] Historically, a combination of prednisone, N-acetylcysteine, and azathioprine has been used in IPF; however, it was shown that this regimen actually increases mortality, hospitalizations, and serious adverse events (SAEs) compared with N-acetylcysteine or placebo.[23] The 2015 ATS/ERS/JRS/ALAT guidelines also recommend against the use of anticoagulation (warfarin); imatinib; combination prednisone, azathioprine, and N-acetylcysteine; and

selective endothelin receptor antagonist (ambrisentan), and conditionally recommends against the use of phosphodiesterase-5 inhibitor (sildenafil) and dual endothelin receptor antagonists (macitentan, bosentan) for the treatment of IPF.[22] According to the clinical expert involved in this review, use of corticosteroids as monotherapy or combination therapy with an immunomodulator (e.g., azathioprine or cyclophosphamide) or use of bosentan, interferon-gamma, etanercept, colchicine, or cyclosporine A is not recommended in IPF.[1]

Several non-pharmacological therapies are recommended for IPF patients. Such therapies include long-term oxygen therapy, especially for patients with resting hypoxia, and mechanical ventilation and pulmonary rehabilitation in select IPF patients.[1] Lung transplantation is also recommended for IPF patients; however, there are no clear data to guide precise timing and eligibility of transplantation.[1]

Before nintedanib, pirfenidone (Esbriet, 267 mg oral capsules) was the only pharmacologic therapy approved in Canada specifically for the treatment of IPF. Pirfenidone is thought to act by suppressing pulmonary inflammation and excessive collagen disposition via the inhibition of collagen synthesis induced by transforming growth factor beta and via inhibition of tumour necrosis factor alpha.[24] Pirfenidone is indicated for the treatment of mild to moderate IPF in adults.[24] Upon initiating treatment, the dose of pirfenidone should be titrated to the recommended daily dose of nine capsules per day (2,403 mg/day) over a 14-day period.[24] In April 2015, following review of a resubmission for Esbriet, the Canadian Drug Expert Committee recommended that pirfenidone be listed for the treatment of adults with mild to moderate IPF if the following clinical criteria are met: mild to moderate IPF (defined as forced vital capacity [FVC] ≥ 50% of predicted), stable disease (defined as FVC not decreased by ≥ 10% during the previous 12 months), and treatment discontinued if FVC declines by ≥ 10% within any 12-month period while receiving therapy.[25] Furthermore, the patient is to be under the care of a specialist with experience in the diagnosis and management of patients with IPF.[25]

1.3 Drug

Nintedanib is an intracellular inhibitor of multiple tyrosine kinases, including fibroblast growth factor receptors 1, 2, and 3, platelet-derived growth factors alpha and beta, and vascular endothelial growth factor receptors 1, 2, and 3.[4] Nintedanib competitively binds to the adenosine triphosphate binding pocket of these receptors, blocks substrate binding, and inhibits a number of downstream signalling cascades that interfere with fibroblast proliferation, migration and differentiation, and the secretion of extracellular matrix.[4] Ofev (nintedanib) 100 mg and 150 mg oral capsules are indicated for the treatment of IPF.[5] The recommended dose is 150 mg twice daily, administered approximately 12 hours apart with food.[5] The management of adverse events (AEs) associated with nintedanib may require dose reduction to 100 mg twice daily or temporary treatment interruption until the specific AE has resolved to levels that allow continuation of therapy.[5] Treatment with nintedanib may be resumed at the full proposed dose (150 mg twice daily) or at a reduced dose (100 mg twice daily); however, if a patient does not tolerate 100 mg twice daily, treatment with nintedanib should be discontinued.[5]

Indication under review
Ofev (nintedanib) is indicated for the treatment of IPF.
Listing criteria requested by sponsor
For adult patients who have a diagnosis of IPF confirmed by a respirologist and an HRCT scan within the previous 24 months with an FVC ≥ 50% of predicted normal.

FVC = forced vital capacity; HRCT = high-resolution computed tomography; IPF = idiopathic pulmonary fibrosis.

TABLE 2: KEY CHARACTERISTICS OF NINTEDANIB AND PIRFENIDONE

	Nintedanib	Pirfenidone
Mechanism of Action	Nintedanib is a small-molecule TKI including the receptors PDGFR alpha and beta, FGFR 1 to FGFR 3, and VEGFR 1 to VEGFR 3. Nintedanib binds competitively to the ATP binding pocket of these receptors and blocks the intracellular signalling, resulting in decreased proliferation and migration of lung fibroblasts or myofibroblasts.	Pirfenidone attenuates fibroblast proliferation, production of fibrosis-associated proteins and cytokines, and the increased biosynthesis and accumulation of extracellular matrix in response to cytokine growth factors such as TGF-beta and PDGF. It has also been shown to reduce the accumulation of inflammatory cells in response to various stimuli.
Indication[a]	Treatment of IPF	Treatment of mild to moderate IPF in adults
Route of Administration	Oral capsules	Oral capsules
Recommended Dose	150 mg b.i.d. approximately 12 hours apart with food	Dose titrated to daily dose of 9 capsules (3 × 267 mg capsules t.i.d. or 2,403 mg/day) over a 14-day period. Doses should be taken with food.
Serious Side Effects or Safety Issues	Contraindicated in patients with hypersensitivity to peanut or soya ATEs, GI AEs	Drug interaction with CYP1A2 inhibitors and other CYP isoenzymes (e.g., fluvoxamine, ciprofloxacin) Angioedema, GI AEs
Other	Safety and efficacy in pediatric patients has not been studied; not recommended for use in patients under 18 years of age	Not studied in pediatric patients and not recommended for use in this population

AE = adverse event; ATE = arterial thrombotic event; ATP = adenosine triphosphate; b.i.d. = twice daily; CYP1A2 = cytochrome P450, family 1, subfamily A, polypeptide 2; FGFR = fibroblast growth factor receptor; GI = gastrointestinal; IPF = idiopathic pulmonary fibrosis; PDGFR = platelet-derived growth factor receptor; TGF-beta = transforming growth factor beta; t.i.d. = three times daily; TKI = tyrosine kinase inhibitor; VEGFR = vascular endothelial growth factor receptor.

[a] Health Canada indication.

Source: Ofev Product Monograph,[5] Esbriet Product Monograph.[24]

2. OBJECTIVES AND METHODS

2.1 Objectives

To perform a systematic review of the beneficial and harmful effects of nintedanib 100 mg and 150 mg capsules for the treatment of IPF in adult patients.

2.2 Methods

Studies selected for inclusion in the systematic review included the pivotal studies provided in the manufacturer's submission to CADTH Common Drug Review (CDR) as well as those meeting the selection criteria presented in Table 3.

TABLE 3: INCLUSION CRITERIA FOR THE SYSTEMATIC REVIEW

Patient Population	Adult patients (≥ 18 yrs) with diagnosis of IPF Subgroups: age, gender, IPF severity (% predicted FVC), emphysema, smoking status, number of acute exacerbations
Intervention	Nintedanib 100 mg or 150 mg capsules twice daily
Comparators	• Pirfenidone • Supportive care
Outcomes	**Key efficacy outcomes:** • Mortality[a] (i.e., all-cause and IPF-related) • Lung function (i.e., FVC decline, FVC responders) • Quality of life[a] (e.g., SGRQ) • Health care resource utilization (e.g., hospitalization, ER visits, lung transplants) **Other efficacy outcomes:** • Acute exacerbations • Six-minute walk test • Progression-free survival[a] • Lung transplantation • Symptoms (e.g., dyspnea, cough, SOB, fatigue)[a] • Patient adherence • Mental health/psychological well-being[a] • Functional capacity/productivity[a] **Harms outcomes:** • AEs • SAEs • WDAEs • AEs of special interest: ATEs, GI AEs
Study Design	Published and unpublished phase 3 RCTs

AE = adverse event; ATE = arterial thrombotic event; ER = emergency room; FVC = forced vital capacity; GI = gastrointestinal; IPF = idiopathic pulmonary fibrosis; RCT = randomized controlled trial; SAE = serious adverse event; SGRQ = St. George's Respiratory Questionnaire; SOB = shortness of breath; WDAE = withdrawal due to adverse event; yrs = years.
[a] Identified in Patient Input Summary.

The literature search was performed by an information specialist using a peer-reviewed search strategy.

Published literature was identified by searching the following bibliographic databases: MEDLINE (1946–) with in-process records and daily updates via Ovid; Embase (1974–) via Ovid; and PubMed. The search strategy consisted of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concept was nintedanib (Ofev).
No filters were applied to limit the retrieval by study type. Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results.

The initial search was completed on May 15, 2015. Regular alerts were established to update the search until the meeting of the Canadian Drug Expert Committee on September 16, 2015. Regular search updates were performed on databases that do not provide alert services.

Grey literature (literature that is not commercially published) was identified by searching relevant websites from the following sections of the CADTH Grey Matters checklist (https://www.cadth.ca/resources/finding-evidence/grey-matters-practical-search-tool-evidence-based-medicine): Health Technology Assessment Agencies, Health Economics, Clinical Practice Guidelines, Drug Regulatory Approvals, Advisories and Warnings, Drug Class Reviews, and Databases (free). Google and other Internet search engines were used to search for additional Web-based materials. These searches were supplemented by reviewing the bibliographies of key papers and through contacts with appropriate experts. In addition, the manufacturer of the drug was contacted for information regarding unpublished studies.

Two CDR clinical reviewers independently selected studies for inclusion in the review based on titles and abstracts, according to the predetermined protocol. Full-text articles of all citations considered potentially relevant by at least one reviewer were acquired. Reviewers independently made the final selection of studies to be included in the review, and differences were resolved through discussion. The included studies are presented in Table 4; there were no excluded studies.

3. RESULTS

3.1 Findings From the Literature

A total of two studies were identified from the literature for inclusion in the systematic review (Figure 1). The included studies are summarized in Table 4 and described in Section 3.2. There were no excluded studies.

FIGURE 1: QUOROM FLOW DIAGRAM FOR INCLUSION AND EXCLUSION OF STUDIES

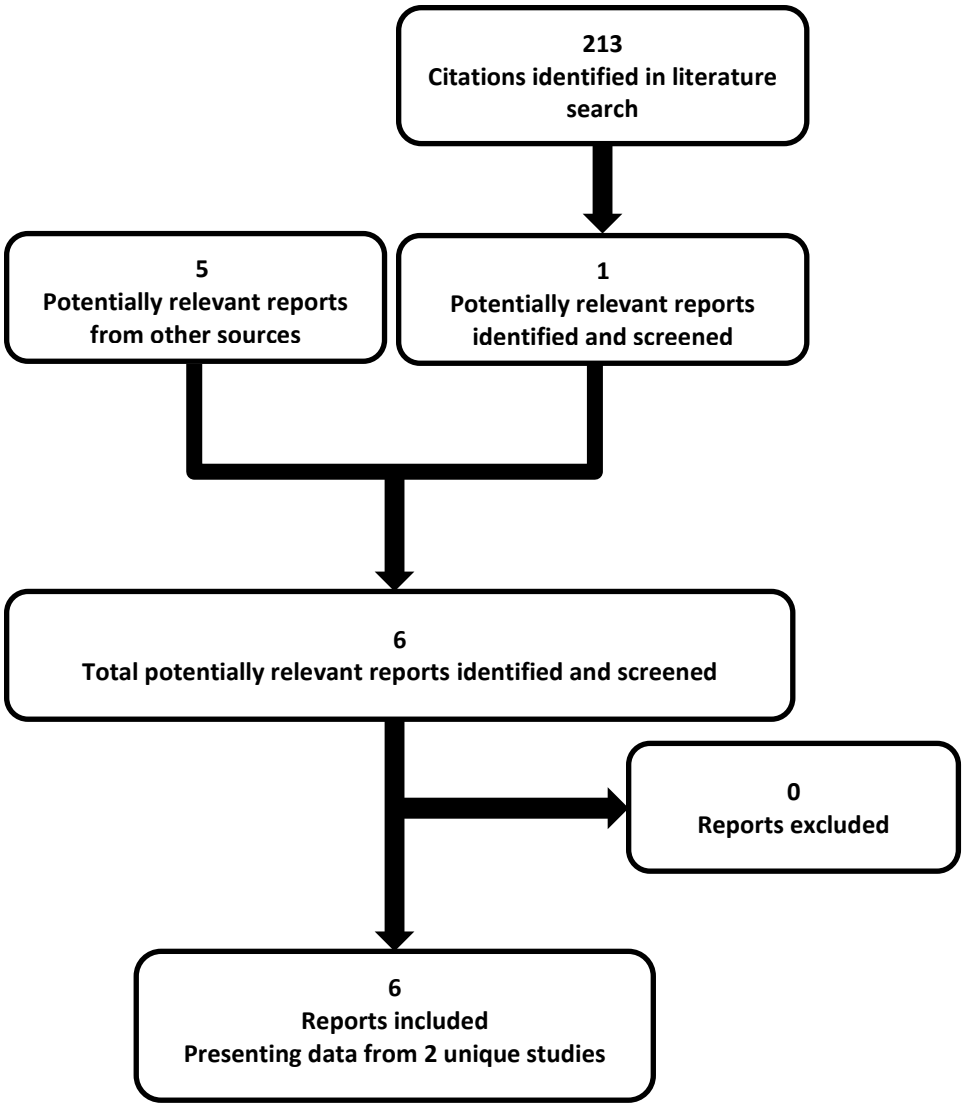


TABLE 4: DETAILS OF INCLUDED STUDIES

		INPULSIS-1 (1199.32)	INPULSIS-2 (1199.34)
DESIGNS AND POPULATIONS	**Study Design**	DB, MC, PC, phase 3 RCT	
	Locations	98 sites in 13 countries (Americas, Europe, Asia, Japan, India, Australia)	107 sites in 17 countries (Americas, Europe, Asia, Japan, India, Mexico)
	Randomized[a] (N)	515	551
	Inclusion Criteria	Pts ≥ 40 yrs, diagnosed with IPF based upon the most recent ATS/ERS/JRS/ALAT guidelines, BL DL_{CO} value 30% to 79% of predicted, and BL FVC ≥ 50% of predicted	
	Exclusion Criteria	AST, ALT, and bilirubin > 1.5 ULN, pre-bronchodilator FEV_1/FVC < 0.7, cardiac disease (MI with 6 months or unstable angina within 1 month of randomization), pts likely to have a lung transplant during trial, bleeding or thrombotic risk or prior IPF treatment within 2 to 8 weeks	
DRUGS	**Intervention**	Nintedanib 150 mg b.i.d. PO	
	Comparator	Placebo PO	
	Phase		
	Run-in	None	
	Double-blind	52 weeks	
	Follow-up	28 days	
OUTCOMES	**Primary End Point**	Annual rate of decline in FVC (mL) over 52 weeks	
	Other End Points	Change from BL to 52 weeks in SGRQ total score, time to first acute IPF exacerbation, FVC, FVC per cent predicted, proportion of FVC responders, time to death or lung transplant	
NOTES	**Publications**	Richeldi et al. 2014[14]	

ALAT = Latin American Thoracic Society; ALT = alanine aminotransferase; AST = aspartate aminotransferase; ATS = American Thoracic Society; b.i.d. = twice daily; BL = baseline; DB = double-blind; DL_{CO} = diffusion capacity of lung for carbon monoxide; ERS = European Respiratory Society; FEV_1 = forced expiratory volume in one second; FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; JRS = Japanese Respiratory Society; MC = multi-centre; MI = myocardial infarction; PC = placebo-controlled; PO = orally; pts = patients; RCT = randomized controlled trial; SGRQ = St. George's Respiratory Questionnaire; ULN = upper limit of normal; yrs = years.

[a] In INPULSIS-1, 513 patients received one dose of study medication as 2 patients were randomized in error. In INPULSIS-2, 548 patients received one dose of study medication as 3 patients were randomized in error.

Note: Three additional reports were included (manufacturer submission,[26] FDA medical review,[27] and statistical review[28]).

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

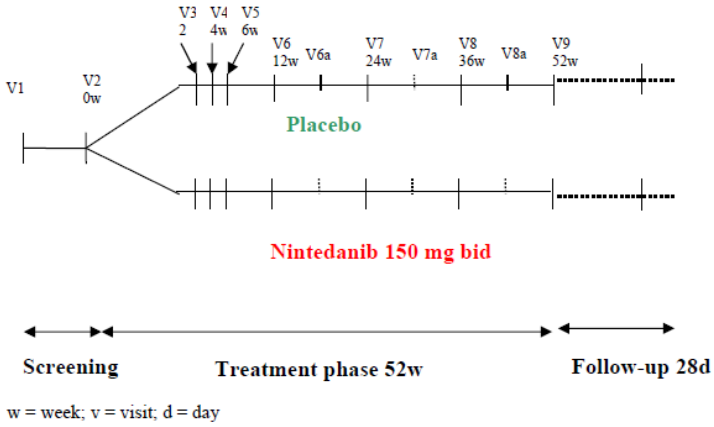
3.2 Included Studies

3.2.1 Description of studies

Two prospective, double-blind, parallel-group, multi-centre, placebo-controlled, phase 3, randomized controlled trials of 52 weeks' duration met the selection criteria for inclusion in the systematic review: INPULSIS-1 (N = 515)[18] and INPULSIS-2 (N = 551).[19] The trials were replicative trials as illustrated in Figure 2. The only differences between the trials were select geographic locations. Overall, the trials were performed at 205 sites in 24 countries in the Americas, Europe, Asia, and Australia. The primary objective of both trials was to assess the reduction in lung function decline with nintedanib 150 mg

twice daily compared with placebo in patients with IPF, as measured by a change in the annual rate of decline in FVC.

FIGURE 2: STUDY DESIGN OF INPULSIS-1 AND INPULSIS-2 TRIALS



Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

Following screening, if patients met all inclusion criteria and did not violate any exclusion criteria, they were randomized directly into the 52-week double-blind treatment phase. Patients were randomized in a 3:2 ratio using an interactive voice and Web response system. Randomization was not stratified. The randomization code was generated using a validated system involving a pseudo-random number generator and was checked by a trial-independent statistician. Patients were followed up 28 days after the last study visit or after end of treatment for patients who prematurely discontinued the trial. All patients who completed 52 weeks of treatment and the follow-up visit were eligible to enter an open-label active treatment extension trial (Study No. 1199.33). No information on the extension phase was provided.

An adjudication committee independent of the investigators evaluated the primary cause of death in reports of fatal SAEs and adjudicated all clinical reports of acute IPF exacerbations. All reported events of acute exacerbation of IPF were assessed in order to determine whether it was a confirmed or suspected case or "not IPF exacerbation." Adjudication was a three-step process; first, the clinical trial monitor provided all the information necessary and available for the evaluation; second, each and every fatal event or reported case of acute IPF exacerbation was adjudicated separately by each of the committee members; third, the committee discussed the cases at a central adjudication meeting and through email exchange for the last cases. Disagreements were adjudicated by using one of the following methods:

1. Cause of death: The independent adjudication committee discussed the case in an effort to achieve a consensus on the primary cause of death.
2. Assessment of acute IPF exacerbation: The independent adjudication committee discussed the case in an effort to achieve a consensus on the diagnosis of confirmed or suspected acute exacerbation of IPF, or on the rejection of this diagnosis.
3. If the independent adjudication committee believed that additional supporting information was required to determine the primary cause of death or to assess the diagnosis of acute IPF exacerbation, the chairman requested the additional data from the sponsor.

In situations where agreement by all members was not possible, the primary cause of death or the assessment of acute IPF exacerbation was determined by a majority vote. The committee reviewed cases in a blinded manner with respect to treatment groups. In addition, an independent data monitoring committee was responsible for reviewing unblinded safety data, monitoring the benefit/risk ratio, and addressing any specific safety issues during the conduct of the trials. Chest HRCT results, lung biopsies, and spirometry measurements were reviewed and rated centrally.

3.2.2 Populations

a) Inclusion and exclusion criteria

Inclusion criteria were identical between the two trials as shown in Table 4. Both trials included patients 40 years of age or older with a diagnosis of IPF within the past five years based upon the most recent ATS/ERS/JPS/ALAT guidelines.[1] Eligible patients were also required to have a chest HRCT performed within 12 months of the first study visit and to have a combination of an HRCT pattern and, if available, a surgical biopsy pattern consistent with a diagnosis of IPF. The HRCT images and lung biopsy specimens, if available, were reviewed centrally by a single radiologist and a single pathologist. To qualify for entry based on chest HRCT if a surgical lung biopsy was not available, Criteria A and B and C, or Criteria A and C, or Criteria B and C had to be met as follows:

- Criterion A: Definite honeycomb lung destruction with basal and peripheral predominance
- Criterion B: Presence of reticular abnormality and traction bronchiectasis consistent with fibrosis with basal and peripheral predominance
- Criterion C: Atypical features are absent, specifically nodules and consolidation. Ground-glass opacity, if present, is less extensive than reticular opacity pattern.

In addition, patients were required to have a diffusion capacity of lung for carbon monoxide (DL_{CO}) corrected for hemoglobin of 30% to 79% of the predicted value and a FVC of 50% or more of the predicted value.

Patients were excluded if certain laboratory parameters were abnormal (i.e., aspartate aminotransferase, alanine aminotransferase, and bilirubin > 1.5 upper limit of normal [ULN]), if there was airway obstruction (i.e., pre-bronchodilator forced expiratory volume in one second (FEV_1)/FVC < 0.7), or if a patient was likely to have a lung transplant during the trial. Patients with cardiac disease (e.g., myocardial infarction or unstable angina) or bleeding risk (e.g., genetic predisposition; patients requiring fibrinolysis, anticoagulation, or antiplatelet therapy; active gastrointestinal bleeding or ulcers; abnormal coagulation parameters) or at thrombotic risk (e.g., inherited predisposition, history of thrombotic events such as stroke or transient ischemic attacks) were also excluded. Patients who received previous treatment with nintedanib for more than four weeks, other investigational therapy in the past eight weeks, N-acetylcysteine or prednisone > 15 mg/day or equivalent within two weeks, or pirfenidone, azathioprine, cyclophosphamide, or cyclosporine A within the past eight weeks were also excluded.

b) Baseline characteristics

There was a predominance of male patients in both trials (i.e., 77.8% to 81.2% across treatment groups) as detailed in Table 5. Caucasian patients formed the largest proportion (49.2% to 66.2%), followed by Asian patients (20.1% to 39.3%). Mean and median values for age, which ranged from 66 to 68 years, were similar across all treatment groups. The distribution of patients by age categories showed some imbalances, with higher proportions of younger (< 65 years) and elderly patients (≥ 75 years) and a lower proportion of patients from 65 to < 75 years in the nintedanib group than in the placebo group.

Most patients were ex-smokers or current smokers, while approximately one-quarter had never smoked.

Mean baseline efficacy variables were generally comparable across treatment groups. Almost all patients (> 97%) had a radiologic assessment consistent with UIP. Overall, 38.2% to 41.3% of patients had emphysema at baseline. More patients in INPULSIS-2 (23.7% to 25.5%) than in INPULSIS-1 (16.2% to 19.4%) had lung biopsy specimens available at baseline. In INPULSIS-1, a mean difference of approximately 88 mL was noted in baseline FVC between nintedanib (2,756.8 mL) and placebo (2,844.5 mL). The manufacturer investigated the difference in means through a post hoc box-plot of the FVC distribution in both treatment groups, and it was found that there were more outliers in the placebo group (1.5%) than in the nintedanib group (0.6%). The clinical expert consulted for this review advised that a difference of this magnitude would not be considered to be clinically significant. Baseline St. George's Respiratory Questionnaire (SGRQ) total scores were comparable across both treatment groups.

3.2.3 Interventions

The intervention was nintedanib 150 mg or 100 mg soft gelatin capsules administered orally twice daily. The dose of nintedanib 150 mg twice daily was based on the results of the phase 2 TOMORROW safety and dose-finding trial (Study No. 1199.30).[29] In the case of AEs requiring dose reduction, nintedanib 100 mg twice daily was to be given at the request of the investigator.

TABLE 5: SUMMARY OF BASELINE CHARACTERISTICS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Gender, n (%)				
Male	251 (81.2)	163 (79.9)	256 (77.8)	171 (78.1)
Female	58 (18.8)	41 (20.1)	73 (22.2)	48 (21.9)
Race, n (%)				
Caucasian	198 (64.1)	135 (66.2)	162 (49.2)	113 (51.6)
Black	0	0	2 (0.6)	0
Asian	66 (21.4)	41 (20.1)	128 (38.9)	86 (39.3)
Missing	45 (14.6)	28 (13.7)	37 (11.2)	19 (8.7)
Age				
Mean (SD)	66.9 (8.4)	66.9 (8.2)	66.4 (7.9)	67.1 (7.5)
Median (min, max)	68.0 (42, 85)	68.0 (45, 87)	66.0 (42, 89)	68.0 (42, 88)
< 65 yrs, n (%)	119 (38.5)	71 (34.8)	139 (42.2)	74 (33.8)
≥ 65 yrs to < 75 yrs, n (%)	130 (42.1)	102 (50.0)	133 (40.4)	114 (52.1)
≥ 75 yrs, n (%)	60 (19.4)	31 (15.2)	57 (17.3)	31 (14.2)
BMI (kg/m^2)				
Mean (SD)	28.6 (4.5)	28.1 (4.6)	27.6 (4.6)	27.2 (4.5)
Smoking History, n (%)				
Never smoked	71 (23.0)	51 (25.0)	103 (31.3)	71 (32.4)
Ex-smoker	217 (70.2)	144 (70.6)	218 (66.3)	139 (63.5)
Current smoker	21 (6.8)	9 (4.4)	8 (2.4)	9 (4.1)
Time Since IPF Diagnosis, Yrs				
Mean (SD)	1.7 (1.4)	1.6 (1.4)	1.6 (1.3)	1.6 (1.27)
Median (min, max)	1.3 (0, 5.2)	1.2 (0, 5.0)	1.3 (0, 4.9)	1.2 (0.1, 5.0)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Centrilobular Emphysema, n (%)				
No	191 (61.8)	126 (61.8)	193 (58.7)	131 (59.8)
Yes	118 (38.2)	78 (38.2)	136 (41.3)	88 (40.2)
Lung Biopsy Specimen, n (%)	60 (19.4)	33 (16.2)	84 (25.5)	52 (23.7)
Radiologic Assessment, n (%)				
Not evaluable	0 (0)	0 (0)	0 (0)	0 (0)
Consistent with UIP	301 (97.4)	198 (97.1)	321 (97.6)	215 (98.2)
Possible UIP	8 (2.6)	6 (2.9)	8 (2.4)	4 (1.8)
Definitely not UIP	0 (0)	0 (0)	0 (0)	0 (0)
Missing	0 (0)	0 (0)	0 (0)	0 (0)
Systemic CS Therapy, n (%)	68 (22.0)	43 (21.1)	68 (20.7)	46 (21.0)
FVC, mL				
Mean (SD)	2,756.8 (735.1)	2,844.5 (820.1)	2,672.8 (776.0)	2,619.0 (787.3)
FEV_1:FVC, %				
Mean (SD)	81.5 (5.4)	80.8 (6.1)	81.8 (6.3)	82.4 (5.7)
DL_{CO}, mmol/min/kPa				
Mean (SD)	4.0 (1.2)	4.0 (1.1)	3.7 (1.2)	3.7 (1.3)
SpO_2, %				
Mean (SD)	95.9 (2.0)	95.9 (1.9)	95.8 (2.6)	95.7 (2.1)
SGRQ Total Score[a]				
Mean (SD)	39.6 (17.6)	39.8 (18.5)	39.5 (20.5)	39.4 (18.7)

BMI = body mass index; CS = corticosteroid; DL_{CO} = diffusion capacity of lung for carbon monoxide; FEV_1 = forced expiratory volume in one second; FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; NTB = nintedanib; PL = placebo; SD = standard deviation; SGRQ = St. George's Respiratory Questionnaire; SpO_2 = oxygen saturation of peripheral blood; UIP = usual interstitial pneumonia; yrs = years.

[a] In INPULSIS-1, the total score on the SGRQ was available for 298 patients in the NTB group and 202 patients in the PL group; in INPULSIS-2, the SGRQ total score was available for 326 patients in the NTB group and 217 patients in the PL group.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report,[19] Richeldi et al. 2014.[14]

Patients experiencing AEs requiring temporary interruption were permitted to restart trial medication at a reduced dose (100 mg twice daily) or at the same dose if the AE was not determined to be drug-related. In the case of severe toxicity or when the reduced dose was again not tolerated, drug treatment was to be stopped and was not to be restarted. If the reduced dose was well tolerated, re-escalation was possible within four weeks. The comparator was matched placebo soft gelatin capsules administered orally twice daily (i.e., 150 mg or 100 mg as applicable).

Prohibited therapies included pirfenidone, any other experimental IPF therapy, and fibrinolysis or full-dose therapeutic anticoagulation (e.g., vitamin K antagonists, dabigatran, heparin, hirudin, etc.) or high-dose antiplatelet therapy. If any of the anticoagulation therapies were medically indicated during the trial, a four-week washout of the trial medication was to be observed before their use. Restricted therapies were not allowed during the first six months of treatment except in the case of acute IPF exacerbations; these therapies included azathioprine, cyclophosphamide, cyclosporine A, N-acetylcysteine, and prednisone at > 15 mg/day (or > 30 mg every two days or equivalent dose of other oral corticosteroid).

Concomitant therapy with prednisone (if taken at a steady dose of ≤ 15 mg/day [or ≤ 30 mg every two days] or equivalent dose of other oral corticosteroid if dose had been stable for eight weeks) or an antidiarrheal drug such as loperamide as clinically needed for diarrhea (which was a frequent AE associated with nintedanib identified in the preceding phase 2 trial[29]) was permitted. In the case of acute IPF exacerbations, any treatments (e.g., high-dose prednisone, azathioprine, cyclophosphamide, cyclosporine A, or N-acetylcysteine) could be freely initiated or increased at the investigator's discretion, except for pirfenidone. In the case of deterioration after six months, all treatments as listed above, except pirfenidone, could be freely initiated or increased at the investigator's discretion. A ≥ 10 % decrease in the absolute value of FVC per cent predicted or a ≥ 15% decrease in DL_{CO} per cent predicted was to be considered to be deterioration (i.e., disease progression). Prophylactic low-dose heparin or heparin flush as needed for maintenance of an indwelling intravenous device (e.g., enoxaparin 4,000 IU daily) as well as prophylactic use of antiplatelet therapy (e.g., acetylsalicylic acid up to 325 mg/day, clopidogrel 75 mg/day, or equivalent doses of other antiplatelet therapy) were also allowed.

As detailed in Table 6, more nintedanib-treated patients used antidiarrheal medication (34.0% and 35.9%) compared with placebo-treated patients (3.9% and 6.4%). More placebo-treated patients used antitussives (10.3% and 18.3%) and oxygen (10.3% and 11.4%) compared with nintedanib-treated patients (6.8% and 15.2%, and 7.8% and 9.1%, respectively). Approximately one-fifth of all patients (14.2% to 23.6%) used proton pump inhibitors and H_2-receptor antagonists, which is not unexpected given the strong association between gastroesophageal reflux disease and IPF.[1]

TABLE 6: CONCOMITANT ON-TREATMENT THERAPIES (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Pts With ≥ 1 Concomitant On-Treatment Therapy, n (%)	219 (70.9)	125 (61.3)	252 (76.6)	154 (70.3)
Antidiarrheal	105 (34.0)	8 (3.9)	118 (35.9)	14 (6.4)
P-glycoprotein inhibitors	87 (28.2)	67 (32.8)	105 (31.9)	72 (32.9)
Systemic corticosteroid	77 (24.9)	62 (30.4)	66 (20.1)	61 (27.9)
PPI and H_2-receptor antagonists	73 (23.6)	29 (14.2)	70 (21.3)	48 (21.9)
Antioxidants/expectorants	41 (13.3)	32 (15.7)	57 (17.3)	44 (20.1)
Bronchodilator	40 (12.9)	38 (18.6)	49 (14.9)	48 (21.9)
Oxygen	24 (7.8)	21 (10.3)	30 (9.1)	25 (11.4)
Antitussive	21 (6.8)	21 (10.3)	50 (15.2)	40 (18.3)
Antiplatelet	20 (6.5)	12 (5.9)	21 (6.4)	15 (6.8)
Anticoagulant	14 (4.5)	12 (5.9)	16 (4.9)	16 (7.3)
Thrombolytic	0 (0)	0 (0)	7 (2.1)	3 (1.4)
P-glycoprotein inducers	9 (2.9)	1 (0.5)	6 (1.8)	11 (5.0)
Immunosuppressant	6 (1.9)	5 (2.5)	6 (1.8)	6 (2.7)
Phosphodiesterase 5 inhibitor	3 (1.0)	1 (0.5)	2 (0.6)	2 (0.9)
Endothelin receptor antagonist	1 (0.3)	1 (0.5)	1 (0.3)	2 (0.9)

H_2 = histamine$_2$; NTB = nintedanib; PL = placebo; PPI = proton pump inhibitor; pts = patients.
Note: Concomitant therapies were started between first and last trial drug intake; a patient may be counted in more than one category.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

3.2.4 Outcomes

a) Mortality

All mortality end points were measured as time to death. The survival analyses were a secondary end point in the INPULSIS trials and included time to death, time to death due to respiratory cause (adjudicated), and time to on-treatment death, in addition to the composite outcomes of time to death or lung transplant, and time to death or lung transplant or qualifying for lung transplant, all over 52 weeks. A patient was considered to qualify for a lung transplant if he or she fulfilled the criteria: FVC < 45% predicted or DL_{CO} < 30% predicted or oxygen saturation of peripheral blood (SpO_2) < 88% at rest, at sea level (to be adapted for other elevations).

b) Lung function

The annual rate of decline in FVC (measured in mL over 52 weeks) was the primary end point in the INPULSIS trials. Spirometry was done according to ATS/ERS criteria. Each test was conducted in triplicate and the best result was reported. All spirometric measurements were performed on machines provided by the sponsor, and the results electronically transmitted and confirmed by central review, with training and ongoing feedback provided for the site investigators.

Further analyses on FVC were secondary end points in the trials and included absolute change from baseline in FVC (mL) and in per cent predicted FVC over 52 weeks, the proportion of FVC responders using a 5% threshold at 52 weeks, defined as patients with absolute decline in FVC per cent predicted ≤ 5% and with an FVC evaluation at 52 weeks and also using 10% threshold at 52 weeks, defined as patients with absolute decline in FVC per cent predicted ≤ 10% and with an FVC evaluation at 52 weeks.

c) Quality of life

There were two key secondary end points pre-specified in the INPULSIS trials. The first was the change from baseline in the SGRQ total score at 52 weeks. The SGRQ is a self-administered questionnaire that assesses health-related quality of life (HRQoL). It comprises three domains (symptoms, activity, and impact). The total score and the score for each domain range from 0 to 100, with higher scores indicating worse HRQoL. A minimal clinically important difference (MCID) in the SGRQ total score in IPF has been reported to be 6 to 13 points, as detailed in Appendix 5: VALIDITY OF OUTCOME MEASURES. Other secondary end points based on the SGRQ included the proportion of SGRQ responders at 52 weeks (defined as absolute change from baseline at 52 weeks in SGRQ total score ≤ –4 points, although as noted previously an MCID in SGRQ total score in IPF is reported to be 6 to 13 points), change from baseline in SGRQ domains over 52 weeks, and change from baseline in an IPF-specific version of the SGRQ total score (i.e., an exploratory calculation from SGRQ data) over 52 weeks.

Another measure of HRQoL included as a secondary end point in the INPULSIS trials was the change from baseline in the EuroQol 5-Dimensions Questionnaire (EQ-5D) up to 52 weeks. The EQ-5D is a generic, utility-based measure of HRQoL and consists of the EQ-5D descriptive system that assesses health status based on one-day recall on five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and the visual analogue scale (VAS), which asks patients to rate their health on that day on a vertical line based on a range of “worst imaginable health state” equal to 0 and “best imaginable health state” equal to 100. The change from baseline in EQ-5D to 52 weeks was reported only as the change in EQ-5D VAS scores in the INPULSIS trials. The MCID for EQ-5D VAS in IPF is unknown.

d) Acute exacerbations

The second key secondary end point was the time to first acute IPF exacerbation (days) over 52 weeks. Acute exacerbations were reported as AEs by the investigator and were defined as follows:

Otherwise unexplained clinical features including all of the following:

- unexplained worsening or development of dyspnea within 30 days:
  - new diffuse pulmonary infiltrates on chest X-ray, or new HRCT parenchymal abnormalities with no pneumothorax or pleural effusion (new ground-glass opacities) since the last visit
- exclusion of infection as per routine clinical practice and microbiological studies
- exclusion of alternative causes as per routine clinical practice, including the following:
  - left heart failure
  - pulmonary embolism
  - identifiable cause of acute lung injury.

e) Patient-reported outcomes

A number of patient-reported outcomes were included as secondary end points in the INPULSIS trials, one of which was the change from baseline in the University of California, San Diego Shortness of Breath Questionnaire (SOBQ) over 52 weeks. The SOBQ is a 24-item measure that assesses self-reported shortness of breath while patients perform a variety of activities of daily living.[30] It has been validated in patients with moderate to severe lung disease (e.g., chronic obstructive pulmonary disease, cystic fibrosis, and post-lung transplant),[30] and there is limited evidence for a weak correlation between the SOBQ and changes in FVC.[17] Patients rate the severity of shortness of breath during 21 activities associated with varying levels of exertion and answer three additional questions relating to limitations, fear of harm, and fear of shortness of breath for a total sum of responses from 24 items ranging in score from 0 to 120, with higher scores indicating worse results. The MCID for the SOBQ in patients with IPF is unknown.

The change from baseline in cough impact and cough symptom of the Cough and Sputum Assessment Questionnaire (CASA-Q) score over 52 weeks was also a secondary end point. The CASA-Q includes 20 items assessing four concepts (cough symptoms, cough impact, sputum symptoms, and sputum impact) that are rated based on frequency and severity.[31] The CASA-Q has been validated in three countries in patients with chronic (obstructive) bronchitis.[31] A higher score is associated with fewer symptoms due to cough or less impact due to cough; however, the range of scores for the CASA-Q and the MCID in patients with IPF were not reported.

Last, another secondary patient-reported outcome end point was the proportion of patients who were considered to be Patient Global Impression of Change (PGI-C) questionnaire responders at 52 weeks. Responders were defined as being “Very much better,” “Much better,” “A little better,” or “No change” at 52 weeks. No additional details were provided on the other categories of the PGI-C that would apply to patients who worsened, nor was information provided on how the PGI-C was administered (e.g., recall period). The remaining secondary end points were change from baseline in SpO_2 and in DL_{CO} at rest over 52 weeks.

f) Patient adherence

Adherence with study medication was assessed by the investigator by physical count of returned trial medication at each study visit. Adherence was calculated as the number of capsules actually taken multiplied by 100 and divided by the number of capsules that should have been taken. Treatment interruptions due to AEs were taken into account in the calculation by subtracting the duration of interruption due to the AEs.

There were no outcomes reported in the INPULSIS trials for health care resource utilization, six-minute walk test, progression-free survival, mental health/psychological well-being, or functional capacity/productivity as identified in the protocol for the systematic review.

g) Safety

Safety end points included AEs, vital signs, physical examination, weight, and 12-lead electrocardiogram (Japanese centres only). Other end points included pharmacokinetic end points and exploratory end points (e.g., pharmacogenomic evaluation, biomarker evaluation).

3.2.5 Statistical analysis

The sample size for each INPULSIS trial was calculated on the basis of providing 90% power to detect a between-group difference of 100 mL in the primary end point of the annual rate of FVC decline. According to the manufacturer, the difference of 100 mL in the annual rate of decline in FVC was selected in consultation with the Committee for Medicinal Products for Human Use for the European Medicines Agency. The Committee advised that 100 mL was considered clinically relevant because it represents 50% of the mean FVC decline observed in recent clinical trials in IPF. It is also approximately 3% of predicted FVC, which is considered clinically relevant. On the basis of data from the phase 2 TOMORROW trial,[29] the standard deviation for the change in FVC from baseline was assumed to be 300 mL in both groups. Assuming that it would not be possible to evaluate data for 2% of patients, the sample size was calculated as 194 patients in the placebo group and 291 patients in the nintedanib group for a two-group t-test at a one-sided significance level of 2.5%. Since the primary analysis was based on a random coefficient regression model that included adjustment for several variables and took into account information across time, it was expected that the power would be greater than the 90% calculated for the t-test.

For the survival analyses, a log-rank test was used to compare treatment groups and a Cox proportional hazards model adjusted for sex, age, and height was used to determine hazard ratios and 95% confidence intervals (CIs). Since the number of deaths was expected to be low, the protocol pre-specified that survival analyses would also be performed on pooled data from the trials as an additional supportive analysis. The statistical methods used were the same, but with the addition of trial as a fixed effect or covariate in the models. For the pooled analyses, the Cox model was also adjusted for trial. The annual rate of decline in FVC was analyzed using random coefficient regression with fixed effects for trial, treatment, sex, age, and height, and random effect of patient-specific intercept and time. The treatment effect was determined by using estimated slopes for each study group (on the basis of the time-by-treatment interaction term from the mixed model). The decrease in FVC was assumed to be linear within each patient over 52 weeks. All available FVC values from baseline to week 52 were used in the primary model, including FVC measurements at the follow-up visit for patients who discontinued the study medication prematurely and did not complete the study visits through week 52. The same statistical methods were used for the pre-specified pooled analysis. Significance tests were two-sided, with an alpha value of 0.05.

The absolute change from baseline in FVC or per cent predicted FVC over 52 weeks was analyzed using a mixed model for repeated measures, with treatment and visit as fixed effects; baseline value, sex, age, and height as covariates; and treatment-by-visit and baseline-by-visit as interaction terms. The patient effect was assumed to be random. The treatment effect was obtained using the model estimates of the change from baseline in FVC at 52 weeks for each treatment group (using the treatment-by-visit interaction term from the mixed model).

The proportion of FVC responders (i.e., patients with absolute decline in FVC percent predicted ≤ 5% or 10% at week 52) was analyzed using logistic regression with treatment term and age, sex, height, and baseline FVC per cent predicted as covariates. Odds ratios and 95% CIs were calculated. Patients with no FVC value at week 52 were considered to be nonresponders.

Changes from baseline in SGRQ total and domain scores over 52 weeks were analyzed using a mixed model for repeated measures, with treatment and visit as fixed effects, baseline score as a covariate, and treatment-by-visit and baseline-by-visit as interaction terms. The patient effect was assumed to be random. In the pre-specified pooled analysis, changes from baseline in SGRQ total and domain scores over 52 weeks were analyzed using a mixed model for repeated measures, with fixed effects for trial, treatment, visit, treatment by visit, baseline SGRQ total score, and baseline SGRQ total score by visit; and random effect for patient. All other patient-reported outcomes were analyzed over 52 weeks using a mixed model for repeated measures, as described for SGRQ total score.

For the analysis of the time to first acute IPF exacerbation, the time-to-event information was used to produce Kaplan–Meier plots for each treatment group and was primarily analyzed using the log-rank test, with hazard ratios and CIs obtained using the Cox's proportional hazards model adjusted for sex, height, and age. A pre-specified sensitivity analysis of pooled data based on the time to the first adjudicated acute IPF exacerbation (i.e., confirmed or suspected as determined by an independent adjudication committee) was also conducted. For acute IPF exacerbations, the risk in a treatment group of having at least one IPF acute exacerbation was defined as the number of patients with at least one exacerbation divided by their total time at risk, with its asymptotic two-sided 95% CI and associated two-sided *P* value. Risk ratio was calculated as the ratio of risk of exacerbation rates in both treatment groups. The log of the risk ratio was assumed to follow a normal distribution with variance equal to the sum of the reciprocals of the number of patients with at least one exacerbation in each treatment group.

A hierarchical procedure was used to demonstrate the superiority of nintedanib over placebo for the primary and key secondary end points (i.e., change from baseline in the SGRQ total score and time to first investigator-reported acute IPF exacerbation) to account for multiple comparisons. The consecutive steps of the hierarchy were considered only if the previous step was significant at the one-sided 2.5% level and the results were in favour of nintedanib. Given the positive outcome on the primary end point in both trials, it was possible to continue the pre-specified confirmatory testing procedure for the key secondary end points in both trials. According to the Clinical Study Reports for the INPULSIS trials,[18,19] there were two alternative hierarchies of end points with different orders of the key secondary end points for US and European Union submissions. For US registration, the first key secondary end point was the time to first investigator-reported acute IPF exacerbation and the second was the change from baseline in SGRQ total score, whereas for European and other regulators, including Health Canada, the opposite order was used (i.e., the first key secondary end point was SGRQ and time to first investigator-reported acute IPF exacerbation was second).

In INPULSIS-2, the result for time to first acute IPF exacerbation and for change from baseline in SGRQ total score at week 52 were both statistically significant, hence a formal confirmatory testing could proceed for both key secondary end points. In INPULSIS-1, neither the result for time to first acute exacerbation nor the change from baseline in SGRQ total score at week 52 was statistically significant, hence the statistical testing was made only on a nominal basis for the second key secondary end point.

Various methods for handling missing data were undertaken in the INPULSIS trials. The primary analysis allowed for missing data, assuming they were missing at random and thus were not imputed. Sensitivity

analyses were conducted to demonstrate the robustness of the results of the main analysis and in particular for missing data. Data collected from patients who discontinued and continued to attend study visits were included in the primary analysis. For secondary end points, the statistical model used for the analysis of continuous end points (e.g., change from baseline in SGRQ total score) allowed for missing data, assuming they were missing at random. If there were more than 2, 4, or 6 missing items in the SGRQ symptoms, activity, or impacts component, respectively, the component was set to missing. Sensitivity analyses, including sensitivity to missing data handling, were conducted for the SGRQ primary analysis. In the analyses of the time-to-event end points, missing or incomplete data were managed by standard survival analysis techniques (i.e., censoring). Missing or incomplete start or end dates for IPF acute exacerbations were imputed according to the manufacturer's standards for AEs. In the analysis of the categorical secondary end points (SGRQ, FVC, PGI-C responders), missing data were imputed using the worst case analysis (i.e., set to the category "nonresponder"). For the SOBQ, missing questions were not imputed and the score was set to "missing." No imputation was planned for EQ-5D. For CASA-Q, if there was a missing item, it was imputed by the average score across completed items for that scale (rounded to the nearest score), if the minimum number of items was reached. No imputation was planned for missing questionnaires.

All safety analyses were descriptive only.

a) Analysis populations

The efficacy and safety analyses were based on the treated set, which included randomized patients who were dispensed trial medication and were documented to have taken at least one dose of study drug. There were no intention-to-treat or per-protocol analyses.

3.3 Patient Disposition

More patients in the nintedanib treatment groups (25.2% and 23.7%) prematurely discontinued the trials than in the placebo treatment groups (17.6% and 20.1%); however, for both treatment groups the main reason for discontinuation was AEs.

TABLE 7: PATIENT DISPOSITION

	INPULSIS-1		INPULSIS-2	
	NTB	PL	NTB	PL
Screened, N	718		794	
Randomized, N	309	206	331	220
Not Treated,[a] N	0	2	2	1
Treated (Treated Set), N (%)	309 (100)	204 (100)	329 (100)	219 (100)
Discontinued, N (%)	78 (25.2)	36 (17.6)	78 (23.7)	44 (20.1)
AE	65 (21.0)	24 (11.8)	62 (18.8)	35 (16.0)
Protocol non-compliance	2 (0.6)	3 (1.5)	2 (0.6)	1 (0.5)
Lost to follow-up	0	0	0	1 (0.5)
Pt refusal to continue study drug	9 (2.9)	7 (3.4)	11 (3.3)	6 (2.7)
Other	2 (0.6)	2 (1.0)	3 (0.9)	1 (0.5)
Safety, N	309	204	329	219

AE = adverse event; NTB = nintedanib; PL = placebo; pt = patient.

[a] In INPULSIS-1, 2 patients were randomized in error, and in INPULSIS-2, 3 patients were randomized in error. In each case the patient was randomized by the interactive voice and Web response system but did not meet eligibility criteria. These patients did not receive the study drug and were not included in the treated set in either trial.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

3.4 Exposure to Study Treatments

Overall, the mean duration of exposure was similar between the nintedanib and placebo treatment groups in both trials, with the majority of patients (65.3% to 73.1%) across treatment groups exposed to the study drug for 6 to 12 months (Table 8).

TABLE 8: EXPOSURE TO STUDY MEDICATION (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Duration of Exposure, months				
Mean (SD)	10.28 (3.34)	10.88 (2.83)	10.29 (3.40)	10.78 (2.81)
Median (min, max)	11.93 (0.0, 12.5)	11.93 (0.5, 13.0)	111.93 (0.0, 12.7)	11.93 (0.0, 13.1)
Duration of Exposure Category, n (%)				
≤ 3 months	27 (8.7)	12 (5.9)	31 (9.4)	10 (4.6)
> 3 to ≤ 6 months	18 (5.8)	8 (3.9)	16 (4.9)	9 (4.1)
> 6 to ≤ 12 months	202 (65.4)	143 (70.1)	215 (65.3)	160 (73.1)
> 12 to ≤ 13 months	62 (20.1)	41 (20.1)	67 (20.4)	39 (17.8)
> 13 months	0 (0.0)	0 (0.0)	0 (0)	1 (0.5)
Total Dose,[a] g				
Mean (SD)	88.50 (31.36)	98.29 (26.12)	88.69 (31.47)	97.67 (25.97)
Median (min, max)	107.10 (0.3, 114.0)	109.20 (4.5, 118.5)	107.10 (0.3, 116.4)	108.90 (0.3, 119.7)
Dose Intensity,[b] %				
Mean (SD)	93.45 (11.95)	98.87 (4.55)	93.91 (10.58)	99.00 (4.40)
Median (min, max)	100.00 (35.8, 102.2)	100.00 (69.0, 106.2)	100.00 (56.9, 104.3)	100.00 (62.7, 107.3)

NTB = nintedanib; PL = placebo; SD = standard deviation.
[a] For each patient, total dose was exposure (days) multiplied by actual dose (g).
[b] Dose intensity was defined as the amount of drug received over the trial divided by the amount of drug that would have been administered if there had been no dose reductions or treatment interruptions.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

More patients in the nintedanib treatment groups in both trials required one or more dose reductions (26.5% and 29.2%) compared with patients who received placebo treatment (4.9% and 2.7%) (Table 9). There did not appear to be a clear pattern with regard to the time to first dose reduction. The main reason for dose reductions across all treatment groups was a related AE. Overall, 75.1% to 77.7% of nintedanib-treated patients had a last dose consistent with that recommended (i.e., 150 mg twice daily).

TABLE 9: SUMMARY OF DOSE CHANGES (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Number of Dose Reductions Per Pt, n (%)				
0	227 (73.5)	194 (95.1)	233 (70.8)	213 (97.3)
1	75 (24.3)	10 (4.9)	88 (26.7)	6 (2.7)
2	6 (1.9)	0 (0)	7 (2.1)	0 (0)
> 2	1 (0.3)	0 (0)	1 (0.3)	0 (0)
Number of Pts With ≥ 1 Dose Reduction, n (%)	82 (26.5)	10 (4.9)	96 (29.2)	6 (2.7)
Time (days) to First Dose Reduction, n (%)				
≤ 31 days	10 (3.2)	1 (0.5)	12 (3.6)	0 (0)
> 31 days to ≤ 61 days	12 (3.9)	1 (0.5)	14 (4.3)	3 (1.4)
> 61 days to ≤ 92 days	7 (2.3)	0 (0)	13 (4.0)	0 (0)
> 92 days to ≤ 183 days	31 (10.0)	4 (2.0)	32 (9.7)	1 (0.5)
> 183 days	22 (7.1)	4 (2.0)	25 (7.6)	2 (0.9)
Total Number of Dose Reductions[a]	90	10	105	6
Reasons for Dose Reduction,[a] n (%)	0 (0)	0 (0)	0 (0)	0 (0)
Acute exacerbation	86 (95.6)	9 (90.0)	101 (96.2)	5 (83.3)
Related AE	3 (3.3)	1 (10.0)	3 (2.9)	1 (16.7)
Unrelated AE	1 (1.1)	0 (0)	1 (1.0)	0 (0)
Other				
Number of Pts With ≥ 1 Dose Increase, n (%)	20 (6.5)	5 (2.5)	20 (6.1)	2 (0.9)
Number of Pts With Last Dose of 150 mg b.i.d., n (%)	240 (77.7)	199 (97.5)	247 (75.1)	215 (98.2)

AE = adverse event; b.i.d. = twice daily; NTB = nintedanib; PL = placebo; pt = patient.
[a] All reductions were included. A patient may have had several reductions and therefore more than one reason for dose reduction. Percentage is based on the total number of dose reductions.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

There was also a higher proportion of nintedanib-treated patients (25.6% and 21.9%) who required one or more treatment interruptions compared with placebo-treated patients (11.8% and 8.2%) (Table 10). There did not seem to be a clear pattern regarding the time to first interruption. The main reasons for interruptions were related AEs in the nintedanib groups and unrelated AEs in the placebo groups. As also noted in Section 3.3 and Table 7, more nintedanib-treated patients prematurely discontinued treatment compared with placebo-treated patients.

TABLE 10: SUMMARY OF TREATMENT INTERRUPTIONS AND PREMATURE DISCONTINUATIONS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Number of Interruptions Per Pt, n (%)				
0	230 (74.4)	180 (88.2)	257 (78.1)	201 (91.8)
1	57 (18.4)	22 (10.8)	56 (17.0)	17 (7.8)
2	17 (5.5)	2 (1.0)	13 (4.0)	1 (0.5)
> 2	5 (1.6)	0 (0)	3 (0.9)	0 (0)
Number of Pts With ≥ 1 Interruption, n (%)	79 (25.6)	24 (11.8)	72 (21.9)	18 (8.2)
Time (days) to First Interruption, n (%)				
≤ 31 days	19 (6.1)	2 (1.0)	13 (4.0)	2 (0.9)
> 31 days to ≤ 61 days	8 (2.6)	0 (0)	10 (3.0)	3 (1.4)
> 61 days to ≤ 92 days	11 (3.6)	5 (2.5)	9 (2.7)	1 (0.5)
> 92 days to ≤ 183 days	23 (7.4)	5 (2.5)	25 (7.6)	4 (1.8)
> 183 days	18 (5.8)	12 (5.9)	15 (4.6)	8 (3.7)
Total Number of Interruptions[a]	108	26	91	19
Reasons for Interruptions[a], n (%)				
Acute exacerbation	0 (0)	0 (0)	1 (1.1)	0 (0)
Related AE	70 (64.8)	11 (42.3)	66 (72.5)	3 (15.8)
Unrelated AE	20 (18.5)	11 (42.3)	18 (19.8)	9 (47.4)
Other	18 (16.7)	4 (15.4)	6 (6.6)	7 (36.8)
Number of Pts With Premature Treatment Discontinuations,[b] n (%)	78 (25.2)	36 (17.6)	78 (23.7)	44 (20.1)
Time (days) to Premature Treatment Discontinuation, n (%)				
≤ 31 days	7 (2.3)	5 (2.5)	10 (3.0)	5 (2.3)
> 31 days to ≤ 61 days	10 (3.2)	3 (1.5)	8 (2.4)	1 (0.5)
> 61 days to ≤ 92 days	10 (3.2)	5 (2.5)	13 (4.0)	4 (1.8)
> 92 days to ≤ 183 days	18 (5.8)	7 (3.4)	16 (4.9)	9 (4.1)
> 183 days	33 (10.7)	16 (7.8)	31 (9.4)	25 (11.4)

AE = adverse event; NTB = nintedanib; PL = placebo; pt = patient.
[a] Only interruptions > 7 consecutive days were reported in the electronic case report form. A patient may have had more than one reason for interruption. Percentage was based on the total number of interruptions.
[b] Premature discontinuations were always permanent.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

3.5 Critical Appraisal

3.5.1 Internal validity

- The methods used for randomization (i.e., interactive voice and Web response system and computer-generated randomization lists) and methods of allocation concealment appear to be appropriate to avoid selection bias. Based on the findings from the TOMORROW phase 2 dose-finding trial[29] that gastrointestinal AEs and premature discontinuations occurred more frequently with the dose of nintedanib 150 mg twice daily, randomization in the INPULSIS trials was unbalanced with a 3:2 ratio of nintedanib versus placebo.

- Despite the use of matched placebo capsules, it is possible that the treatment-emergent AE profile of nintedanib could have resulted in unblinding, especially due to the relatively high frequency of diarrhea associated with nintedanib (61.5% to 63.2%) as compared with placebo (18.3% to 18.6%). Similarly, the disproportionate use of antidiarrheals by nintedanib-treated patients could also have been an unplanned source of unblinding (i.e., 34.0% to 35.9% of nintedanib-treated patients compared with 3.9% to 6.4% of placebo-treated patients used antidiarrheal medication).
- Adequate sample sizes appear to have been determined in both INPULSIS trials based on a priori power calculations. The sample size for each trial was calculated on the basis of being able to detect a between-group difference of 100 mL in the primary end point of the annual rate of FVC decline. The manufacturer did not provide any rationale to support the between-group difference of 100 mL used in the sample size calculation. Nonetheless, the clinical expert consulted for this review concurred that this difference was clinically meaningful.
- Although more nintedanib-treated patients (23.7% to 25.2%) prematurely discontinued the trials, the magnitude of the difference was not substantial when compared with placebo patients (17.6% to 20.1%). Measures were put in place to minimize the impact of missing data (as detailed in Section 3.2.5: Statistical Analysis); however, for certain outcomes (e.g., SGRQ responders, PGI-C responders, and EQ-5D VAS scores), the large amount of missing data at week 52 may call the validity of the results into question. Missing data for SGRQ responders and PGI-C responders were imputed as the worst case scenario (nonresponders), so, overall, the treatment effect may have been underestimated for these outcomes. In contrast, missing results for EQ-5D were not imputed, so the large amount of missing data at week 52 is problematic; however, no comparisons between groups were made between EQ-5D VAS scores, and only descriptive statistics were reported.
- In both INPULSIS trials, the efficacy and safety analyses were based on the treated set, defined as randomized patients who were dispensed trial medication and were documented to have taken at least one dose of study drug. This is not a true intention-to-treat population but rather a modified intention-to-treat population; however, the treated set comprised more than 99% of the randomized patients in each trial, so this is unlikely to have affected the validity of the trial outcomes. No per-protocol analyses were done in either trial.
- A hierarchical stepwise testing procedure was used in order to demonstrate superiority of nintedanib over placebo for the primary and two key secondary end points to account for multiple comparisons. The consecutive steps of the hierarchy were considered only if the previous step was significant in favour of nintedanib at the one-sided 2.5% level. Any end points outside this hierarchy would be at risk of type 1 error and should be considered as exploratory.
- Sensitivity analyses and pooled analyses of data from INPULSIS-1 and INPULSIS-2 reported in this clinical review report were all pre-specified analyses. For the key secondary outcome of time to first acute IPF exacerbation, the results were to be based on investigator-reported events. An additional pre-specified analysis was based on pooled data from the two trials for the time to the first adjudicated acute IPF exacerbation (i.e., including only exacerbations adjudicated as confirmed or suspected by the independent adjudication committee). Each case of an acute IPF exacerbation was adjudicated separately by each of the committee members and then discussed at a central adjudication meeting and through email exchange. This implies that the decision of the independent adjudication committee was based on consensus; however, this was not explicitly stated.

3.5.2 External validity

- The INPULSIS trials were conducted at 205 sites in 24 countries; however, there appear to have been only three Canadian sites and only 14 Canadian patients enrolled between the two trials.
- Various baseline patient and disease characteristics may affect the generalizability of the results of the INPULSIS trials to Canadian patients with IPF. There appears to be an underrepresentation of

female patients and an overrepresentation of male patients (i.e., 77.8% to 81.2% across treatment groups). Although IPF occurs more commonly in males, the clinical expert consulted for this review advised that the distribution of male patients to female patients is more likely in the order of 60:40.

- The inclusion criteria for the INPULSIS trials specified that patients were required to have a baseline DL_{CO} value of 30% to 79% of predicted and FVC ≥ 50% of predicted. Given the high unmet need in this patient population, there is potential for nintedanib to be used in more severe patients with IPF (e.g., FVC < 50% predicted), in patients with probable or possible IPF, or in patients with early IPF or at risk of developing IPF, in the hope of impacting disease progression. There is also potential for nintedanib to be used in combination with pirfenidone due to the different mechanism of action of the drugs. To date, there is no efficacy information on combination use; however, based on a randomized, double-blind, phase 2 dose-escalation study[32] of nintedanib alone and when added to pirfenidone, there may be lower exposure (i.e., maximum plasma concentration and area under the curve) of nintedanib and its metabolites in the presence of pirfenidone, whereas nintedanib appears to have no effect on the pharmacokinetics of pirfenidone. An exclusion criterion for entry into the INPULSIS trials was if a patient was likely to have a lung transplant during the trial. This probably accounts, in part, for the low rate of lung transplants that occurred during the trials and makes it difficult to generalize the results to patients who are candidates for lung transplantation.
- Potential subgroups of interest identified in the protocol for the systematic review included age, gender, IPF severity (per cent predicted FVC), emphysema, smoking status, and number of acute exacerbations at baseline. There were no pre-specified subgroup analyses conducted in the INPULSIS trials, with the exception of a subgroup analysis of the primary and key secondary end points and select safety analyses in Japanese versus non-Japanese patients.
- Both INPULSIS trials were placebo-controlled trials; however, with the availability of pirfenidone, a direct head-to-head active comparator trial would yield useful information on the relative efficacy and safety of the two drugs. Differences in recruitment criteria between the pivotal nintedanib and pirfenidone trials (which support that patients included in the INPULSIS trials may have had less severe disease) preclude the ability to make meaningful comparisons between the trials.[33,34]
- There are no data for a number of outcomes specified in the protocol for the systematic review (e.g., health care resource utilization, six-minute walk test, progression-free survival, mental health/psychological well-being, and functional capacity/productivity). There also was no specific measurement of fatigue, loss of energy, or reduced physical activity in the trials, which are important disease manifestations as identified by patients (see APPENDIX 1: PATIENT INPUT SUMMARY).
- Inconsistent results were observed between the two INPULSIS trials for both key secondary end points (i.e., SGRQ total score and time to first acute exacerbation). The differences could not be explained by differences in baseline characteristics between the two trials.[14] With regard to investigator-reported acute IPF exacerbations, it was speculated that the relatively rare occurrence of exacerbations in patients with IPF and the difficulty of assessing and categorizing exacerbations could have been responsible for some of the heterogeneity in the findings.[14] The latter may be true given the different results obtained in the analysis of adjudicated acute IPF exacerbations.
- The duration of the trials (52 weeks) was not adequate to show a survival benefit with nintedanib. It is expected that an adequately powered trial with longer treatment duration would be required to demonstrate a survival benefit with nintedanib.
- The change from baseline in EQ-5D to 52 weeks appears to have been reported as change only in EQ-5D VAS scores. While it appears that the full EQ-5D was administered (because the distributions of the five dimensions of the EQ-5D descriptive system were included in the Clinical Study Reports for the INPULSIS trials),[18,19] it does not appear that the ratings on each dimension were used to

produce overall utility or index scores. There was no explanation provided of the underlying methodology used to administer, score, or analyze the change in EQ-5D.

- SGRQ responders were defined as the proportion of patients who had an absolute change from baseline to 52 weeks in the SGRQ total score ≤ –4 points. The MCID of 4 points has been established for patients with chronic obstructive pulmonary disease; however, in patients with IPF, the MCID is reported to be much higher (i.e., 6 to 13 points) (see Appendix 5: VALIDITY OF OUTCOME MEASURES).

3.6 Efficacy

Only those efficacy outcomes identified in the review protocol are reported below in Section 0, Table 3. See Appendix 4: DETAILED OUTCOME DATA for detailed efficacy data.

3.6.1 Mortality

Survival analyses were a secondary end point in the INPULSIS trials, and all were based on time-to-event analyses. The results for the individual trials are reported in Table 13 and the pre-specified pooled survival analysis is reported in Table 14. The proportion of patients who died from any cause over the 52-week trial period was 4.2% and 6.7% (nintedanib) and 6.4% and 9.1% (placebo) in the INPULSIS-1 and INPULSIS-2 trials, respectively. In the pre-specified pooled analysis, 5.5% of patients treated with nintedanib and 7.8% of patients treated with placebo died due to any cause over the trial duration, corresponding with a hazard ratio of 0.70 (95% CI, 0.43 to 1.12; *P* = 0.1399). Similarly, there were no statistically significant between-group differences in death due to a respiratory cause (adjudicated) or time to on-treatment death. There were also no between-group differences in the composite measures of time to death or lung transplant, or time to death or lung transplant or qualifying for a lung transplant, over the 52 weeks of the trials.

3.6.2 Lung function

In both INPULSIS-1 and INPULSIS-2, the adjusted rate of decline in FVC (the primary efficacy end point), was statistically significantly lower in the nintedanib group than in the placebo group (Table 21). In INPULSIS-1, the rate was −114.65 mL per year (nintedanib) compared with −239.91 mL per year with placebo, representing a between-group difference of 125.26 mL per year (95% CI, 77.68 to 172.84); *P* < 0.0001. In INPULSIS-2, the rate of FVC decline was −113.59 mL per year (nintedanib) compared with −207.32 mL per year (placebo), representing a difference of 93.73 mL per year (95% CI, 44.78 to 142.68; *P* = 0.0002). The results of the pre-specified pooled analysis were consistent with those in the individual trials (i.e., between-group difference of 109.94 mL per year [95% CI, 75.85 to 144.03; *P* < 0.0001]). The clinical expert consulted for this review concurred that a difference in FVC decline of 100 mL was considered to be clinically meaningful.

Other measures of lung function assessed as secondary end points were consistent with the results of decline in FVC. As detailed in Table 22, the absolute change from baseline in FVC over 52 weeks was statistically significantly lower in the nintedanib-treated groups (–95.07 mL and –95.26 mL) than in the placebo-treated groups (–205.00 mL and –205.03 mL) (*P* < 0.0001 for the adjusted mean difference in each trial). Similarly, the absolute change from baseline in FVC per cent predicted was statistically significantly lower in the nintedanib-treated groups (–2.76 and –3.09) than in the placebo-treated groups (–5.98 and –6.15) (*P* < 0.0001 for the adjusted mean difference in each trial) (Table 23).

A responder analysis based on no absolute decline in FVC per cent predicted greater than 5% or 10% demonstrated that the results using a 5% threshold were consistent with other results for FVC; however, the trials were inconsistent when a 10% threshold was used (Table 24). In both INPULSIS trials, a

statistically significantly greater proportion of nintedanib-treated patients (52.75% and 53.19%), compared with placebo-treated patients (38.24% and 39.27%) had an FVC response defined as no absolute decline in the percentage of predicted FVC > 5% at week 52 (*P* = 0.001 for both trials). In INPULSIS-1, when patients were defined as no absolute decline of > 10% at week 52, a statistically significantly greater proportion of patients in the nintedanib group (70.55%) compared with the placebo group (56.86%) achieved a response (*P* = 0.0007). In contrast, in INPULSIS-2, 69.60% (nintedanib) compared with 63.93% (placebo) of patients achieved a response at the 10% threshold and the difference between groups was not statistically significant. According to the clinical expert consulted for this review, a 5% decline in FVC may be clinically significant, with 10% being clearly clinically significant.

3.6.3 Quality of life

The SGRQ total score was a key secondary outcome in the INPULSIS trials. Results for SGRQ total score for the individual trials as well as the pre-specified pooled analysis are presented in Table 15. Additional analyses on SGRQ such as the proportion of SGRQ responders at 52 weeks (defined as absolute change from baseline at 52 weeks in SGRQ total score ≤ –4 points, although as noted previously an MCID in SGRQ total score in IPF is reported to be 6 to 13 points), change from baseline in SGRQ domains over 52 weeks, and change from baseline in the IPF-specific version of the SGRQ total score are reported in Table 16 and Table 17.

In INPULSIS-1, there was no significant between-group difference in the adjusted mean change in the total SGRQ score from baseline to week 52 (increase of 4.34 points in the nintedanib group and 4.39 points in the placebo group; difference −0.05 points [95% CI, −2.50 to 2.40; *P* = 0.9657]). In contrast, in INPULSIS-2, there was a statistically significantly smaller increase in the total SGRQ score at week 52 (consistent with less deterioration in HRQoL) in the nintedanib group than in the placebo group (increase of 2.80 points versus 5.48 points; difference −2.69 points [95% CI, −4.95 to −0.43; *P* = 0.0197]). In the pre-specified pooled analysis, there was no significant difference between the treatment groups in total SGRQ score (difference between nintedanib and placebo of −1.43 points [95% CI, −3.09 to 0.23; *P* = 0.2081]). The additional SGRQ analysis, including SGRQ responders (defined as change from baseline in SGRQ total score ≤ –4 points at 52 weeks), change from baseline in SGRQ domain scores, and IPF-specific SGRQ total score were consistent with the changes in the total SGRQ score in each trial (Table 16 and Table 17). Of note, all results for the change from baseline in SGRQ total score were below the MCID of 6 to 13 points in patients with IPF.

In both INPULSIS trials, the change (reduction) from baseline in EQ-5D VAS score at week 52 was less with nintedanib (–2.46 and –2.52) when compared with placebo (–5.88 and –5.60) in INPULSIS-1 and INPULSIS-2, respectively (Table 29). This implies less worsening in HRQoL with nintedanib; however, no statistical comparisons between groups were made. Caution must be exercised in the interpretation of these results; as noted previously, there was a large amount of missing data at week 52 for this outcome.

No data were reported for the key efficacy outcome of health care resource utilization identified in the review protocol.

3.6.4 Other efficacy outcomes

a) Acute exacerbations

Time to first IPF exacerbation was another key secondary outcome in the INPULSIS trials. Results for the time to first exacerbation in the individual trials as well as in the pre-specified pooled analysis (both based on investigator-reported events) are provided in Table 18. In INPULSIS-1, there was no significant

difference between the nintedanib and placebo groups in the time to the first acute IPF exacerbation (hazard ratio 1.15 [95% CI, 0.54 to 2.42; *P* = 0.6728]). The proportion of patients with at least one investigator-reported acute exacerbation was similar in the nintedanib (6.1%) and placebo (5.4%) groups. In contrast, in INPULSIS-2, there was a significant increase in the time to the first acute exacerbation in the nintedanib group compared with the placebo group (hazard ratio 0.38 [95% CI, 0.19 to 0.77; *P* = 0.005]). The proportion of patients with at least one investigator-reported acute exacerbation was lower in the nintedanib group (3.6%) compared with the placebo group (9.6%).

In the pre-specified pooled analysis of investigator-reported events there was no significant difference between treatment groups in the time to first acute exacerbation (hazard ratio 0.64 [95% CI, 0.39 to 1.05; *P* = 0.0823]) (Table 18). The proportion of patients with at least one investigator-reported acute exacerbation was 4.9% (nintedanib) and 7.6% (placebo). On the other hand, in another pre-specified analysis of pooled data on the time to the first adjudicated acute exacerbation (confirmed or suspected events as adjudicated by the independent adjudication committee), it was shown that nintedanib was associated with a statistically significant increase in the time to first acute IPF exacerbation (hazard ratio 0.32 [95% CI, 0.16 to 0.65; *P* = 0.001]) (Table 19). Overall, 63 patients had at least one investigator-reported exacerbation event (31 nintedanib-treated patients and 32 placebo-treated patients experienced a total of 69 exacerbations). As detailed in Table 19, during the adjudication process, two events were deemed to have insufficient data for adjudication and 31 events were found to not be acute exacerbations. There were more exacerbations considered to not be acute exacerbations in the nintedanib group (n = 19) compared with the placebo group (n = 12) and more suspected exacerbations in the placebo group (n = 20) compared with the nintedanib group (n = 8) following adjudication by the independent adjudication committee (Table 19).

The incidence of patients with at least one acute IPF exacerbation over the 52-week period was consistent with the results of the investigator-reported time to first acute IPF exacerbation (Table 20). The risk ratio for the incidence of exacerbation rate per 100 patient-years was not statistically significant in INPULSIS-1 (1.17 [95% CI, 0.56 to 2.46; *P* = 0.6793]); however, it was statistically significant in INPULSIS-2 (0.38 [95% CI, 0.19 to 0.77; *P* = 0.007]).

b) Lung transplantation

The proportion of patients who received a lung transplant was low across all treatment groups in the INPULSIS trials (0% to 1.3% with nintedanib and 0.5% to 0.9% with placebo) as shown in Table 27. This is not surprising, as an exclusion criterion for entry into the INPULSIS trials was if a patient was likely to have a lung transplant during the trial. The proportion of patients qualifying for a lung transplant based on FVC < 45% predicted or DL_{CO} < 30% predicted or SpO_2 < 88% at rest (sea level) was similar between the treatment groups (11.0% to 14.6% with nintedanib and 12.7% to 16.9% with placebo). No statistical comparisons were made.

c) Symptoms

There were no statistically significant differences between treatment groups in any of the other patient-reported outcomes reported in either INPULSIS-1 or INPULSIS-2 after 52 weeks (Table 16 and Table 17). This included the SOBQ total score, the proportion of PGI-C responders, and the CASA-Q cough score (symptoms and impact domains).

d) Patient adherence

Patient adherence with study medication was high across all treatment groups (Table 28). Overall, mean compliance (defined as the actual number of capsules taken multiplied by 100 divided by the total

number of capsules that should have been taken) was > 96% for both the nintedanib and placebo treatment groups in both trials.

e) SpO_2 and DL_{CO}

There were no statistically significant between-group differences in either the absolute change from baseline in SpO_2 (Table 25) or DL_{CO} (Table 26).

No data were reported for a number of other efficacy outcomes identified in the review protocol (i.e., six-minute walk test, progression-free survival, mental health/psychological well-being or functional capacity/productivity).

3.7 Harms

Only those harms identified in the review protocol (Section 2.2) are reported in Table 11. See Appendix 4: DETAILED OUTCOME DATA for detailed harms data.

TABLE 11: HARMS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Deaths				
Pts with AE Leading to Death,[a] n (%)	12 (3.9)	10 (4.9)	25 (7.6)	21 (9.6)
AEs				
Pts With ≥ 1 AE, n (%)	298 (96.4)	181 (88.7)	311 (94.5)	198 (90.4)
Most Frequent AEs (> 10% in Any Group), n (%)				
Diarrhea	190 (61.5)	38 (18.6)	208 (63.2)	40 (18.3)
Nausea	70 (22.7)	12 (5.9)	86 (26.1)	16 (7.3)
Nasopharyngitis	39 (12.6)	34 (16.7)	48 (14.6)	34 (15.5)
Cough	47 (15.2)	26 (12.7)	38 (11.6)	31 (14.2)
Progression of IPF	31 (10.0)	21 (10.3)	33 (10.0)	40 (18.3)
Bronchitis	36 (11.7)	28 (13.7)	31 (9.4)	17 (7.8)
Upper respiratory tract infection	28 (9.1)	18 (8.8)	30 (9.1)	24 (11.0)
Dyspnea	22 (7.1)	23 (11.3)	27 (8.2)	25 (11.4)
Decreased appetite	26 (8.4)	14 (6.9)	42 (12.8)	10 (4.6)
Vomiting	40 (12.9)	4 (2.0)	34 (10.3)	7 (3.2)
Weight loss	25 (8.1)	13 (6.4)	37 (11.2)	2 (0.9)
SAEs				
Pts With ≥ 1 SAE, n (%)	96 (31.1)	55 (27.0)	98 (29.8)	72 (32.9)
Most Frequent SAEs (≥ 2% in Any Group), n (%)				
Pneumonia	5 (1.6)	5 (2.5)	18 (5.5)	11 (5.0)
Squamous cell carcinoma	1 (0.3)	4 (2.0)	–	–
IPF	20 (6.5)	11 (5.4)	22 (6.7)	28 (12.8)
Pulmonary hypertension	5 (1.6)	6 (2.9)	6 (1.8)	3 (1.4)
Respiratory failure	0 (0)	3 (1.5)	2 (0.6)	5 (2.3)
Result of SAEs				
Fatal	12 (3.9)	10 (4.9)	25 (7.6)	21 (9.6)
Life-threatening	4 (1.3)	2 (1.0)	5 (1.5)	4 (1.8)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Disability or incapacity	2 (0.6)	1 (0.5)	2 (0.6)	1 (0.5)
Requires hospitalization	85 (27.5)	45 (22.1)	85 (25.8)	58 (26.5)
Prolonged hospitalization	3 (1.0)	5 (2.5)	8 (2.4)	10 (4.6)
Other	20 (6.5)	16 (7.8)	15 (4.6)	15 (6.8)
WDAEs				
Pts With ≥ 1 WDAE, n (%)	65 (21.0)	22 (10.8)	58 (17.6)	33 (15.1)
Most Common Reason by SOC for Withdrawal (≥ 2% in Any Group), n (%)	26 (8.4)	3 (1.5)	21 (6.4)	2 (0.9)
Gastrointestinal disorders	12 (3.9)	10 (4.9)	8 (2.4)	18 (8.2)
Respiratory, thoracic, and mediastinal disorders	10 (3.2)	1 (0.5)	8 (2.4)	1 (0.5)
Investigation results[b]	5 (1.6)	4 (2.0)	2 (0.6)	3 (1.4)
Cardiac disorders	8 (2.6)	3 (1.5)	2 (0.6)	1 (0.5)
General disorders and administration site conditions				

AE = adverse event; IPF = idiopathic pulmonary fibrosis; NTB = nintedanib; PL = placebo; pts = patients; SAE = serious adverse event; SOC = system organ class; WDAE = withdrawal due to adverse event.
[a] The most common AE leading to death in both INPULSIS-1 and INPULSIS-2 was IPF.
[b] Investigation results includes results of clinical laboratory tests, radiologic tests, physical examination, and physiologic tests (e.g., weight decreased, hepatic enzyme increased, etc.).
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report,[19] Richeldi et al. 2014.[14]

3.7.1 Adverse events

Most patients in both treatment groups experienced AEs (94.5% to 96.4% with nintedanib compared with 88.7% to 90.4% with placebo) as shown in Table 11. The most frequent treatment-emergent AE in the nintedanib groups was diarrhea (61.5% and 63.2%) when compared with the placebo groups (18.6% and 18.3%). Similarly, other gastrointestinal AEs (e.g., nausea, decreased appetite, vomiting, and weight loss) occurred more frequently with nintedanib and are described in detail in Table 12 and Section 3.7.4.

3.7.2 Serious adverse events

In both trials, the proportion of patients with SAEs was similar between treatment groups (31.1% versus 27.0% and 29.8% versus 32.9%) for nintedanib versus placebo in the INPULSIS-1 and INPULSIS-2 trials, respectively. There did not appear to be a higher frequency of any one type of SAE or the result of the SAE associated with either group. Across all treatment groups, the most frequent SAE in both the nintedanib and placebo groups was IPF (Table 11).

3.7.3 Withdrawals due to adverse events

More patients in the nintedanib groups (21.0% and 17.6%) withdrew due to an AE compared with patients in the placebo groups (10.8% and 15.1%) (Table 11). More patients withdrew due to gastrointestinal AEs in the nintedanib groups (8.4% and 6.4%) than in the placebo groups (1.5% and 0.9%).

3.7.4 Notable harms

Notable harms included gastrointestinal AEs, cardiac disorder and coagulation AEs, and liver enzyme and bilirubin elevations (Table 12). The most frequent treatment-emergent AEs associated with nintedanib

were gastrointestinal in nature; however, the majority of events were of mild to moderate intensity. The most frequent gastrointestinal AEs associated with nintedanib were diarrhea, nausea, and vomiting. Weight decrease and decreased appetite also appeared to occur more frequently with nintedanib; however, the proportions of patients affected were relatively less than for the other gastrointestinal AEs. Very few patients experienced dehydration.

Cardiac disorder AEs or SAEs and ischemic heart disease or serious ischemic heart disease occurred in similar proportions of patients in both the nintedanib and placebo groups. There appeared to be an imbalance in myocardial infarctions across the treatment groups, as myocardial infarction was reported in 1.6% and 1.5% of nintedanib-treated patients compared with 0.5% of placebo-treated patients in INPULSIS-1 and INPULSIS-2, respectively. In total, two myocardial infarction events in the nintedanib groups (one patient with a documented history of myocardial infarction, and both infarctions occurring after discontinuation of study drug but still within the defined follow-up period) and one myocardial infarction event in the placebo groups were fatal. There appeared to be no clinically significant differences between groups in coagulation AEs.

In both trials, nintedanib was associated with higher proportions of patients experiencing elevated liver enzymes compared with placebo-treated patients. In INPULSIS-1, 15 patients in the nintedanib group (4.9%) and one patient in the placebo group (0.5%) had alanine aminotransferase or aspartate aminotransferase levels, or both, that were three or more times the ULN. In INPULSIS-2, 17 patients in the nintedanib group (5.2%) and two patients in the placebo group (0.9%) had such elevations. Elevations in total bilirubin and alkaline phosphatase that were 1.5 or more times the ULN also occurred in more patients treated with nintedanib than patients treated with placebo.

TABLE 12: NOTABLE HARMS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Gastrointestinal AEs				
Pts With Any Diarrhea AE, n (%)	190 (100.0)	38 (100.0)	208 (100.0)	40 (100.0)
Mild	103 (54.2)	29 (76.3)	123 (59.1)	31 (77.5)
Moderate	75 (39.5)	9 (23.7)	75 (36.1)	7 (17.5)
Severe	11 (5.8)	0 (0)	10 (4.8)	2 (5.0)
Pts With Any Nausea AE, n (%)	70 (100.0)	12 (100.0)	86 (100.0)	16 (100.0)
Mild	48 (68.6)	12 (100.0)	68 (79.1)	14 (87.5)
Moderate	20 (28.6)	0 (0)	18 (20.9)	2 (12.5)
Severe	2 (2.9)	0 (0)	0 (0)	0 (0)
Pts With Any Vomiting AE, n (%)	40 (100.0)	4 (100.0)	34 (100.0)	7 (100.0)
Mild	26 (65.0)	2 (50.0)	23 (67.6)	7 (100.0)
Moderate	10 (25.0)	2 (50.0)	11 (32.4)	0 (0)
Severe	4 (10.0)	0 (0)	0 (0)	0 (0)
Pts With Any Weight Decrease AE, n (%)	25 (100.0)	13 (100.0)	37 (100.0)	2 (100.0)
Mild	16 (64.0)	7 (53.8)	23 (62.2)	1 (50.0)
Moderate	8 (32.0)	6 (46.2)	14 (37.8)	1 (50.0)
Severe	1 (4.0)	0 (0)	0 (0)	0 (0)
Pts With Any Decreased Appetite AE, n (%)	26 (100.0)	14 (100.0)	42 (100.0)	10 (100.0)
Mild	18 (69.2)	13 (92.9)	29 (69.0)	7 (70.0)
Moderate	8 (30.8)	1 (7.1)	13 (31.0)	2 (20.0)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Severe	0 (0)	0 (0)	0 (0)	1 (10.0)
Pts With Any Dehydration AE, n (%)	2 (100.0)	1 (100.0)	3 (100.0)	0 (0)
Mild	1 (50.0)	0 (0)	2 (566.7)	0 (0)
Moderate	1 (50.0)	0 (0)	0 (0)	0 (0)
Severe	0 (0)	1 (100.0)	1 (33.3)	0 (0)
Cardiac Disorder and Coagulation AEs				
Pts With Any Cardiac Disorder AE, n (%)	30 (9.7)	19 (9.3)	34 (10.3)	26 (11.9)
Pts With Any Cardiac Disorder SAE, n (%)	14 (4.5)	11 (5.4)	18 (5.5)	12 (5.5)
Pts With Fatal Cardiac Disorder AE, n (%)	1 (0.3)	2 (1.0)	2 (0.6)	4 (1.8)
Pts With Ischemic Heart Disease, n (%)	13 (4.2)	10 (4.9)	14 (4.3)	7 (3.2)
Pts With Serious Ischemic Heart Disease, n (%)	8 (2.6)	7 (3.4)	7 (2.1)	3 (1.4)
Pts With MI,[a] n (%)	5 (1.6)	1 (0.5)	5 (1.5)	1 (0.5)
Transition From Within Normal Range at Baseline to Greater Than Normal Range in > 10% of Pts, n (%)	89 (33.3)	50 (26.9)	107 (37.0)	67 (38.1)
aPTT elevation	74 (26.8)	47 (24.7)	80 (27.2)	69 (36.3)
PT-INR elevation				
Liver Enzyme and Bilirubin Elevation AEs				
Maximum Elevation, n (%)				
ALT				
≥ 3 × ULN	12 (3.9)	1 (0.5)	16 (4.9)	2 (0.9)
≥ 5 × ULN	4 (1.3)	0 (0)	6 (1.8)	0 (0)
≥ 8 × ULN	2 (0.6)	0 (0)	2 (0.6)	0 (0)
AST				
≥ 3 × ULN	10 (3.2)	0 (0)	11 (3.3)	1 (0.5)
≥ 5 × ULN	4 (1.3)	0 (0)	4 (1.2)	1 (0.5)
≥ 8 × ULN	1 (0.3)	0 (0)	3 (0.9)	1 (0.5)
ALT or AST or Both				
≥ 3 × ULN	15 (4.9)	1 (0.5)	17 (5.2)	2 (0.9)
≥ 5 × ULN	6 (1.9)	0 (0)	8 (2.4)	1 (0.5)
≥ 8 × ULN	2 (0.6)	0 (0)	3 (0.9)	1 (0.5)
Bilirubin (Total)				
≥ 1.5 × ULN	5 (1.6)	1 (0.5)	19 (3.0)	2 (0.9)
≥ 2 × ULN	1 (0.3)	0 (0)	2 (0.6)	2 (0.9)
ALKP				
≥ 1.5 × ULN	16 (5.2)	0 (0)	21 (6.4)	4 (1.8)
≥ 2 × ULN	8 (2.6)	0 (0)	9 (2.7)	1 (0.5)

AE = adverse event; ALT = alanine aminotransferase; aPTT = activated partial thromboplastin time; AST = aspartate aminotransferase; ALKP = alkaline phosphatase; MI = myocardial infarction; NTB = nintedanib; PL = placebo; PT-INR = prothrombin time – international normalized ratio; pts = patients; SAE = serious adverse event; ULN = upper limit of normal.

[a] MI is also included in ischemic heart disease (i.e., based on standard MedDRA query).

Note: Patients may be counted more than once.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

4. DISCUSSION

4.1 Summary of Available Evidence

Two replicative, placebo-controlled, double-blind, phase 3, randomized controlled trials of 52 weeks' duration met the selection criteria for inclusion in the systematic review: INPULSIS-1 (N = 515) and INPULSIS-2 (N = 551). The trials enrolled patients who were at least 40 years of age with a diagnosis of IPF within the past five years according to the most recent ATS/ERS/JRS/ALAT guidelines,[1] with baseline FVC ≥ 50% predicted and DL_{CO} of 30% to 79% predicted. All efficacy and safety end points were analyzed using the treated set, which included all randomized patients who were documented to have taken at least one dose of study drug. The primary efficacy end point was the annual decline in FVC over 52 weeks. The two key secondary end points were the change from baseline to 52 weeks in the SGRQ total score and the time to first investigator-reported acute IPF exacerbation. Key limitations are the inconsistent results observed for these key secondary end points in the included trials and the lack of data for a number of outcomes pre-specified in the review protocol as being of importance to patients (e.g., fatigue, mental health/psychological well-being, functional capacity/productivity, and caregiver burden). Various baseline patient and disease characteristics may also limit the generalizability of the results to Canadian patients with IPF, and there was no adjustment for multiplicity for the outcomes outside the pre-specified hierarchy.

4.2 Interpretation of Results

4.2.1 Efficacy

Mortality (all-cause and IPF-related) was the key efficacy outcome in the review protocol, and data for this outcome are available from various survival analyses. Neither of the INPULSIS trials demonstrated a statistically significant mortality benefit for nintedanib compared with placebo for death due to any cause, death due to a respiratory cause (adjudicated), or on-treatment death, or for the composite outcomes of time to death or lung transplant, or time to death or lung transplant or qualifying for a lung transplant. Similarly, in pre-specified pooled analyses of these survival outcomes, no statistically significant differences were observed between nintedanib and placebo. The INPULSIS trials were not powered to demonstrate a statistically significant reduction in mortality. While an effect on mortality would be the most unequivocal and clinically important measure of efficacy against a progressive and ultimately fatal disease such as IPF, it is acknowledged that it is impractical to design studies to examine mortality in IPF, given the large numbers of patients and long duration that would be required.[6,16] Although a statistically significant difference in time to death was not observed in the INPULSIS trials, a numerically smaller proportion of patients died in the nintedanib-treated groups. In the pre-specified pooled analysis, 5.5% of patients treated with nintedanib compared with 7.8% of patients treated with placebo died due to any cause over the trial duration, corresponding with a hazard ratio of 0.70.

The adjusted rate of decline in FVC over 52 weeks was the primary efficacy end point in the INPUSLSIS trials. In both trials, a statistically significant reduction in the rate of FVC decline over 52 weeks was observed with nintedanib compared with placebo. The result of the pre-specified pooled analysis was also consistent with those of the individual trials, and the magnitude of the between-group difference (~110 mL per year) is considered to be clinically meaningful according to the clinical expert consulted on this review. FVC has been considered to be a useful prognostic indicator in IPF, and based on natural history studies, a decline of ≥ 10% in an individual's FVC is considered to be a sign of disease progression.[1] For this reason, FVC has been widely used as an end point in phase 3 clinical trials in IPF; however, it is a surrogate end point and, before the trials of nintedanib and pirfenidone, had not been validated as a reliable predictor of clinically meaningful outcomes such as mortality in patients with

IPF.[16] In its reviews of pirfenidone and nintedanib, the US FDA took the position that because IPF causes a progressive decline in pulmonary function in a restrictive or scarring pattern, it is logical to monitor for change in a surrogate lung function parameter such as FVC that reflects such changes.[6] Due to the relationship between FVC and mortality trends observed in the trials for nintedanib and pirfenidone, the US FDA concurred that FVC was a clinically relevant efficacy measure in IPF.[6] As detailed in Appendix 5: VALIDITY OF OUTCOME MEASURES, a change in per cent predicted FVC of ≥ 10% is predictive of mortality in IPF.[1,7-10]

The effect of nintedanib on FVC was shown to be robust, as other secondary outcomes based on FVC demonstrated similar findings, although there was no adjustment for multiplicity for these outcomes. The absolute change from baseline in FVC and in FVC per cent predicted over 52 weeks was statistically significantly lower in the nintedanib-treated groups in both INPULSIS trials compared with the placebo-treated groups. A responder analysis based on no absolute decline in per cent predicted FVC > 5% was also consistent with other results for FVC. In both INPULSIS trials, statistically significantly more nintedanib-treated patients (52.75% and 53.19%) had an FVC response at the 5% threshold compared with placebo-treated patients (38.24% and 39.27%). When the threshold was set at 10%, the trials demonstrated inconsistent results. In INPULSIS-1, a statistically significantly greater proportion of patients in the nintedanib group (70.55%) than the placebo group (56.86%) had an FVC response at the 10% threshold; however, in INPULSIS-2, the difference between nintedanib-treated patients (69.60%) and placebo-treated patients (63.93%) was not statistically significant. The MCID for per cent predicted FVC is reported to be between 2% and 6%.[7] According to the clinical expert consulted for this review, a 5% decline in FVC may be clinically significant in clinical practice, with 10% being clearly clinically significant. It did not appear that the reduced decline in FVC correlated with improvements in HRQoL or in symptoms as measured by the patient-reported outcomes reported in the INPULSIS trials.

The effect of nintedanib on HRQoL in the INPULSIS trials was inconsistent. One of the key secondary outcomes in the trials was the change from baseline to week 52 in the total SGRQ score, an instrument validated in asthma and chronic obstructive pulmonary disease for the measurement of HRQoL and for which the MCID in chronic obstructive pulmonary disease is considered to be a change in 4 points. The MCID for the SGRQ total score in IPF is reported to be 6 to 13 points, as detailed in Appendix 5: VALIDITY OF OUTCOME MEASURES. In INPULSIS-1, there was no statistically significant between-group difference in the adjusted mean change (–0.05 points) in the total SGRQ score from baseline to week 52. In INPULSIS-2, there was a statistically significantly smaller increase in the total SGRQ score at week 52 (consistent with less deterioration in HRQoL) with nintedanib than with placebo, corresponding with an adjusted mean change of −2.69 points. In the pre-specified pooled analysis of the SGRQ total score, the adjusted mean change was –1.43 points, which was not statistically significant. None of the between-group comparisons exceeded the MCID for the SGRQ total score in IPF. Additional analyses, including SGRQ responders (defined as ≤ –4 point change from baseline in SGRQ total score at 52 weeks), change from baseline in SGRQ domain scores, and IPF-specific SGRQ total score, were aligned with the changes in the SGRQ total score in each trial. The differences in results between INPULSIS-1 and INPULSIS-2 could not be explained by differences in baseline characteristics between the trials.[14] Another measure of HRQoL used in the INPULSIS trials was the EQ-5D VAS score. An MCID has not been established in IPF for EQ-5D. In addition, various methodological issues with EQ-5D in the INPULSIS trials (e.g., missing data) put the validity of EQ-5D results into question. Nonetheless, although no statistical comparisons were conducted, the change in EQ-5D VAS score from baseline to 52 weeks was less (suggesting less deterioration of HRQoL) with nintedanib (–2.46 and –2.52) compared with placebo (–5.88 and –5.60) in both INPULSIS-1 and INPULSIS-2, respectively.

Acute IPF exacerbations are unpredictable events that result in acute deterioration in respiratory status and are associated with high morbidity and mortality.[35] The effect of nintedanib was also inconsistent for the time to first investigator-reported acute IPF exacerbation, the other key secondary outcome in the INPULSIS trials. There was no significant difference in INPULSIS-1 between the nintedanib and placebo groups in the time to the first investigator-reported acute IPF exacerbation (hazard ratio 1.15 [95% CI, 0.54 to 2.42]). In contrast, in INPULSIS-2, there was a significant increase in the time to the first investigator-reported acute exacerbation in the nintedanib group versus the placebo group (hazard ratio 0.38 [95% CI, 0.19 to 0.77]). The pre-specified pooled analysis of investigator-reported events demonstrated no significant difference between treatment groups. It may be that the heterogeneity in the findings is due to the relatively rare occurrence of exacerbations in IPF patients in clinical trials and the difficulty in assessing and categorizing the events.[14] In another pre-specified analysis of pooled data for the time to the first adjudicated acute exacerbation (confirmed or suspected events as adjudicated by the independent adjudication committee), it was shown that nintedanib was associated with a statistically significant increase in the time to first acute IPF exacerbation (hazard ratio 0.32 [95% CI, 0.16 to 0.65]). During the adjudication process, two events were deemed to have insufficient data for adjudication and 31 events were found to not be acute exacerbations. Acute exacerbation of IPF is reported to be a clinically meaningful end point because it is a direct measure of symptoms and is associated with high mortality and impaired functional status.[16] In many cases, documentation of acute IPF exacerbation events in a clinical trial requires central adjudication, although the adjudication process itself can be subject to measurement error.[16] Another possible reason for the inconsistency between the INPULSIS trials may be a treatment effect dilution.[15] Due to the apparent high number of false exacerbations identified during adjudication, inaccuracy is introduced by the random addition of non-events in the investigator-reported analysis, leading to a dilution of the drug effect and the result of no statistically significant difference.[15] In the analysis of adjudicated events, the non-events are removed and the result becomes statistically significant because the number of false exacerbations is equally distributed between groups.[15] The clinical expert consulted for this review also advised that the investigation of exacerbations may have been very centre-specific and this may, in part, also explain the discrepancy.

There were limited data from the INPULSIS trials for the outcome of lung transplantation, as the proportion of patients who received a lung transplant was very low across all treatment groups (0% to 1.3% with nintedanib and 0.5% to 0.9% with placebo) with no statistical comparisons made between groups. The proportion of patients qualifying for a lung transplant was also similar across groups. This is not surprising, as an exclusion criterion for entry into the INPULSIS trials was if a patient was likely to have a lung transplant during the trial. Although lung transplantation might be a clinically meaningful end point, it is not one that is necessarily related to the natural history of IPF due to disease-independent factors such as donor availability, age, comorbidity, social support, insurance status, and differences in allocation regimens across centres.[16]

There were no statistically significant differences observed between treatment groups in any of the patient-reported outcomes used to measure symptoms in either INPULSIS-1 or INPULSIS-2. This included the SOBQ total score, the proportion of PGI-C responders, and the CASA-Q cough score (symptoms and impact domains). Patients identified symptoms and the severity of symptoms as important outcomes in IPF, and the lack of a statistically significant treatment effect of nintedanib on these outcomes is troubling. None of the patient-reported outcome instruments used in the INPULSIS trials appear to have been validated in patients with IPF (there is limited evidence for a weak correlation between SOBQ and changes in FVC).[17] In addition, the use of the CASA-Q instrument may not have been appropriate in this patient population as, according to the clinical expert consulting on this review, sputum is not generally

a factor in IPF. There were also methodological issues in the measurement of patient-reported outcomes (e.g., missing data, incorrect MCID for SGRQ responders in IPF). Patient-reported outcomes can be clinically meaningful, but because the patient-reported outcomes used in the INPULSIS trials were not validated for use in IPF, it is unknown if these outcomes accurately measured disease symptoms (e.g., dyspnea, cough) or broader constructs such as health status in patients with IPF or if they were sensitive enough to detect treatment effects.[16] There also was no measurement of specific symptoms associated with IPF that were identified by patients as particularly important, such as fatigue, loss of energy, reduced physical activity, functional status, mental well-being, or caregiver burden (see APPENDIX 1: PATIENT INPUT SUMMARY).

Patient adherence with study medication was high across all treatment groups (mean compliance was > 96% for both the nintedanib and the placebo treatment groups in both trials). Similarly, overall exposure to study drug was similar between groups (> 65% of patients in both treatment groups were exposed to study drug for six to 12 months). Although roughly five times more nintedanib-treated patients required one or more dose reductions and twice as many nintedanib-treated patients required one or more treatment interruptions compared with placebo-treated patients, the differences in the proportions of patients who prematurely discontinued the trials was not great. Overall, 25.2% and 23.7% of nintedanib-treated patients and 17.6% and 20.1% of placebo-treated patients prematurely discontinued the INPULSIS-1 and INPULSIS-2 trials, respectively, with the main reason being an AE regardless of treatment group. This suggests that the regimen of reducing the dose of nintedanib to 100 mg twice daily or temporarily interrupting therapy was effective at managing the AE profile of nintedanib and minimizing treatment discontinuations.

Nintedanib is the second drug to be specifically approved in Canada for the treatment of IPF. The first drug, pirfenidone (Esbriet), was approved by Health Canada for the treatment of mild to moderate IPF, although there do not appear to be specific clinical criteria available for differentiating between different severities of IPF. Pirfenidone was studied in two placebo-controlled randomized controlled trials (CAPACITY-1 and CAPACITY-2) that demonstrated inconsistent results with respect to statistical significance of improvements in the rate of FVC decline and the six-minute walk test.[36] These findings, in addition to insufficient evidence of an effect on mortality or quality of life, led to a "Do Not List" recommendation from the Canadian Drug Expert Committee in 2013.[36] In 2015, the Committee considered a resubmission for pirfenidone in which results from the ASCEND randomized controlled trial were presented that showed clinically significant improvements in per cent predicted FVC compared with placebo and a reduction in all-cause mortality and IPF-related mortality based on a pre-specified meta-analysis of the CAPACITY-1, CAPACITY-2, and ASCEND trials.[25] Although the ASCEND trial and the two INPULSIS trials all demonstrated that treatment resulted in relative reductions in the rate of decline of FVC compared with placebo, the patient populations in the trials may not be directly comparable due to differences in recruitment criteria.[33] The INPULSIS trials did not exclude patients with normal values for FVC, therefore, the mean baseline FVC was higher in the INPULSIS trials (i.e., approximately 80% of predicted value [Table 23]) compared with approximately 68% of the predicted value for FVC in the ASCEND trial, thus implying that patients in the INPULSIS trials had less severe disease than those in the ASCEND trial.[33]

In the absence of a direct head-to-head randomized controlled trial comparing nintedanib and pirfenidone, it was anticipated that the indirect comparisons identified in the literature by the CDR review team[37] and submitted by the manufacturer[26] would yield useful information on the relative efficacy and safety of the two drugs. As summarized and critically appraised in Appendix 6: SUMMARY OF INDIRECT COMPARISONS, due to the differences in study populations between the nintedanib and

pirfenidone trials (particularly in baseline FVC and potential IPF severity), the underlying assumption that the placebo groups in the trials are similar is probably invalid. In addition, the sparse networks and variability in study durations, and therefore outcome assessments, were other key limitations of the analyses. Therefore, given these limitations and the corresponding high degree of uncertainty in the comparative results, no meaningful conclusions can be drawn from the indirect comparison analyses regarding the relative efficacy and safety of nintedanib and pirfenidone.

4.2.2 Harms

Most patients in both treatment groups in the INPULSIS trials experienced AEs. The most frequent AEs were gastrointestinal in nature, with the majority of patients who received nintedanib experiencing diarrhea of mild to moderate intensity. In keeping with this, more nintedanib-treated patients than placebo-treated patients also received concomitant antidiarrheal therapy. Other gastrointestinal AEs (e.g., nausea, decreased appetite, vomiting, and weight loss) also occurred more frequently with nintedanib. In the nintedanib product monograph, it is recommended that nintedanib be taken with food to reduce the incidence of gastrointestinal effects and it is warned that diarrhea is the most frequent gastrointestinal AE and should be treated with adequate hydration and antidiarrheal medication (e.g., loperamide) or, if required, treatment interruption.[5]

The proportions of patients with SAEs were similar between treatment groups, and the most frequent SAE in both groups was IPF. Cardiac disorder AEs or SAEs or ischemic heart disease or serious ischemic heart disease occurred in similar proportions of patients in both the nintedanib and placebo groups. Despite this, there appeared to be an imbalance in myocardial infarctions, as almost three times as many nintedanib-treated patients compared with placebo-treated patients had myocardial infarctions. Of these, two myocardial infarction events in the nintedanib groups and one myocardial infarction event in the placebo groups were fatal. The clinical significance of this finding is unknown, and further observation in larger cohorts is needed.[14] There did not appear to be any clinically significant differences between groups in coagulation AEs. In both trials, nintedanib was also associated with higher proportions of patients with elevations in alanine aminotransferase or aspartate aminotransferase, or both, that were three or more times the ULN, and total bilirubin and alkaline phosphatase liver enzymes that were 1.5 or more times the ULN compared with placebo. The proportion of withdrawals due to adverse events was higher in the nintedanib treatment group, with the most common reason for withdrawal being gastrointestinal disorders.

5. CONCLUSIONS

Two replicative, placebo-controlled, double-blind, phase 3, randomized controlled trials met the selection criteria for inclusion in the systematic review (INPULSIS-1 and INPULSIS-2). Data from these trials provide evidence that, compared with placebo, nintedanib reduced the rate of decline in FVC in patients with IPF over the 52-week duration of the trials. The magnitude of the reduction was sufficiently large to be clinically meaningful. There were no statistically significant differences between nintedanib and placebo in mortality (all-cause or due to adjudicated respiratory causes) in the individual trials or in the pre-specified pooled analyses of survival data. The effect of nintedanib on HRQoL (measured as the change from baseline to week 52 in the SGRQ total score) and on the time to the first investigator-reported acute IPF exacerbation was discordant between the trials. Statistically significant differences in favour of nintedanib were observed for both outcomes in INPULSIS-2 but not in INPULSIS-1. Pre-specified pooled analyses also did not demonstrate statistically significant differences between groups for either of these outcomes. There were no statistically significant between-group differences in the patient-reported outcomes that measured patient's symptoms in either trial. The most frequent AEs associated with nintedanib were gastrointestinal in nature, with most nintedanib-treated patients experiencing diarrhea of mild to moderate intensity. Dose reduction or temporarily interrupting therapy appeared to be effective at managing the AE profile of nintedanib.

APPENDIX 1: PATIENT INPUT SUMMARY

This section was summarized by CADTH Common Drug Review (CDR) staff based on the input provided by patient groups.

1. Brief Description of Patient Groups Supplying Input

Four patients groups submitted patient input.

The British Columbia Lung Association is a charitable organization aimed at improving respiratory health as well as patient and caregiver quality of life through programs, education, research, training, treatment, advocacy, and the prevention of lung disease. The Association has received health educator's grants from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Grifols, InterMune, Merck Frosst, Novartis, and Pfizer.

The Canadian Pulmonary Fibrosis Foundation is a charitable organization with a mandate to support patients diagnosed with idiopathic pulmonary fibrosis (IPF) and their caregivers, to raise public awareness about IPF, to be the "patient voice" for IPF, and to raise funds for IPF research. The following manufacturers and organizations have sponsored the Foundation: Boehringer Ingelheim, Roche, Dynamic Funds, Emerson Canada, and Investors Group.

The Lung Association of Saskatchewan is a not-for-profit organization that aims to support and improve the lives of patients living with lung disease (including IPF) and their caregivers though programs, education, research, training, treatment, advocacy, and the prevention of lung disease. The Association has received unrestricted grants for educational initiatives from AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Grifols, InterMune, Merck Frosst, Novartis, Nycomed, Pfizer, Roche, and Takeda.

The Ontario Lung Association is a charitable organization whose mandate is to advocate for and provide programs and services to patients and health care providers while being the voice and primary resource in the prevention and control of respiratory illness. The Association has received sponsorships and grants to support educational and research initiatives from the following pharmaceutical companies: Actelion, Astellas Pharma, AstraZeneca, Bayer, Boehringer Ingelheim, Eli Lilly, GlaxoSmithKline, Grifols, InterMune, Johnson & Johnson, Merck, Novartis, Pfizer, Roche, Rx&D, Takeda, and Valeant Pharmaceuticals. In addition, they have also received program funding from the Ontario Home Respiratory Services Association.

None of the aforementioned groups declared any conflicts of interests with regard to the preparation of this submission.

2. Condition-Related Information

The information for all of the remaining sections was gathered though consultations, online surveys (patients with IPF and their caregivers), and the personal experience of a respondent.

IPF is a progressive and life-threatening condition that has no cure. All patients diagnosed with IPF experience many problems and symptoms as a result of the illness. Depending on the stage of the disease, patients will have different symptoms and increasing severity of symptoms. The most common symptoms include breathlessness, fatigue, loss of energy, reduced physical activity, and a chronic cough. Additional symptoms can also include chest pain or tightness, rapid weight loss, leg swelling, wheezing,

and difficulty fighting infections. In addition to the symptoms, patients have to manage psychosocial issues associated with their prognosis. The stress of the diagnosis and prognosis greatly affects patients' quality of life and mental well-being, even among those with mild disease and minimal symptoms. The symptoms of IPF may limit patients' ability to stay physically active, to work, to participate in social activities, to travel, and to continue their usually daily activities and household chores. For some, leaving the house takes a lot of planning and energy; therefore, taking trips is often avoided, leading to social isolation.

Complete families, along with the primary caregivers, are also affected by IPF. Caregivers worry about their loved one's prognosis in addition to feeling helpless, depressed, anxious, stressed, exhausted, and isolated. They may limit their own social activities and employment in order to take on more of the household chores and to provide care and attend medical appointments for their loved one. In addition, the worries and increase in care are compounded as the patients' disease worsens. There is also constant worry associated with coverage of medications, particularly if health plans cease to pay for medications.

3. Current Therapy-Related Information

There are few supportive treatments available to manage symptoms of IPF. Patients have reported using oxygen (sometimes at night only, while others with moderate to advanced disease often require daily use, particularly when performing physical activity), N-acetylcysteine, prednisone, azathioprine, and pulmonary rehabilitation. These treatments were somewhat effective in some patients but were also accompanied with unwanted side effects (such as weight gain, mood swings, confusion, difficulty sleeping, and bowel issues), which added to their stress. Lung transplant is an option that is available to few patients due to the requirement that the patient meet specific criteria to be a suitable candidate for this procedure.

4. Expectations About Ofev

While patients realize that Ofev is not a cure, there is hope that this new medication may offer them an alternate medication choice and potentially increased tolerability, if current treatment is not well tolerated. Patients are also hopeful that Ofev will slow the progression of IPF and, at least partially, address their most problematic symptoms. Patients are also willing to deal with some side effects as long as they are not worse than what they are currently experiencing and are not irreversible. In addition, the twice daily dosing schedule is attractive to patients.

Patients who had experience with Ofev indicated various benefits, including reduced cough, fatigue, and shortness of breath, and an improvement in appetite and energy level. Some patients also indicated that they experienced fewer or no side effects with Ofev and that, when compared with other drug treatments, Ofev was superior in administration, cost burden, side effect profile, and treating PF or IPF. In addition, most patients on Ofev reported an improvement in their quality of life. Although most Ofev experiences were reported as being positive, some patients indicated that they felt worse on it or that it was too early to tell if it was making a difference.

APPENDIX 2: LITERATURE SEARCH STRATEGY

OVERVIEW	
Interface:	Ovid
Databases:	Embase 1974 to May 14, 2015 MEDLINE Daily and MEDLINE 1946 to present MEDLINE In-Process & Other Non-Indexed Citations **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	May 15, 2015
Alerts:	Biweekly search updates until September 16, 2015
Study Types:	No search filters were applied
Limits:	No date or language limits were used Conference abstracts were excluded

SYNTAX GUIDE	
/	At the end of a phrase, searches the phrase as a subject heading
.sh	At the end of a phrase, searches the phrase as a subject heading
MeSH	Medical Subject Heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
.pt	Publication type
.kw	Keywords defined by the author
.ti	Title
.ab	Abstract
.ot	Original title
.hw	Heading word; usually includes subject headings and controlled vocabulary
.rn	CAS registry number
.nm	Name of substance word
pmez	Ovid database code; MEDLINE In-Process & Other Non-Indexed Citations, MEDLINE Daily and Ovid MEDLINE 1946 to Present
oemezd	Ovid database code; Embase 1974 to present, updated daily

MULTI-DATABASE STRATEGY		
Line #	**Search Strategy**	**Results**
1	(Ofev* or nintedanib* or intedanib* or vargatef* or BIBF-1120 or BIBF1120 or G6HRD2P839 or 42F62RTZ4G or 656247-17-5 or 928326-83-4).ti,ot,ab,kw,sh,rn,hw,nm,rn.	953
2	1 use pmez	165
3	*nintedanib/	137
4	(Ofev* or nintedanib* or intedanib* or vargatef* or BIBF-1120 or BIBF1120 or G6HRD2P839 or 42F62RTZ4G or 656247-17-5 or 928326-83-4).ti,ab.	463
5	or/3-4	475
6	5 use oemezd	324
7	6 not conference abstract.pt.	181
8	2 or 7	346
9	remove duplicates from 8	217

OTHER DATABASES	
PubMed	Same MeSH, keywords, limits, and study types used as per MEDLINE search, with appropriate syntax used.
Trial registries (Clinicaltrials.gov and others)	Same keywords, limits used as per MEDLINE search.

Grey Literature

Date of Search:	May 14, 2015
Keywords:	Ofev, nintedanib, BIBF-1120, BIBF1120, idiopathic pulmonary fibrosis
Limits:	No date or language limits used

Relevant websites from the following sections of the CADTH grey literature checklist, *Grey matters: a practical tool for evidence-based searching* (https://www.cadth.ca/resources/finding-evidence/grey-matters-practical-search-tool-evidence-based-medicine) were searched:

- Health Technology Assessment Agencies
- Health Economics
- Clinical Practice Guidelines
- Drug Regulatory Approvals
- Advisories and Warnings
- Drug Class Reviews
- Databases (free)
- Internet Search

APPENDIX 3: EXCLUDED STUDIES

There were no excluded studies.

APPENDIX 4: DETAILED OUTCOME DATA

TABLE 13: SURVIVAL ANALYSIS OVER 52 WEEKS: INDIVIDUAL TRIALS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Time to Death Over 52 Weeks[a]				
Patients with event, n (%)	13 (4.2)	13 (6.4)	22 (6.7)	20 (9.1)
Probability of being event-free	0.958	0.935	0.933	0.908
Comparison vs. PL[b]				
Hazard ratio[c]	0.63		0.74	
95% CI	0.29 to 1.36		0.40 to 1.35	
P value[d]	0.2880		0.2995	
Time to Death Due to Respiratory Cause Over 52 Weeks[e]				
Patients with event, n (%)	10 (3.2)	10 (4.9)	14 (4.3)	11 (5.0)
Probability of being event-free	0.967	0.950	0.956	0.949
Comparison vs. PL[b]				
Hazard ratio[c]	0.61		0.86	
95% CI	0.25 to 1.47		0.39 to 1.90	
P value[d]	0.3515		0.6654	
Time to On-Treatment Death[f]				
Patients with event, n (%)	8 (2.6)	9 (4.4)	16 (4.9)	17 (7.8)
Probability of being event-free	NE	NE	NE	NE
Comparison vs. PL				
Hazard ratio[c]	0.68		0.68	
95% CI	0.26 to 1.82		0.34 to 1.35	
P value[d]	0.4869		0.1664	
Time to Death or Lung Transplant Over 52 Weeks				
Patients with event, n (%)	16 (5.2)	14 (6.9)	22 (6.7)	22 (10.0)
Probability of being event-free	0.947	0.930	0.933	0.899
Comparison vs. PL[b]				
Hazard ratio[c]	0.73		0.67	
95% CI	0.36 to 1.51		0.37 to 1.21	
P value[d]	0.4430		0.1664	
Time to Death or Lung Transplant or				

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Qualifying for Lung Transplant[g] Over 52 Weeks				
Patients with event, n (%)	46 (14.9)	37 (18.1)	64 (19.5)	52 (23.7)
Probability of being event-free	0.848	0.817	0.801	0.761
Comparison vs. PL[b]				
Hazard ratio[c]	0.81		0.80	
95% CI	0.52 to 1.25		0.55 to 1.16	
P value[d]	0.3558		0.2123	

CI = confidence interval; DL_{CO} = diffusion capacity of lung for carbon monoxide; FVC = forced vital analysis; NE = not estimated; NTB = nintedanib; PL = placebo; SpO_2 = oxygen saturation of peripheral blood; vs. = versus.
[a] In INPULSIS-1, there were 33 deaths reported during the trial. Of these, 26 deaths occurred during the 52-week treatment period (372 days after randomization) and are included in the time-to-death analysis. The remaining 7 patients died after 52 weeks, including one patient who died after receiving a lung transplant. In INPULSIS-2, there were 54 deaths reported during the trial. Of these, 42 occurred within the 52-week treatment period (372 days after randomization) and are included in the time-to-death analysis. The remaining 12 patients died after 52 weeks.
[b] Based on data collected up to 372 days after randomization (52 weeks + 7 days margin).
[c] Based on Cox's regression model with terms for treatment, gender, age, and height.
[d] Based on log-rank test.
[e] According to adjudication.
[f] On-treatment deaths include deaths having occurred within 28 days following the last trial drug intake.
[g] Qualifying for lung transplant was defined as FVC < 45% predicted, or DL_{CO} < 30% predicted, or SpO_2 < 88% at rest, at sea level.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report,[19] Common Technical Document.[26]

TABLE 14: SURVIVAL ANALYSIS OVER 52 WEEKS: MANUFACTURER-POOLED ANALYSIS (TREATED SET)

	INPULSIS-1	
	NTB 150 mg N = 638[a]	PL N = 423[a]
Time to Death Over 52 Weeks		
Patients with event, n (%)	35 (5.5)	33 (7.8)
Comparison vs. PL[b]		
Hazard ratio[c]	0.70	
95% CI	0.43 to 1.12	
P value[d]	0.1399	
Time to Death Due to Respiratory Cause Over 52 Weeks[e]		
Patients with event, n (%)	24 (3.8)	21 (5.0)
Comparison vs. PL[b]		
Hazard ratio[c]	0.74	
95% CI	0.41 to 1.34	
P value[d]	0.3435	
Time to On-Treatment Death[f]		
Patients with event, n (%)	24 (3.8)	26 (6.1)
Comparison vs. PL		
Hazard ratio[c]	0.68	

	INPULSIS-1	
	NTB 150 mg N = 638[a]	**PL N = 423[a]**
95% CI	0.39 to 1.19	
P value[d]	0.1599	
Time to Death or Lung Transplant Over 52 Weeks		
Patients with event, n (%)	38 (6.0)	36 (8.5)
Comparison vs. PL[b]		
Hazard ratio[c]	0.70	
95% CI	0.44 to 1.10	
P value[d]	0.1185	
Time to Death or Lung Transplant or Qualifying for Lung Transplant[g] Over 52 Weeks		
Patients with event, n (%)	110 (17.2)	89 (21.0)
Comparison vs. PL[b]		
Hazard ratio[c]	0.80	
95% CI	0.60 to 1.06	
P value[d]	0.1233	

CI = confidence interval; DL_{CO} = diffusion capacity of lung for carbon monoxide; FVC = forced vital analysis; NTB = nintedanib; PL = placebo; SpO_2 = oxygen saturation of peripheral blood; vs. = versus.
[a] Cumulative number of patients entering the respective time interval.
[b] Based on data collected up to 372 days (52 weeks + 7 days margin), including data collected after premature discontinuation of study medication.
[c] Based on log-rank test.
[d] Based on Cox's regression model with terms for trial, treatment, gender, age, and height.
[e] According to adjudication.
[f] On-treatment deaths are those that occurred on-treatment or during the follow-up period of 28 days.
[g] Qualifying for lung transplant was defined as FVC < 45% predicted, or DL_{CO} < 30% predicted, or SpO_2 < 88% at rest, at sea level.
Source: Common Technical Document.[26]

TABLE 15: ABSOLUTE CHANGE FROM BASELINE IN ST. GEORGE'S RESPIRATORY QUESTIONNAIRE TOTAL SCORE AT WEEK 52 (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	298	202	326	217
Mean (SD)	39.55 (17.628)	39.79 (18.478)	39.46 (20.471)	39.39 (18.647)
Week 52				
N	240	163	267	179
Mean (SD)	42.67 (20.187)	42.36 (21.391)	40.33 (21.977)	42.24 (21.113)
Change From Baseline to Week 52				
N	233	161	264	178
Mean (SD)	3.91 (15.995)	3.73 (15.348)	2.16 (15.195)	5.34 (16.101)
Adjusted[a,b] mean (SE)	4.34 (0.799)	4.39 (0.960)	2.80 (0.730)	5.48 (0.891)
95% CI	2.77 to 5.90	2.51 to 6.27	1.36 to 4.23	3.74 to 7.23
Comparison vs. PL				
Adjusted[a,b] mean difference (SE)	–0.05 (1.248)		–2.69 (1.151)	
95% CI	–2.50 to 2.40		–4.95 to –0.43	
P value	0.9657		0.0197	

Manufacturer-Pooled Analysis From INPULSIS-1 and INPULSIS-2		
	NTB 150 mg N = 638	PL N = 423
Change From Baseline to Week 52		
N	497	339
Mean (SD)	2.98 (15.584)	4.58 (15.746)
Adjusted[a] mean (SE)	3.53 (0.540)	4.96 (0.654)
95% CI	2.47 to 4.59	3.68 to 6.24
Comparison vs. PL		
Adjusted[a,b] mean difference (SE)	–1.43 (0.848)	
95% CI	–3.09 to 0.23	
P value	0.0923	
Treatment by time by trial interaction *P* value[b]	0.2081	

CI = confidence interval; IPF = idiopathic pulmonary fibrosis; MMRM = mixed model for repeated measures; NTB = nintedanib; PL = placebo; SD = standard deviation; SE = standard error; SGRQ = St. George's Respiratory Questionnaire; vs. = versus.
[a] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline SGRQ total score, and baseline SGRQ total score by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix.
[b] Adjusted mean was based on all analyzed patients in the model (not only patients with a change from baseline to week 52).
Note: Two alternative hierarchies with different orders of the key secondary end points were required for US and European Union submissions. For US registration, the first key secondary end point was the time to first investigator-reported acute IPF exacerbation and the second was the change from baseline in SGRQ total score, whereas for European registration, the opposite order was used. It is not known what hierarchical order Health Canada considered because the Health Canada reviewer's report was not available. In INPULSIS-2, since both key secondary end points were statistically significant, formal confirmatory testing could proceed for both key secondary end points. In INPULSIS-1, neither the result for time to first acute exacerbation nor the change from baseline in SGRQ total score at week 52 was statistically significant; hence the statistical testing was made only on a nominal basis for the second key secondary end point.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 16: INPULSIS-1 SUMMARY OF PATIENT-REPORTED OUTCOME ASSESSMENTS UP TO WEEK 52 (TREATED SET)

Patient-Reported Outcome	NTB 150 mg N = 309	PL N = 204	Comparison vs. PL, Adjusted mean (SE)	95% CI	*P* Value
	Adjusted Mean (SE)	Adjusted Mean (SE)			
SGRQ Responders,[a] n (%)	63 (20.39)	49 (24.02)	0.840[b]	0.55 to 1.29	0.4298[c]
SGRQ Scores					
Total[d,e]	4.34 (0.799)	4.39 (0.960)	–0.05 (1.248)	–2.50 to 2.40	0.9657
Symptoms[e]	1.56 (1.104)	3.89 (1.351)	–2.32 (1.744)	–5.74 to 1.10	0.1832
Activity[e]	4.62 (0.906)	5.81 (1.103)	–1.19 (1.427)	–3.99 to 1.61	0.4049
Impact[e]	4.87 (0.923)	4.01 (1.113)	0.86 (1.446)	–1.97 to 3.70	0.5510
SGRQ-I Total Score[f]	4.30 (0.824)	5.08 (0.992)	–0.78 (1.289)	–3.31 to 1.75	0.5446
SOBQ Total Score[g]	6.73 (1.113)	7.61 (1.376)	–0.88 (1.770)	–4.35 to 2.60	0.6203
PGI-C Responders,[h] n (%)	188 (60.84)	112 (54.90)	1.276[c]	0.89 to 1.83	0.1818[i]
CASA-Q Cough Score					
Symptoms[e,j]	–0.76 (1.136)	–0.52 (1.400)	–0.24 (1.803)	–3.78 to 3.30	0.8942
Impact[e,j]	–2.36 (1.006)	–4.00 (1.240)	1.64 (1.596)	–1.49 to 4.77	0.3042

CASA-Q = Cough and Sputum Assessment Questionnaire; CI = confidence interval; MMRM = mixed effects model for repeated measures; NTB = nintedanib; PGI-C = Patient Global Impression of Change; PL = placebo; SE = standard error; SGRQ = St. George's Respiratory Questionnaire; SGRQ-I = St. George's Respiratory Questionnaire specific to idiopathic pulmonary fibrosis; SOBQ = Shortness of Breath Questionnaire.

[a] Responders defined as ≤ –4 points in change from baseline in SGRQ total score at 52 weeks.

[b] Odds ratio.

[c] Based on logistic regression

[d] Adjusted mean was based on all analyzed patients in the model (not only patients with a change from baseline to week 52).

[e] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline SGRQ total score/symptoms/activity/impact component, and baseline SGRQ total score/symptoms/activity/impact component by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix. Note: Lower score indicates better health.

[f] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline SGRQ-I total score, and baseline SGRQ-I total score by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix.

[g] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline SOBQ total score, and baseline SOBQ total score by visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix. Note: A lower score indicates less shortness of breath.

[h] Responders at 52 weeks defined as very much better, much better, a little better, or no change.

[i] Based on logistic regression with term treatment.

[j] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline CASA-Q cough symptoms/impact score, and baseline CASA-Q cough symptoms/impact score by visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix. A higher score is associated with fewer symptoms due to cough or less impact due to cough. Note: The actual number of patients analyzed can differ for each of the end points.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 17: INPULSIS-2 SUMMARY OF PATIENT-REPORTED OUTCOME ASSESSMENTS UP TO WEEK 52 (TREATED SET)

Patient-Reported Outcome	NTB 150 mg N = 309	PL N = 204	Comparison vs. PL, Adjusted mean (SE)	95% CI	*P* Value
	Adjusted Mean (SE)	Adjusted Mean (SE)			
SGRQ Responders,[a] n (%)	83 (25.23)	37 (16.89)	1.664[b]	1.08 to 2.57	0.0218[c]
SGRQ Scores					
Total[d,e]	2.80 (0.730)	5.48 (0.891)	–2.69 (1.151)	–4.95 to –0.43	0.0197
Symptoms[e]	2.03 (1.061)	3.43 (1.297)	–1.40 (1.675)		0.4019
Activity[e]	3.89 (0.863)	7.20 (1.052)	–3.31 (1.360)	–4.69 to 1.88	0.0152
Impact[e]	2.85 (0.852)	5.93 (1.036)	–3.08 (1.342)	–5.97 to –0.64	0.0220
SGRQ-I Total Score[f]	2.72 (0.757)	5.84 (0.921)	–3.12 (1.192)	–5.71 to –0.45	0.0089
SOBQ Total Score[g]	6.69 (1.073)	9.07 (1.300)	–2.38 (1.685)	–5.46 to –0.79	0.1587
PGI-C Responders,[h] n (%)	203 (61.70)	118 (53.88)	1.379[b]	–5.68 to 0.93	0.0690[i]
CASA-Q Cough Score				0.98 to 1.95	
Symptoms[e,j]	–0.33 (1.087)	–2.38 (1.325)	2.05 (1.713)	–1.31 to 5.41	0.2326
Impact[e,j]	–2.58 (0.991)	–4.39 (1.209)	1.81 (1.564)	–1.26 to 4.88	0.2475

CASA-Q = Cough and Sputum Assessment Questionnaire; CI = confidence interval; MMRM = mixed effects model for repeated measures; NTB = nintedanib; PGI-C = Patient Global Impression of Change; PL = placebo; SE = standard error; SGRQ = St. George's Respiratory Questionnaire; SGRQ-I = SGRQ specific to idiopathic pulmonary fibrosis; SOBQ = Shortness of Breath Questionnaire.

[a] Responders defined as ≤ –4 points in change from baseline in SGRQ total score at 52 weeks.

[b] Odds ratio.

[c] Based on logistic regression

[d] Adjusted mean was based on all analyzed patients in the model (not only patients with a change from baseline to week 52).

[e] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline SGRQ total score/symptoms/activity/impact component, and baseline SGRQ total score/symptoms/activity/impact component by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix. Note: Lower score indicates better health.

[f] Based on MMRM, with fixed effects treatment, visit treatment by visit, baseline SGRQ-I total score, and baseline SGRQ-I total score by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix.

[g] Based on MMRM, with fixed effects treatment, visit treatment by visit, baseline SOBQ total score, and baseline SOBQ total score-by-visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix. Note: A lower score indicates less shortness of breath.

[h] Responders at 52 weeks defined as very much better, much better, a little better, or no change.

[i] Based on logistic regression with term treatment.

[j] Based on MMRM, with fixed effects for treatment, visit treatment by visit, baseline CASA-Q cough symptoms/impact score, and baseline CASA-Q cough symptoms/impact score by visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix. A higher score is associated with fewer symptoms due to cough or less impact due to cough.

Note: The actual number of patients analyzed can differ for each of the end points.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 18: TIME TO FIRST EXACERBATION (DAYS) OVER 52 WEEKS; INVESTIGATOR-REPORTED EVENTS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Patients With Events, n (%)	19 (6.1)	11 (5.4)	12 (3.6)	21 (9.6)
Probability of Being Event-Free	0.934	0.944	0.961	0.898
Comparison vs. PL[a]				
Hazard ratio[b]	1.15		0.38	
95% CI	0.54 to 2.42		0.19 to 0.77	
P value[c]	0.6728		0.0050	
Manufacturer-Pooled Analysis From INPULSIS-1 and INPULSIS-2				
	NTB 150 mg N = 638		PL N = 423	
Patients With Events, n (%)	31 (4.9)		32 (7.6)	
Comparison vs. PL[a]				
Hazard ratio[b,d]	0.64			
95% CI; *P* value[c]	0.39 to 1.05; 0.0823			
Treatment by trial interaction *P* value[e]	0.0320			

CI = confidence interval; NTB = nintedanib; PL = placebo; vs. = versus.
[a] Based on data collected up to 372 days after randomization (52 weeks + 7 days margin).
[b] Based on a Cox's regression model with terms for treatment, gender, age, and height.
[c] Based on a log-rank test.
[d] Trial is included as a fixed effect to the model.
[e] Based on a Cox's regression model with terms for treatment, gender, age, height, trial, and treatment by trial.
Note: Two alternative hierarchies with different orders of the key secondary end points were required for US and European Union submissions. For US registration, the first key secondary end point was the time to first investigator-reported acute IPF exacerbation and the second was the change from baseline in SGRQ total score, whereas for European registration, the opposite order was used. It is not known what hierarchical order Health Canada considered because the Health Canada reviewer's report was not available. In INPULSIS-2, since both key secondary end points were statistically significant, formal confirmatory testing could proceed for both key secondary end points. In INPULSIS-1, neither the result for time to first acute exacerbation nor the change from baseline in SGRQ total score at week 52 was statistically significant; hence the statistical testing was made only on a nominal basis for the second key secondary end point.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 19: TIME TO FIRST EXACERBATION (DAYS) OVER 52 WEEKS; ADJUDICATED EVENTS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Patients With ≥ 1 Acute Exacerbation (Investigator-Reported), n (%)	19 (6.1)	11 (5.4)	12 (3.6)	21 (9.6)
Number (%) of Adjudicated Acute Exacerbation Events				
Confirmed	2 (0.6)	1 (0.5)	2 (0.6)	3 (1.4)
Suspected	5 (1.6)	7 (3.4)	3 (0.9)	13 (5.9)
No acute exacerbation	12 (3.9)	5 (2.5)	7 (2.1)	7 (3.2)
Insufficient data for adjudication	1 (0.3)	0 (0)	0 (0)	1 (0.5)
Manufacturer-Pooled Analysis From INPULSIS-1 and INPULSIS-2				
	NTB 150 mg N = 638		PL N = 423	
Patients With ≥ 1 Acute Exacerbation (Investigator-Reported), n (%)	31 (4.9)		32 (7.6)	
Number (%) of Adjudicated Acute Exacerbation Events				
Confirmed	4 (0.6)		4 (0.9)	
Suspected	8 (1.3)		20 (4.7)	
No acute exacerbation	19 (3.0)		12 (2.8)	
Insufficient data for adjudication	1 (0.2)		1 (0.2)	
Comparison vs. PL				
Hazard ratio	0.32			
95% CI; *P* value	0.16 to 0.65; 0.001			

CI = confidence interval; NTB = nintedanib; PL = placebo; vs. = versus.
Source: Richeldi et al. 2014 Supplementary Appendix.[38]

TABLE 20: INCIDENCE OF PATIENTS WITH AT LEAST ONE ACUTE IDIOPATHIC PULMONARY FIBROSIS EXACERBATION OVER 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Number of Patients With ≥ 1 Acute IPF Exacerbation,[a] n (%)	19 (6.1)	11 (5.4)	12 (3.6)	21 (9.6)
Total Number of Years at Risk	289.9	196.3	311.1	205.0
Incidence of Exacerbation Rates[b] Per 100 Patient-Years	6.6	5.6	3.9	10.2
Comparison vs. PL				
Risk ratio	1.17		0.38	
95% CI	0.56 to 2.46		0.19 to 0.77	
P value[c]	0.6793		0.0070	

CI = confidence interval; IPF = idiopathic pulmonary fibrosis; NTB = nintedanib; PL = placebo; vs. = versus.
[a] Based on investigator-reported adverse events.
[b] Calculated as the number of patients with at least one acute IPF exacerbation divided by the total number of years at risk.
[c] Based on the statistic of ln (risk ratio) having a normal distribution.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 21: RATE OF DECLINE IN FORCED VITAL CAPACITY (ML/YR) OVER 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Rate of Decline Over 52 Weeks				
Adjusted rate[a] (SE)	−114.65 (15.327)	−239.91 (18.709)	−113.59 (15.726)	−207.32 (19.309)
95% CI	−144.78 to −84.53	−276.68 to −203.14	−144.50 to −82.69	−245.27 to −169.38
Comparison vs. PL				
Adjusted rate difference[a] (SE)	125.26 (24.209)		93.73 (24.907)	
95% CI	(77.68 to 172.84)		(44.78 to 142.68)	
P value	< 0.0001		0.0002	
Manufacturer-Pooled Analysis From INPULSIS-1 and INPULSIS-2				
	NTB 150 mg N = 638		PL N = 423	
Rate of Decline Over 52 Weeks				
Adjusted rate[a] (SE)	−113.59 (10.984)		−223.53 (13.448)	
95% CI	−135.14 to −92.03		−249.92 to −197.13	
Comparison vs. PL				
Adjusted rate difference[a] (SE)	109.94 (17.368)			
95% CI	75.85 to 144.03			
P value	< 0.0001			
Treatment by time by trial interaction *P* value[b]	0.4482			

CI = confidence interval; NTB = nintedanib; PL = placebo; SE = standard error; vs. = versus; yr = year.

[a] Based on a random coefficient regression with fixed effects for treatment, gender, age, and height; and random effect of patient-specific intercept and time. Within-patient errors were modelled by an unstructured variance-covariance matrix. Inter-individual variability was modelled by a variance-components variance-covariance matrix.

[b] Based on a random coefficient regression with fixed effects for treatment, gender, age, height, trial, and treatment by time by trial interaction; and random effect of patient-specific intercept and time.

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 22: ABSOLUTE CHANGE FROM BASELINE IN FORCED VITAL CAPACITY (ML) OVER 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	309	204	329	219
Mean (SD)	2,756.8 (735.12)	2,844.5 (820.11)	2,672.8 (775.96)	2,619.0 (787.35)
Week 52				
N	250	165	269	180
Mean (SD)	2,669.0 (772.04)	2,664.4 (834.01)	2,637.3 (811.80)	2,512.5 (821.44)
Change from Baseline to Week 52				
N	250	165	269	180
Mean (SD)	–90.93 (242.704)	–201.81 (305.869)	–86.94 (283.398)	–204.04 (280.543)
Adjusted[a,b] mean (SE)	–95.07 (14.375)	–205.00 (16.544)	–95.26 (14.200)	–205.03 (16.629)
95% CI	–123.27 to –66.86	–237.45 to –172.54	–123.12 to –67.40	–237.66 to –172.41
Comparison vs. PL				
Adjusted[a,b] mean (SE)	109.93 (19.708)		109.77 (19.808)	
95% CI	71.27 to 148.59		70.92 to 148.62	
P value	< 0.0001		< 0.0001	

CI = confidence interval; FVC = forced vital capacity; MMRM = mixed model for repeated measures; NTB = nintedanib; PL = placebo; SD = standard deviation; SE = standard error; vs. = versus.

[a] Based on MMRM, with fixed effects for treatment, visit, gender, age, height, treatment by visit, baseline FVC, and baseline FVC by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix.

[b] Adjusted mean was based on all analyzed patients in the model (not only patients with a change from baseline to week 52).

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 23: ABSOLUTE CHANGE FROM BASELINE IN FORCED VITAL CAPACITY (PER CENT PREDICTED) OVER 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	309	204	329	219
Mean (SD)	79.47 (17.028)	80.53 (17.341)	79.99 (18.081)	78.09 (18.968)
Week 52				
N	250	165	269	180
Mean (SD)	76.55 (17.506)	75.70 (18.080)	77.86 (18.997)	73.60 (20.530)
Change from Baseline to Week 52				
N	250	165	269	180
Mean (SD)	–2.65 (7.057)	–5.90 (8.746)	–2.78 (8.677)	–6.07 (8.026)
Adjusted[a,b] mean (SE)	–2.76 (0.408)	–5.98 (0.474)	–3.09 (0.433)	–6.15 (0.505)
95% CI	–3.56 to –1.95	–6.91 to –5.05	–3.94 to –2.24	–7.14 to –5.16
Comparison vs. PL				
Adjusted[a,b] mean (SE)	3.22 (0.564)		3.06 (0.607)	
95% CI	2.11 to 4.33		1.87 to 4.25	
P value	< 0.0001		< 0.0001	

CI = confidence interval; FVC = forced vital capacity; MMRM = mixed model for repeated measures; NTB = nintedanib; PL = placebo; SD = standard deviation; SE = standard error; vs. = versus.

[a] Based on MMRM with fixed effects for treatment, visit, gender, age, height, treatment by visit, baseline FVC (per cent predicted), and baseline FVC (per cent predicted) by visit; and random effect for patient. Within-patient errors were modelled by compound symmetry covariance matrix.

[b] Adjusted mean was based on all analyzed patients in the model (not only patients with a change from baseline to week 52).

Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 24: PROPORTION OF FORCED VITAL CAPACITY RESPONDERS USING 5% AND 10% THRESHOLDS AT 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Responders at 5% Threshold				
FVC Responders[a]				
n (%)	163 (52.75)	78 (38.24)	175 (53.19)	86 (39.27)
95% CI	47.18 to 58.25	31.84 to 45.06	47.79 to 58.52	33.04 to 45.87
Comparison vs. PL				
Odds ratio	1.847		1.794	
95% CI	1.28 to 2.66		1.26 to 2.55	
P value[b]	0.0010		0.0011	
Responders at 10% Threshold				
FVC Responders[a]				
n (%)	218 (70.55)	116 (56.86)	229 (69.60)	140 (63.93)
95% CI	65.24 to 75.36	50.00 to 63.47	64.43 to 74.33	57.38 to 70.00
Comparison vs. PL				
Odds ratio	1.914		1.286	
95% CI	1.32 to 2.79		0.89 to 1.86	
P value[b]	0.0007		0.1833	

CI = confidence interval; FVC = forced vital capacity; NTB = nintedanib; PL = placebo; vs. = versus.
[a] Responder patients were those with no absolute decline greater than 5% or 10%, respectively, in FVC % predicted and in patients with an FVC evaluation at 52 weeks.
[b] Based on logistic regression with treatment, age, gender, height, and baseline FVC % predicted included. Patients with missing data were considered to be nonresponders.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 25: ABSOLUTE CHANGE FROM BASELINE IN OXYGEN SATURATION OF PERIPHERAL BLOOD OVER 52 WEEKS

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	309	204	329	219
Mean (SD)	95.9 (1.97)	95.9 (1.94)	95.8 (2.56)	95.7 (2.14)
Week 52				
N	248	166	270	181
Mean (SD)	95.7 (2.21)	95.4 (2.51)	95.5 (2.60)	95.3 (2.71)
Change from Baseline to Week 52				
N	248	166	270	181
Mean (SD)	–0.34 (2.097)	–0.65 (2.292)	–0.44 (2.675)	–0.60 (2.455)
Adjusted[a,b] mean (SE)	–0.24 (0.129)	–0.53 (0.150)	–0.39 (0.149)	–0.66 (0.174)
95% CI	–0.49 to 0.01	–0.82 to –0.23	–0.68 to –0.10	–1.00 to –0.32
Comparison vs. PL				
Adjusted[a,b] mean (SE)	0.29 (0.181)		0.27 (0.213)	
95% CI	–0.07 to 0.64		–0.15 to 0.69	
P value	0.1138		0.2032	

CI = confidence interval; MMRM = mixed model for repeated measures; NTB = nintedanib; PL = placebo; SD = standard deviation; SE = standard error; SpO_2 = oxygen saturation of peripheral blood; vs. = versus.
[a] Based on MMRM with fixed effects for treatment, visit, gender, age, height, treatment by visit, baseline SpO_2, and baseline SpO_2 by visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix.
[b] Adjusted mean is based on all analyzed patients in the model (not only patients with a change from baseline to week 52).
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 26: ABSOLUTE CHANGE FROM BASELINE IN DIFFUSION CAPACITY OF LUNG FOR CARBON MONOXIDE (MMOL/MIN/KPA) OVER 52 WEEKS

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	309	204	329	218
Mean (SD)	3.964 (1.2014)	3.963 (1.1126)	3.772 (1.2335)	3.748 (1.3164)
Week 52				
N	243	158	252	168
Mean (SD)	3.612 (1.3368)	3.740 (1.3835)	3.492 (1.4361)	3.482 (1.5029)
Change from Baseline to Week 52				
N	243	158	252	168
Mean (SD)	–0.373 (0.9277)	–0.354 (0.9684)	–0.354 (1.1389)	–0.485 (1.0795)
Adjusted[a,b] mean (SE)	–0.380 (0.0644)	–0.365 (0.0750)	–0.286 (0.0729)	–0.400 (0.0843)
95% CI	–0.507 to –0.254	–0.512 to –0.218	–0.429 to –0.143	–0.565 to –0.234
Comparison vs. PL				
Adjusted[a,b] mean (SE)	–0.015 (0.0896)		0.113 (0.1005)	
95% CI	–0.191 to 0.161		–0.084 to 0.310	
P value	0.8650		0.2600	

CI = confidence interval; DL_{CO} = diffusion capacity of lung for carbon monoxide; HGB = hemoglobin; MMRM = mixed model for repeated measures; NTB = nintedanib; PL = placebo; SD = standard deviation; SE = standard error; vs. = versus.
[a] Based on MMRM, with fixed effects for treatment, visit, gender, age, height, treatment by visit, baseline DL_{CO} (HGB corrected) (mmol/min/kPa), and baseline DL_{CO} (HGB corrected) (mmol/min/kPa) by visit; and random effect for patient. Within-patient errors are modelled by compound symmetry covariance matrix.
[b] Adjusted mean is based on all analyzed patients in the model (not only patients with a change from baseline to week 52).
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 27: PROPORTIONS OF PATIENTS HAVING RECEIVED OR QUALIFYING FOR A LUNG TRANSPLANT (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Lung Transplant Received, n (%)	4 (1.3)	1 (0.5)	0 (0)	2 (0.9)
Qualifying for Lung Transplant,[a] n (%)	34 (11.0)	26 (12.7)	48 (14.6)	37 (16.9)

DL_{CO} = diffusion capacity of lung for carbon monoxide; FVC = forced vital capacity; NTB = nintedanib; PL = placebo; SpO_2 = oxygen saturation of peripheral blood.
[a] Qualifying for a lung transplant was defined as FVC < 45% predicted, or DL_{CO} < 30% predicted, or SpO_2 < 88% at rest, at sea level.
Note: A patient could be counted in more than one category.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 28: COMPLIANCE WITH STUDY MEDICATION (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Overall Compliance[a] (%)				
N	295	202	320	212
Mean (SD)	96.05 (7.10)	97.07 (6.68)	96.71 (5.88)	96.43 (5.82)
Median (min, max)	98.33 (55.6, 116.9)	99.18 (34.9, 117.6)	98.48 (59.6, 120.0)	98.64 (57.5, 107.7)
Overall Compliance in Class, n (%)				
< 50%	0 (0)	1 (0.5)	0 (0)	0 (0)
≥ 50% to < 80%	11 (3.6)	3 (1.5)	4 (1.2)	4 (1.8)
≥ 80% to ≤ 120%	284 (91.9)	198 (97.1)	316 (96.0)	208 (95.0)
Missing	14 (4.5)	2 (1.0)	9 (2.7)	7 (3.2)

AE = adverse event; NTB = nintedanib; PL = placebo; SD = standard deviation.
[a] Compliance is calculated as the number of capsules actually taken multiplied by 100 divided by the number of capsules that should have been taken. Treatment interruptions due to AE were taken into account in the calculation by subtracting the duration of interruption due to the AE.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

TABLE 29: ABSOLUTE CHANGE FROM BASELINE IN EUROQOL 5-DIMENSIONS QUESTIONNAIRE VISUAL ANALOGUE SCALE SCORE OVER 52 WEEKS (TREATED SET)

	INPULSIS-1		INPULSIS-2	
	NTB 150 mg N = 309	PL N = 204	NTB 150 mg N = 329	PL N = 219
Baseline				
N	306	203	324	218
Mean (SD)	66.71 (17.42)	68.02 (16.34)	69.77 (18.84)	67.75 (16.47)
Median (min, max)	70.0 (8.0, 100.0)	70.0 (19.0, 100.0)	70.0 (19.0, 100.0)	70.0 (20.0, 100.0)
Change From Baseline at 52 Weeks				
N	247	160	265	178
Mean (SD)	–2.46 (18.92)	–5.88 (19.17)	–2.52 (16.95)	–5.60 (17.67)
Median (min, max)	0 (–90.0, 73.0)	–5.0 (–73.0, 51.0)	0 (–61.0, 60.0)	–3.5 (–80.0, 50.0)

NTB = nintedanib; PL = placebo; SD = standard deviation.
Source: INPULSIS-1 Clinical Study Report,[18] INPULSIS-2 Clinical Study Report.[19]

APPENDIX 5: VALIDITY OF OUTCOME MEASURES

Aim

To summarize the validity of the following outcome measures:

- forced vital capacity (FVC)
- six-minute walk test (6MWT)
- St. George's Respiratory Questionnaire (SGRQ).

Findings

TABLE 30: VALIDITY AND MINIMAL CLINICALLY IMPORTANT DIFFERENCE OF OUTCOME MEASURES

Instrument	Type	Validated?	MCID	References
FVC	Measures total expiratory volume; usually reported as a percentage of predicted for persons of same size, age, and sex	Yes	2% to 6%	du Bois et al. 2011[7]
6MWT[a]	Distance (metres) walked in six minutes on a flat surface	Yes	28 m 24 m to 45 m 29 m to 34 m	Swigris et al. 2010[39] du Bois et al. 2011[40] Holland et al. 2009[41]
SGRQ	SGRQ is a disease-specific measure of HRQoL that consists of 50 items with 76 responses. It was developed for patients with chronic airflow limitation. The questionnaire is divided into three dimensions: symptoms, activity, and impacts of the disease. The total score ranges from 0 to 100, where 0 indicates no impairment and 100 indicates greatest impairment.	Yes	6 to 13	Swigris et al. 2014[42] Swigris et al. 2010[13]

6MWT = six-minute walk test; FVC = forced vital capacity; HRQoL = health-related quality of life; SGRQ = St. George's Respiratory Questionnaire.

[a] The 6MWT has been included in the summary of the validity of outcome measures because some of the studies examining the validity, reliability, and correlations of the SGRQ were compared with the 6MWT. Also, it was an important efficacy outcome in the pirfenidone trials.

Forced vital capacity

FVC is the volume of air that can be forcibly exhaled from the lungs after taking the deepest breath possible. It is usually reported as the percentage of the volume predicted for a person of the same size, age, and sex. The test properties of FVC were examined using data from two randomized controlled trials in patients with mild to moderate idiopathic pulmonary fibrosis (IPF).[7] FVC showed good within-person reliability when repeated after a short interval (intraclass correlation 0.93).[7] Validity was tested by comparing FVC with other measures of IPF severity. The per cent predicted FVC was found to be weakly correlated with measures of functional status (6MWT), gas exchange, dyspnea, and health-related quality of life (HRQoL).[7] Similar results were found when the change in percent predicted FVC was correlated with the change in these measures.[7] Baseline FVC has shown an inconsistent relationship with mortality.[1] The change in per cent predicted FVC, however, was found to be predictive of mortality in patients with IPF.[1,7-10] A six-month absolute decrease in the per cent predicted FVC ≥ 10% was

associated with a 2.8-fold to 4.8-fold increase in risk of mortality relative to those with stable disease (defined as change in per cent predicted FVC < 5%) and with a two-fold increased risk of mortality relative to those with < 10% change in per cent predicted FVC.[9] An absolute decline of 5% to < 10% in per cent predicted FVC was associated with a two-fold increased risk of death relative to those with stable FVC values.[7,8] Of note, Richeldi et al.[43] reported that using the relative change in FVC as opposed to using the absolute change in FVC maximized the chance of identifying a ≥ 10% decline in FVC.[43] Using the relative change in FVC also did not sacrifice prognostic accuracy.[43] If, however, a ≥ 5% decline in FVC was the desired outcome, then the absolute change in FVC was observed to be more predictive.[43]

The minimal clinically important difference (MCID) for the per cent predicted FVC was between 2% and 6%, using anchor and distribution-based methods in a mild to moderate IPF population.[7] In mixed interstitial lung disease populations (including those with IPF), the MCID for the per cent predicted FVC was also reported as 6% of baseline.[44]

The enrolment of mixed populations including patients with other forms of interstitial lung disease or those who may not meet the American Thoracic Society (ATS) criteria for IPF could potentially introduce generalizability issues in that this may not be appropriate for strictly IPF patients.[8,9,41,45,46]

Six-minute walk test

6MWT assesses submaximal functional capacity. It measures the distance a patient can walk on a flat surface over six minutes.[39] In two studies, the 6MWT showed good within-person reproducibility when repeated after a short interval.[40,45] The correlation between 6MWT and physiologic, functional, dyspnea, and HRQoL measures were in the expected direction but were weak.[40] Changes in 6MWT showed responsiveness in predicting mortality.[40] The risk of death was 4.3 times higher in patients who had a greater than 50 m decrease in 6MWT and 3.6 times higher in those with a 26 m to 50 m decrease in distance walked, compared with those with up to a 25 m decline.[40] Shorter distances walked during a 6MWT were predictive of mortality in some studies.[46,47]

The MCID of the 6MWT in patients with IPF was evaluated in three studies using anchor and distribution-based methods.[39-41] Although the methods used varied, the MCID estimates were similar across studies and ranged from 24 m to 45 m.[39-41] It should be noted that other factors that are unrelated to pulmonary disease, such as lower extremity pain, may limit the distance walked during a 6MWT.[48]

St. George's Respiratory Questionnaire

SGRQ is a disease-specific measure of HRQoL that was specifically developed for patients with airway obstruction.[49,50] It was developed in 1992 to measure impaired health and perceived well-being in patients with airway disease and to meet the need for a sensitive measure of HRQoL.[51] The instrument has been used worldwide in studies and in clinical settings.[51] The SGRQ questionnaire includes questions regarding sleep disturbances, public embarrassment, and panic (which can be signs of depression or anxiety) as well as feeling like a nuisance to friends and family, employment, and recreation activities (which are indicative of social impact).[52] While the SGRQ was originally designed to measure HRQoL in patients with chronic obstructive pulmonary disease, it has also been used in many clinical trials to assess HRQoL in patients with IPF.[49,53]

The questionnaire contains 50 items and 76 weighted responses that are divided into three subscales: symptoms (eight items measuring the frequency of respiratory symptoms over a preceding period that may range from one month to one year), activity (16 items measuring the disturbances to a patient's

daily physical activity), and impacts (26 items measuring the psychosocial impact of the disease).[11,54,55] Items are weighted using empirically derived weights in order to determine the SGRQ total score, which ranges from 0 to 100, where 0 indicates no impairment and 100 indicates worst possible health.[54,56] A decrease in score indicates an improvement in HRQoL.[11,12]

Component scores for the symptoms, activity, and impacts domains can be calculated (also ranging from 0 to 100) in addition to the total score. In the symptoms domain, patients are asked to rate the appearance, frequency, and severity of respiratory symptoms (wheeze, breathlessness, cough, etc.) on a 5-point scale, where low scores indicate no or less severe symptoms and high scores indicate more severe symptoms.[12] A number of items in the symptoms component relate to the frequency of symptoms over the previous year.[57] Responses on the other two domains are mostly yes-or-no in nature. The activity domain deals with mobility and physical activity problems that either cause or are limited by breathlessness.[54] The impacts domain covers aspects involved in social functioning and psychosocial disturbances resulting from the obstructive airway disease (employment, panic, medication, and side effects).[57]

Changes in HRQoL have been reported using SGRQ, which has been validated in patients with IPF and can distinguish patients with differing IPF severity.[49] However, unlike FVC, this tool has not been shown to predict all-cause mortality, IPF-specific mortality,[49,53] or survival.[53] In IPF patients from China who were followed over a six-month period , changes in HRQoL measured using the SGRQ domains were significantly correlated with changes in total lung capacity (as a percentage of the predicted value).[58] In addition, the dyspnea score and changes within the dyspnea score were significantly correlated with the activity and impacts scores of the SGRQ, but not with the symptoms score.[58] The internal consistency of the SGRQ activity and impacts domains and the total score were reported as excellent in patients with IPF; however, there was lower internal consistency in the symptoms domain, presumably due to the focus of this domain on a range of respiratory symptoms that IPF patients do not experience.[42] It has also been postulated that this domain detracts from the face validity of the SGRQ in IPF and it may be more beneficial to remove this section for use in patients with IPF.[42] However, even in the case of the potential weakness surrounding the symptoms domain, the SGRQ performs reasonably well in IPF patients, mainly due to the fact that this domain contributes the least out of all of the domains to the total score.[42]

MCIDs for the SGRQ based on distribution-based estimates were observed to range from 6 through 13.[13] In the review by Swigris et al.,[42] the MCID were reported as 8, 5, 7, and 7 for the SGRQ symptoms, activity, impacts, and total score, respectively.

Conclusion

FVC appears to be a valid outcome measure in patients with mild or moderate IPF. An absolute change of at least 5% or 10% in the per cent predicted FVC is predictive of mortality. Estimates of the MCID in IPF for the per cent predicted FVC range from 2% to 6%. The SGRQ has excellent internal consistency in its activity and impacts domains along with the total score. The MCID for the SGRQ in IPF was determined to be between 6 and 13.

APPENDIX 6: SUMMARY OF INDIRECT COMPARISONS

Introduction

Background

In the absence of randomized controlled trials directly comparing nintedanib with other pharmaceuticals, this Appendix provides a summary and critical appraisal of:

- A systematic review with indirect comparison of nintedanib, pirfenidone, and N-acetylcysteine provided by the manufacturer as part of its CADTH Common Drug Review (CDR) submission package[26]
- A systematic review with indirect comparison of nintedanib, pirfenidone, and N-acetylcysteine identified in a literature search during the CDR review.[37] This is an update of a mixed treatment comparison of all available drug treatments (nintedanib, pirfenidone, N-acetylcysteine, azathioprine, thalidomide, sildenafil) prepared to support a National Institute for Health and Care Excellence (NICE) health technology assessment (HTA),[59] which was also published with the inclusion of non-drug therapies.[60]

Methods

One indirect comparison was provided by the manufacturer as part of the CDR submission package.[26] Hereafter it will be referred to as Ofev CDR.

A search for published articles and grey literature (see Appendix 2 for details of search terms) conducted in support of the CDR review identified two published articles[37,60] and one HTA report[59] by the same authors. The HTA described an indirect comparison of all drug therapies.[59] The same methods were used in a published indirect comparison of drug plus non-drug therapies[60] and in an update restricting the treatments to nintedanib, pirfenidone, and N-acetylcysteine.[37] The first two reports included only one nintedanib study, the phase 2, dose-ranging TOMORROW study, whereas the update included the two nintedanib pivotal studies. Hereafter the updated analysis will be referred to as Loveman et al.

Description of Indirect Comparisons Identified

The inclusion criteria for the two systematic reviews are summarized in Table 31. The Ofev CDR review omitted detail of the inclusion and exclusion criteria for the systematic review, with the exception of the outcomes. It refers to the draft scope of the NICE HTA for nintedanib in its description of search and outcomes, suggesting that it followed a similar strategy.

TABLE 31: INCLUSION CRITERIA FOR THE PUBLISHED SYSTEMATIC REVIEWS

	Ofev CDR 2015	Loveman et al. 2015
Patient Population	Not reported	People with a confirmed diagnosis of IPF
Interventions and Comparators	Not reported	Treatment with: • nintedanib • pirfenidone • N-acetylcysteine alone or in combination
Outcomes	**Efficacy outcomes** Overall survival, acute exacerbation rates, pulmonary function parameters, progression-free survival, physical function parameters, health-related quality of life Timing and duration of follow-up was as reported in individual studies **Harms outcomes** Adverse events, tolerability	**Efficacy outcomes** Mortality, acute exacerbations, FVC Timing and duration of follow-up was as reported in individual studies **Harms outcomes** Not included in this update *Based on previous review, the authors indicated that other outcomes were not consistently enough reported to meta-analyze.*
Study Design	RCTs; no restriction on duration	RCTs; no restriction on duration

FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; RCT = randomized controlled trial.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Review and Appraisal of Indirect Comparisons

As the two indirect comparisons included an almost identical set of trials, they are discussed together.

Objectives and rationale

In both reports, an indirect comparison was planned and conducted because there were no trials directly comparing nintedanib with other active treatments for idiopathic pulmonary fibrosis (IPF). Two such treatments, N-acetylcysteine and pirfenidone, were identified in a review completed as preparation for a single technology assessment by NICE, referenced by both sets of authors.

Methods

Study eligibility and selection process

The review inclusion criteria are presented in Table 31.

Eleven randomized controlled trials were identified by Ofev CDR: three trials of nintedanib, five of pirfenidone, and three of N-acetylcysteine. Individual patient data from the two pivotal nintedanib trials INPULSIS-1 and INPULSIS-2 were pooled for post hoc analyses of certain outcomes, or pooled results from individual patient data were drawn from the summaries of clinical efficacy and clinical safety. Results from pooled individual patient data from the pivotal pirfenidone trials CAPACITY 004 and CAPACITY 006 were obtained from the European Medicines Agency (EMA) Committee for Medicinal Products for Human Use (CHMP) Assessment Report.

Loveman et al. searched MEDLINE, Embase, the Centre for Research and Dissemination, and The Cochrane Library from June 2013 to May 2014,[61] updating their previous search, which included Science Citation Index, BIOSIS, conference proceedings and grey literature, and trial registration and research-in-progress databases.[62] There were no language restrictions. Trials were also identified by handsearching

reference lists and consulting experts. The review inclusion criteria are presented in Table 31. Two reviewers selected studies with differences resolved by consensus or by involving a third reviewer.

Eleven randomized controlled trials were identified by Loveman et al.: three trials of nintedanib, five of pirfenidone, and three of N-acetylcysteine. Published data were used throughout. For INPULSIS-1 and INPULSIS-2, the results of pooling of individual patient data were available for overall mortality and respiratory mortality, while the two trials were treated separately for the other outcomes. Data were extracted by one reviewer and checked by a second, and differences were resolved by discussion with a third reviewer.

A summary of the design, interventions, and key inclusion and exclusion criteria of the studies identified by either systematic review is presented in Table 32. Tomioka et al. 2012 was included only in Ofev CDR, and Raghu et al. 2012 was included only in Loveman et al.

TABLE 32: DESIGN OF STUDIES INCLUDED IN INDIRECT COMPARISONS OF PHARMACEUTICAL TREATMENT IN IDIOPATHIC PULMONARY FIBROSIS

Reference Study Name	Design, Follow-Up	Interventions[a]	Patients (n)	Inclusion	Exclusion
Richeldi et al., 2014 INPULSIS-1	DB RCT, 52 weeks	Nintedanib 150 mg b.i.d. Placebo oral	309 204	• Age ≥ 40 years • IPF based upon the most recent ATS/ERS/JRS/ALAT guidelines • FVC % predicted ≥ 50% • DL_{CO} 30% to 79% predicted	• Pre-bronchodilator FEV_1/FVC < 0.7 • AST, ALT, and bilirubin > 1.5 ULN • Cardiac disease, bleeding, or thrombotic risk
Richeldi et al., 2014 INPULSIS-2	DB RCT, 52 weeks	Nintedanib 150 mg b.i.d. Placebo oral	329 219		
Richeldi et al., 2011 TOMORROW	DB RCT, 52 weeks	Nintedanib 50 mg q.d. Nintedanib 50 mg b.i.d. Nintedanib 100 mg b.i.d. Nintedanib 150 mg b.i.d. Placebo	86 86 86 85 85	• Age ≥ 40 years • IPF consistent with the ATS/ERS 2000 criteria • < 5 years since diagnosis • HRCT within last year • FVC % predicted ≥ 50% • DL_{CO} 30% to 79% predicted • PaO_2 on room air ≥ 55 mm Hg or ≥ 50 mm Hg (above 1,500 m)	• Life expectancy < 2.5 years for other than IPF • Predisposition to bleeding or thrombosis, concomitant anticoagulation medication • Elevated liver enzymes • Likelihood of lung transplantation during study
King et al., 2014 ASCEND	DB RCT, 52 weeks	Pirfenidone 2,403 mg daily Placebo	278 277	• Age 40 to 80 years • FVC % predicted 50% to 90% • DL_{CO} 30% to 90% predicted	• Post-bronchodilator FEV_1/FVC < 0.8 • Lung transplant expected within 1 year, or on waiting list • Unstable or deteriorating cardiac disease, pulmonary disease other than IPF within 6 months
Noble et al., 2011 CAPACITY 004	DB RCT, 72 weeks	Pirfenidone 1,197 mg daily Pirfenidone 2,403 mg daily Placebo	87 174 174	• Age 40 to 80 years • IPF diagnosed in the previous 48 months • No improvement in measures of disease severity over the preceding year • FVC % predicted ≥ 50% • DL_{CO} ≥ 35% predicted	• On waiting list for lung transplant
Noble et al., 2011 CAPACITY 006	DB RCT, 72 weeks	Pirfenidone 2,403 mg daily Placebo	171 173		

Reference Study Name	Design, Follow-Up	Interventions[a]	Patients (n)	Inclusion	Exclusion
				• Either FVC % predicted or DL_{CO} ≤ 90% • 6MWT ≥ 150 m	
Taniguchi et al., 2010	DB RCT, 52 weeks	Pirfenidone 1,800 mg daily Pirfenidone 1,200 mg daily Placebo	108 55 104	• Age 20 to 75 years • IPF diagnosed using ATS/ERS and clinical diagnostic criteria guidelines for idiopathic interstitial pneumonia in Japan • Independently evaluated HRCT and probable UIP by surgical lung biopsy • ≥ 5% difference between resting SpO_2 and lowest SpO_2 during a 6-minute steady-state exercise test • Lowest SpO_2 during the 6MWT of ≥ 85% on room air	• Decrease in symptoms (preceding 6 months) • Immunosuppressants or oral corticosteroids at > 10 mg/day – 1 (preceding 3 months)
Azuma et al., 2005	DB RCT, 39 weeks	Pirfenidone 1,800 mg daily Placebo	72 35	• Age 20 to 75 years • Histological UIP not mandatory; HRCT evidence of definite or probable UIP required: bibasilar inspiratory crackles, abnormal PFTs, increased serum levels of damaged pneumocyte markers • PaO_2 ≥ 70 mm Hg at rest and demonstrated SpO_2 of ≤ 90% during exertion while breathing air, 1 month before enrolment	• Decrease in symptoms (preceding 6 months) • Immunosuppressants or oral corticosteroids at > 10 mg/day – 1 (preceding 3 months)

Reference Study Name	Design, Follow-Up	Interventions[a]	Patients (n)	Inclusion	Exclusion
Martinez et al., 2014 PANTHER	DB RCT, 60 weeks	N-acetylcysteine 600 mg t.i.d. Placebo	133 131	• Adults 35 to 85 years with IPF per modified criteria of the ATS/ERS (2011) • Diagnosed from HRCT or biopsy ≤ 48 months before enrolment • FVC % predicted ≥ 50% • DL_{CO} % predicted ≥ 30%	• FEV_1/FVC ratio < 0.65 • PaO_2 on room air < 55 mm Hg • Residual volume > 120% predicted • Post-bronchodilator FVC differing by > 11% • Listed for lung transplantation
Raghu et al., 2012[b] PANTHER	DB RCT, 32 weeks[c]	N-acetylcysteine 600 mg t.i.d. (triple therapy)[d] Placebo	77[c] 78		
Homma et al., 2012	OL RCT, 48 weeks	N-acetylcysteine 352 mg b.i.d. by inhalation daily Placebo	38 38	• Age 50 to 79 years • Histological UIP not mandatory; HRCT evidence and other clinical features required • PaO_2 ≥ 70 mm Hg at rest and demonstrated SpO_2 of ≤ 90% on room air during 6MWT	• Improvement in symptoms during the preceding 3 months • Use of N-acetylcysteine, immunosuppressive drugs, oral prednisolone or pirfenidone
Tomioka et al., 2005[e]	OL RCT, 52 weeks	N-acetylcysteine 352 mg b.i.d. by inhalation daily Bromhexine 4 mg daily	10 12	Not reported	Not reported

6MWT = 6-minute walk test; ALAT = Latin American Thoracic Association; ALT = alanine aminotransferase; AST = aspartate aminotransferase; ATS = American Thoracic Society; b.i.d. = twice daily; DB = double-blind; DL_{CO} = diffusion capacity of lung for carbon monoxide; ERS = European Respiratory Society; FEV_1 = forced expiratory volume in one second; FVC = forced vital capacity; HRCT = high-resolution computed tomography; IDC = indirect comparison; IPF = idiopathic pulmonary fibrosis; JRS = Japanese Respiratory Society; OL = open-label; PaO_2 = partial pressure of oxygen in arterial blood; PFT = pulmonary function test; q.d. = once daily; RCT = randomized controlled trial; SpO_2 = oxygen saturation of peripheral blood; t.i.d. = three times daily; UIP = usual interstitial pneumonia; ULN = upper limit of normal.

[a] Dosing is oral unless otherwise indicated.

[b] Only included in Loveman et al. 2015. IDC data extracted from Raghu et al., 2012.[23]

[c] Triple therapy group was discontinued prematurely due to safety concerns.

[ed] Triple therapy consisted of oral N-acetylcysteine 600 mg t.i.d., azathioprine to a maximum 150 mg/day, and prednisone starting at 0.5 mg/kg body weight and tapered to 0.15 mg/kg over 25 weeks.

[e] Only included in Ofev CDR.

Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

TABLE 33: DEFINITIONS OF KEY OUTCOMES USED IN THE INDIRECT COMPARISONS OF PHARMACEUTICAL TREATMENT IN IDIOPATHIC PULMONARY FIBROSIS

Reference Study Name	Follow-Up Duration	Overall Mortality	Acute Exacerbation	Lung Function
Richeldi et al., 2014 INPULSIS-1	52 weeks	Death from any cause during follow-up	Otherwise unexplained clinical features within one month during follow-up, including:[a] • progression of dyspnea over several days to 4 weeks • new diffuse pulmonary infiltrates in chest radiography, HRCT, or new parenchymal abnormalities (ground-glass opacities) without pneumothorax or pleural effusion • exclusion of any known causes of acute worsening in accordance with routine clinical practice	FVC % predicted, FVC in litres Decline in FVC % predicted ≥ 10% at 52 weeks
Richeldi et al., 2014 INPULSIS-2	52 weeks			
Richeldi et al., 2011 TOMORROW	52 weeks	Death from any cause during follow-up	Met all the following criteria during follow-up: • onset or worsening of dyspnea in previous 30 days • new diffuse pulmonary infiltrates in chest radiography, HRCT, or new parenchymal abnormalities (ground-glass opacities) without pneumothorax or pleural effusion • exclusion of any known causes of acute worsening in accordance with routine clinical practice • decrease in $PaO_2 \geq 10$ mm Hg or $PaO_2/FiO_2 < 225$ since last visit • other causes excluded, e.g., infection, congestive heart failure, pulmonary embolism, etc.	FVC % predicted, FVC in litres Decline in FVC % predicted ≥ 10% at 52 weeks
King et al., 2014 ASCEND	52 weeks	Death from any cause during follow-up	Acute exacerbations not reported	FVC in litres Decline in FVC > 10% predicted or death at 52 weeks
Noble et al., 2011 CAPACITY 004	72 weeks	Death from any cause during follow-up	Met all the following criteria within a 4-week period during follow-up: • worsening of PaO_2 (≥ 8 mm Hg drop from recent pre-worsening value) • clinically significant worsening of dyspnea • new ground-glass opacities on HRCT in ≥ 1 lobe • other causes excluded, e.g., cardiac, thrombotic, aspiration, or infectious	FVC in litres Publication reported number of patients with change in FVC ≥ 10% at 72 weeks, possibly including deaths[26] European Medicines Agency report on pirfenidone included FVC % predicted ≥ 10%.[26]
Noble et al., 2011 CAPACITY 006	72 weeks	Death from any cause during follow-up		
Taniguchi et al., 2010	52 weeks	Did not report detailed overall mortality; reported deaths due to respiratory failure	Met all the following criteria within a month during follow-up: • increased dyspnea • new ground-glass opacities on HRCT	FVC in litres, time of measurement not reported

Reference Study Name	Follow-Up Duration	Overall Mortality	Acute Exacerbation	Lung Function
			• PaO_2 in resting arterial blood lower by > 10 mm Hg from previous • other causes excluded, e.g., infection, pneumothorax, cancer, pulmonary embolism, congestive heart failure Serum levels of CRP, LDH, and markers of interstitial pneumonias were usually elevated.	
Azuma et al., 2005	39 weeks	No information about overall mortality; death reported after an acute exacerbation event	Met all the following criteria during follow-up: • worsening, unexplained clinical features within 1 month • progression of dyspnea over a few days to < 5 weeks • new radiographic or HRCT parenchymal abnormalities without pneumothorax or pleural effusion (i.e., new, superimposed ground-glass opacities) • decrease in PaO_2 by ≥ 10 mm Hg • infection excluded	FVC in litres, time of measurement not reported
Martinez et al., 2014 PANTHER	60 weeks	Number of patients dead at the end of the study	Met all the following criteria during follow-up: • aggravation of dyspnea within 1 month • PaO_2/FiO_2 < 225 • newly developing pulmonary infiltrates on chest radiography • absence of apparent infection or heart disease	FVC in litres, time of measurement not reported Decline of FVC > 10% or death
Raghu et al., 2014[b] PANTHER	32 weeks[c]			FVC in litres, time of measurement not reported
Homma et al., 2012	48 weeks	Reported, not defined	No reported definition for exacerbations	FVC in litres, time of measurement not reported
Tomioka et al., 2005[d]	52 weeks	Not reported	No reported definition for exacerbations	Measure not reported

CRP = C-reactive protein; FiO_2 = fraction of inspired oxygen; FVC = forced vital capacity; HRCT = high-resolution computed tomography; LDH = lactate dehydrogenase; PaO_2 = partial pressure of oxygen in arterial blood.

[a] In the trials, investigator-identified exacerbations were reviewed by an adjudication committee. In the IDC, the investigator-identified end points were used.
[b] Only included in Loveman et al. 2015.
[c] Triple therapy group was discontinued prematurely due to safety concerns.
[d] Only included in Ofev CDR 2015.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Data extraction

Table 34 summarizes the dosing and demographic information across the included studies.

The studies recruited adult populations. The three nintedanib studies recruited patients aged ≥ 40 years. Two of the pirfenidone studies recruited patients aged 40 to 80 years, and two recruited patients aged 20 to 75 years. The PANTHER study (Martinez 2014, Raghu 2012) recruited patients aged 35 to 80 years (Table 32). This variation of inclusion criteria was not reflected in the mean ages of the study patients (Table 34). There was a male predominance, with more than three-quarters of the patients being men in most studies, particularly Azuma 2005. This was somewhat higher than the proportion in the overall IPF population, although the rarity and inconsistent diagnostic criteria mean that the epidemiology is uncertain.

Four studies (Tomioka 2012, Homma 2012, Taniguchi 2010, Azuma 2005) were conducted in Japan, with a predominately or exclusively Japanese population. The implications for the dosing of pirfenidone are discussed below. The PANTHER study (Martinez 2014, Raghu 2012) was conducted in the US. The remaining included studies were multinational.

The majority of studies recruited patients with forced vital capacity (FVC) per cent predicted ≥ 50%. The two major pirfenidone studies also imposed an upper limit on FVC per cent predicted. King 2014 required FVC per cent predicted ≤ 90%, and Noble 2011 required that either FVC per cent predicted or diffusion capacity of lung for carbon monoxide (DL_{CO}) be ≤ 90% (Table 32). This had the potential to select for a more advanced population, and the intent of King et al. 2014 was to recruit patients at greater risk of progression. Although the mean disease duration was the same, at around 1.7 years, patients recruited by King 2014 had a lower FVC per cent predicted (67.8% to 68.6%) than those recruited by Richeldi 2014 (78.1% to 80.5%). This would be considered a clinically significant difference, possibly representing more aggressive disease. Noble 2011 had a slightly higher representation of patients with less than one year of disease with slightly poorer mean lung function (73.1% to 74.9%) than Richeldi 2014. Conversely, the dose-ranging study for nintedanib (Richeldi 2011) appears to have recruited patients with shorter duration of disease than Richeldi 2014 (around one year), but with similar lung function.

The patients recruited for the four Japanese studies tended to have a longer disease duration or lower representation of patients with less than one year since diagnosis. Homma 2012 in particular had a mean disease duration of 3.0 to 3.2 years, compared with 0.9 to 1.7 years in the other studies. Ofev CDR review identified this as a potentially important source of variation, given the median survival of three to five years.

The two indirect comparisons took a similar approach to dose selection with nintedanib and pirfenidone:

- From the dose-ranging studies for nintedanib (Richeldi 2011) and pirfenidone (Noble 2011), they selected the patients who received the licensed doses of 150 mg twice daily and 2,403 mg daily, respectively.
- From the studies of pirfenidone in Japanese patients (Azuma 2005, Taniguchi 2010), they selected patients dosed at 1,800 mg/day, which is the approved dose in Japan. This dose had been used to determine the pirfenidone dose for the pivotal trials, based on the difference between the mean body weights of a predominately Japanese and a predominately North American population; they therefore rationalized that this was substantially similar.

- As N-acetylcysteine was not approved for treating IPF, no preconditions were set for dosing. The two indirect comparisons differed in their network models: Ofev CDR assumed that all routes of administration and doses were equivalent, and included the three N-acetylcysteine trials in a single node, while Loveman et al. separated them into three nodes — triple therapy (N-acetylcysteine, azathioprine, prednisone), oral administration, and inhaled (see Evidence Network).

Treatment duration ranged from 32 to 72 weeks overall. The three nintedanib trials were all 52 weeks in duration, the pirfenidone trials lasted from 39 (Azuma 2005) to 72 weeks (Noble 2011), and the N-acetylcysteine trials were planned to last from 48 to 60 weeks. Early termination of the triple therapy group in the PANTHER trial (Raghu 2012) reduced that to 32 weeks.

TABLE 34: SUMMARY OF DEMOGRAPHIC INFORMATION FROM STUDIES INCLUDED IN INDIRECT COMPARISON OF PHARMACEUTICAL TREATMENT OF IDIOPATHIC PULMONARY FIBROSIS

Reference Study Name	Drug, Dose[a]	Age	Gender: Male	Time Since Diagnosis	Disease Duration	FVC % Predicted	DL_{CO} % Predicted	6-Minute Walk Test
		Mean (SD), years	N (%)	Mean (SD), years	N (%) < 1 year	Mean (SD)	Mean (SD)	Mean (SD), metres
Richeldi 2014 INPULSIS-1	Nintedanib 150 mg b.i.d.	66.9 (8.4)	251 (81.2)	1.7 (1.4)	133 (43.0)	79.5 (17.0)	47.5 (12.3)	NR
	Placebo	66.9 (8.2)	163 (79.9)	1.6 (1.4)	92 (45.1)	80.5 (17.3)	47.5 (11.7)	NR
Richeldi 2014 INPULSIS-2	Nintedanib 150 mg b.i.d.	66.4 (7.0)	256 (77.8)	1.6 (1.3)	141 (42.9)	80.0 (18.1)	47.0 (14.5)	NR
	Placebo	67.1 (7.5)	171 (78.1)	1.6 (1.3)	101 (46.1)	78.1 (19.0)	46.6 (14.8)	NR
Richeldi 2011 TOMORROW	Nintedanib 150 mg b.i.d.	65.4 (7.8)	65 (76.5)	1.0 (1.2)	57 (67.2)	79.1 (18.5)	47.5 (11.0)	437.0 (13.69)
	Placebo	64.8 (8.6)	63 (74.1)	1.4 (1.5)	48 (56.5)	81.7 (17.6)	48.4 (12.9)	411.1 (15.90)
King 2014 ASCEND	Pirfenidone 2,403 mg daily	68.4 (6.7)	222 (79.9)	1.7 (1.1)	NR	67.8 (11.2)	43.7 (10.5)	415.0 (98.5)
	Placebo	67.8 (7.3)	213 (76.9)	1.7 (1.1)	NR	68.6 (10.9)	44.2 (12.5)	420.7 (98.1)
Noble 2011 CAPACITY 004	Pirfenidone 2,403 mg daily	66.8 (7.9)	123 (72)	NR	83 (48)	74.9 (13.2)	47.8 (9.8)	411.1 (91.8)
	Placebo	67.0 (7.8)	124 (72)	NR	81 (47)	73.1 (14.2)	47.4 (9.2)	410.0 (90.9)
Noble 2011 CAPACITY 006	Pirfenidone 2,403 mg daily	65.7 (8.2)	118 (68)	NR	100 (58)	74.5 (14.5)	46.4 (9.5)	378.0 (98.5)
	Placebo	66.3 (7.5)	128 (74)	NR	107 (62)	76.2 (15.5)	46.1 (10.2)	420.7 (98.1)
Taniguchi 2010	Pirfenidone 1,800 mg daily	65.4 (6.2)	85 (78.7)	NR	38 (35.2)	77.3 (16.8)[b]	52.1 (16.8)	NR
	Placebo	64.7 (7.3)	81 (77.9)	NR	20 (36.4)	79.1 (17.4)[b]	55.2 (18.2)	NR
Azuma 2005	Pirfenidone 1,800 mg daily	64.0 (7.1)	62 (86)	NR	20 (28)	81.6 (20.3)[b]	57.6 (17.2)	NR
	Placebo	64.3 (7.6)	33 (94)	NR	6 (17)	78.4 (17.2)[b]	57.7 (13.8)	NR
Martinez 2014 PANTHER	N-acetylcysteine 600 mg t.i.d.	68.3 (8.4)	107 (80.5)	1.0 (1.0)	NR	72.2 (15.9)	44.7 (10.8)	371 (116)
	Placebo	67.2 (8.2)	98 (74.8)	1.1 (1.0)	NR	73.4 (14.3)	46.0 (12.2)	375 (105)
Raghu 2012[c] PANTHER	N-acetylcysteine 600 mg t.i.d. (triple therapy)[d]	68.6 (7.3)	59 (77)	0.9 (1.1)	NR	69.3 (15.1)	42.1 (10.2)	362.0 (113.0)
	Placebo	67.9 (8.1)	57 (73)	1.1 (1.0)	NR	72.1 (14.1)	45.3 (12.4)	368.9 (117.3)
Homma 2012	N-acetylcysteine 352 mg b.i.d. by inhalation daily	67.6 (6.4)	29 (76)	3.0 (3.4)	9 (23.7)	89.2 (17.8)	72.3 (17.8)	NR
	Placebo	68.2 (7.7)	29 (76)	3.2 (2.5)	5 (13.2)	88.7 (15.5)	64.4 (20.1)	NR
Tomioka 2005[e]	N-acetylcysteine 352 mg b.i.d. by inhalation daily	70 (4.9)	NR	NR	NR	67.6 (15.7)[b]	64.7 (21.7)	385 (90)
	Bromhexine 4 mg daily	70 (5.3)	NR	NR	NR	76.6 (19.1)[b]	60.7 (16.7)	390 (116)

b.i.d. = twice daily; DL_{CO} = diffusion capacity of lung for carbon monoxide; FVC = forced vital capacity; NR = not reported; SD = standard deviation; t.i.d. = three times daily.

[a] Only doses included in the indirect comparison are given.

[b] Vital capacity (VC) % predicted rather than FVC % predicted.

[c] Only included in Loveman et al. 2015. Indirect comparison data extracted from Raghu 2012.[23]

[d] Triple therapy consisted of oral N-acetylcysteine 600 mg t.i.d., azathioprine to a maximum 150 mg/day, and prednisone starting at 0.5 mg/kg body weight and tapered to 0.15 mg/kg over 25 weeks.

[e] Only included in Ofev CDR 2015.

Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Comparators

Both indirect comparisons included nintedanib, pirfenidone, and N-acetylcysteine as drug treatment comparisons. Pirfenidone has been approved for the treatment of IPF in Europe, US, and Canada. It is taken orally; the approved dose in Canada is 2,403 mg daily (801 mg three times daily).[24]

N-acetylcysteine is used in oral or inhaled form, approved as a mucolytic in patients with respiratory diseases and as an antidote to acetaminophen overdose. Its use in IPF is based on observed abnormalities in the oxidation pathway and deficiencies in glutathione in IPF. It has been tested for use in IPF both in its oral and its inhaled form.[63]

According to the clinical expert involved in the review of nintedanib, pirfenidone is the key comparator and, as such, its inclusion in the indirect comparisons is appropriate for the Canadian setting. The clinical expert noted that although N-acetylcysteine has historically been used in treating patients with IPF, it is less relevant today based on the results of the PANTHER trials and the emergence of pirfenidone and nintedanib.

Outcomes

The efficacy outcomes identified by the authors of Ofev CDR were as follows:[26]

- overall survival, as measured by all-cause mortality
- acute exacerbations, as defined in the individual trials (see Table 33)
- loss of lung function, as a categorical change in FVC per cent predicted ≥ 10% (see Table 33) (some studies included death in their definition of loss of lung function)
- progression-free survival, where progression was defined for the analysis as time to confirmed decline in FVC per cent predicted ≥ 10% or decline in DL_{CO} per cent predicted ≥ 15%, or death
- physical function, as measured by the six-minute walk test (6MWT).

The safety outcomes were as follows:[26]

- withdrawals due to adverse events
- withdrawals for any reason
- serious cardiac adverse events, including arterial thrombotic events including myocardial infarction
- serious gastrointestinal events, including gastrointestinal disorders, gastrointestinal perforation, and elevated liver enzymes and bilirubin.

The efficacy outcomes identified by Loveman et al. were as follows:[37]

- all-cause mortality
- respiratory mortality
- rate of decline of FVC
- loss of lung function, as a categorical change in FVC per cent predicted of 10% (see Table 33)
- acute exacerbations, as defined in the individual trials (see Table 33).

Loveman et al. considered the evidence network too incomplete to conduct indirect comparisons of other outcomes, including safety outcomes.[37]

All the above end points were identified in the CDR review, with the exception of withdrawal for any reason. Mortality and lung function were identified as key end points. Key end points for the CDR review that were not included in either indirect comparison were quality of life and health resource utilization. Additional end points identified for the CDR review were lung transplantation, symptoms (e.g., dyspnea,

cough, shortness of breath, fatigue), patient adherence, mental health/psychological well-being, functional capacity/productivity, all adverse events, and all serious adverse events. The availability of data determined the presentation of end points in the indirect comparisons.

Individual study end points for mortality, exacerbations, and loss of lung function are presented in Table 33.

The pivotal studies for nintedanib and pirfenidone reported death from any causes, whereas other studies reported subsets of mortality that for the purposes of the analysis were assumed to estimate mortality.

The definition of exacerbations varied across studies, although they have common requirements of progression of dyspnea (usually over the previous four weeks), new abnormalities on diagnostic imaging, and the requirement to eliminate other causes for the deterioration. In practice, identification of exacerbations was likely to have varied between sites within the same study, as local practice would influence the extent of investigation into other causes. In the two INPULSIS studies (Richeldi 2015), blinded independent adjudicators confirmed 60% of exacerbations identified for the nintedanib group and 40% of those identified for the placebo group.

Lung function in the form of change in FVC was measured in all studies either as change of percentage of predicted or as change in FVC in litres. Loveman et al. used standard mean difference to convert FVC and vital capacity to a common scale for an analysis of change from baseline, using Hedges' adjusted g method to correct for small sample bias. The standardized mean differences were converted to log odds ratios.

Loss of lung function according to a definition of ≥ 10% decrease in FVC per cent predicted was reported or could be calculated for five studies. Of these, one included patients who had died in the count of those with lost function, and one possibly did so, and one reported FVC rather than FVC percent predicted. Ofev CDR assumed that the changes in FVC and FVC per cent predicted would be proportionate, and therefore the proportions reaching the end point would be similar for both measures. They conducted a sensitivity analysis recalculating all measures of loss of lung function to include deaths.

Progression-free survival was calculated for patients in the nintedanib pivotal trial (Richeldi 2014) according to the methods used in Noble 2011, allowing direct comparison of nintedanib and pirfenidone. Progression-free survival was defined as time to confirmed ≥ 10% decline in FVC per cent predicted, or ≥ 15% decline in DL_{CO} per cent predicted, or death. A hazard ratio was calculated.

The 6MWT was similarly defined across those studies that reported data.

The definition of withdrawals due to adverse events was consistent across the studies that reported data.

Serious cardiac and gastrointestinal events for the indirect comparison were limited to those reported with an incidence > 5%, differing by > 1.5 times between study groups. Serious cardiac and gastrointestinal events were reported as a separate category in Richeldi 2011, Richeldi 2014, Noble 2011, and Martinez 2014. Serious adverse events coded to the broader cardiac and the gastrointestinal

system organ classes were obtained from the manufacturer's submission for pirfenidone (CAPACITY trials). No adjustment was made within the analysis for the heterogeneity of definitions.

Quality assessment of included studies

Ofev CDR assessed quality according to the risk of bias domains identified by the Cochrane Collaboration. Studies were allocated low, moderate, or high risk of bias for sequence generation, allocation concealment, double-blinding, blinding of outcomes, and incomplete outcome data. Risk of bias was considered low for all studies except for Homma 2012 and Tomioka 2010. Homma 2012 was not blinded, and outcomes were not reported for patients excluded from the analysis for pertinent reasons such as due to exacerbation, disease progression, or pneumonia. Tomioka 2010 was also not blinded, had unclear primary and secondary outcomes, and did not report outcomes for patients excluded from the analysis due to respiratory failure, death, loss to follow-up, or other diseases.

No studies were excluded from the primary analysis. Homma 2012 was excluded in sensitivity analyses for end points it contributed to. Azuma 2005 was also excluded in sensitivity analyses due to possible reporting bias not indicated in the risk of bias assessment.

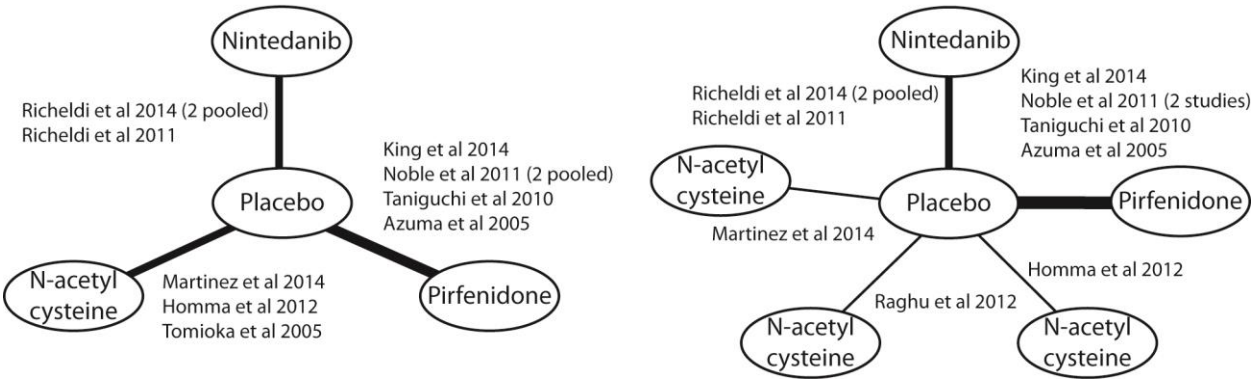
Loveman et al. assessed risk of bias on the basis of adequacy of allocation of concealment before treatment assignment. They assessed all but one study (Homma 2012) as having randomization procedures and blinding (Tomioka 2010 was not included). The majority reported an intention-to-treat analysis. Studies were not excluded on the basis of risk of bias.

Evidence Network

The overall evidence network for the two indirect comparisons for pharmaceutical trials in IPF is shown in Figure 3, indicating the individual trials contributing to each comparison. Depending upon the availability of end point data, the number of trials varied between end point analyses.

The main difference between the two study networks is that Ofev CDR included all the N-acetylcysteine trials (oral therapy, inhaled therapy, and triple therapy) in a single node, while Loveman et al. separated them. In addition, Loveman et al. treated the two CAPACITY trials (Noble 2011) separately, while Ofev CDR pooled the data.

FIGURE 3: OVERALL NETWORK FOR PHARMACEUTICAL TRIALS IN IDIOPATHIC PULMONARY FIBROSIS FOR OFEV CDR 2015 (LEFT) AND LOVEMAN ET AL. 2015 (RIGHT)



CDR = CADTH Common Drug Review.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Indirect comparison methods

Most end points analyzed in Ofev CDR were dichotomous, or were converted to dichotomous, and analyzed as odds ratios; 6MWT was the exception, analyzed as change from baseline. Events (e.g., mortality, exacerbations, adverse events) were collected across the entire duration of follow-up, which varied across studies (see Table 33). Where measurements were repeated (e.g., lung function tests) the timing was not always described, but where it was made explicit (e.g., for loss of lung function in King 2014), the estimate for the time point at end of follow-up was used, based on observed cases and with no correction for missing data.

Direct estimates were obtained for individual treatments versus placebo by pooling results according to standard meta-analytic methods. Both fixed effects (Mantel–Haenszel) and random effects (DerSimonian and Laird) models were used. Statistical heterogeneity of the pooled studies was numerically assessed using Cochrane's Q test (although the authors noted its low power for analyses involving a small number of studies) and the I^2 statistic. For the latter, 0% was interpreted as no heterogeneity, and 25%, 50%, and 75% represented low, medium, and high heterogeneity. Clinical heterogeneity was assessed by appraisal of patient characteristics and study design.

The network meta-analysis was conducted by Bayesian methods using WinBUGS version 1.4.3, with a vague normal prior for variance and a vague non-informative prior for the other parameters, and two chains with different initial values. Following an initial 20,000 iterations, convergence was checked and the burn-in period extended if necessary. If autocorrelation between runs was present, a thinning process was applied. Estimates were based on the results from 50,000 iterations. Both fixed effects and random effects models were analyzed, and the better model for each individual analysis was selected as being the one with the lower calculated deviance information criterion (DIC). Overall "goodness of fit" was assessed using the total residual deviance, which for a good fit should be close to the number of data points entered in the model.

Four of the five end points analyzed in Loveman et al. were dichotomous, or were converted to dichotomous, and analyzed as odds ratios, while change in FVC was analyzed as change from baseline. Loveman et al. also calculated a Bayesian network meta-analysis using WinBUGS version 1.4.3. They selected a vague normal prior for variance (normal distribution, mean 0, variance 10,000). They ran two chains with different initial values for a burn-in period of 20,000 iterations with a thinning interval of two. Estimates were based on results for 30,000 iterations. They assessed convergence using traces, Brooks-Gelman-Rubin, and density plots. They, too, estimated both fixed and random effects, and used DIC to select the better model.

Ofev CDR conducted sensitivity analyses that excluded trials on the basis of patient characteristics (primarily the three studies that were conducted in a Japanese population), route of administration, and potential bias in outcome reporting, and limited the analysis to phase 3 trials. Loveman et al. conducted sensitivity analyses around the measurement scales for FVC and controlling for baseline FVC via meta-regression.

Results

All-cause mortality

Both studies included an indirect comparison of all-cause mortality across eight studies each. Loveman et al. did not include mortality data from the two pirfenidone trials conducted in Japan (Taniguchi 2010 and Azuma 2005) but did include Raghu 2012.

Table 35 shows results of the analysis for all-cause mortality. In Ofev 2015, the lower DIC suggested that the fixed effects model provided a better fit, and the total residual deviance approximated the number of data points, suggesting a good fit. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant. Pirfenidone was not statistically different from placebo. There was no statistical heterogeneity according to the I^2.

In Loveman et al., as above, only the fixed effects model was reported, as having the lower DIC. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant. Pirfenidone showed a lower odds of death than placebo (odds ratio 0.50 [95% credible interval (CrI), 0.29 to 0.84]).

A sensitivity analysis by Ofev CDR excluded the two studies conducted in Japan that Loveman et al. had excluded from their primary analysis without substantially changing the results. However, the data sources and pooling of the pirfenidone registration trials differed between the two indirect comparisons. Ofev CDR pooled the data from CAPACITY 004 and CAPACITY 006 and used unpublished data, including 27 deaths in the pirfenidone group and 34 in the placebo. Loveman et al. did not pool the data and used only published data for these studies, totalling 11 deaths in the pirfenidone group and 22 in the placebo.

Additional sensitivity analyses in Ofev CDR removed studies due to patient characteristics, study phase, and potential bias in outcome reporting without affecting the overall results.

TABLE 35: RESULTS OF INDIRECT COMPARISONS OF ALL-CAUSE MORTALITY (ODDS RATIOS)

	Ofev CDR 2015		Loveman et al. 2015
	Fixed Effects OR (95% CrI) DIC 74.2, TRD 11.4	Random Effects OR (95% CrI) DIC 76.1, TRD 12.1	Fixed Effects OR (95% CrI)
Nintedanib versus placebo	**0.70 (0.45 to 1.10)**	0.70 (0.25 to 2.02)	**0.70 (0.45 to 1.09)**
Nintedanib versus pirfenidone	**1.00 (0.55 to 1.85)**	1.00 (0.23 to 3.59)	**1.39 (0.70 to 2.82)**
Nintedanib versus N-acetylcysteine (all)[a]	**0.32 (0.06 to 1.41)**	0.33 (0.03 to 2.79)	
Pirfenidone versus placebo	**0.70 (0.46 to 1.05)**	0.70 (0.32 to 1.87)	**0.50 (0.29 to 0.84)**
N-acetylcysteine versus placebo (all)[a]	**2.15 (0.53 to 11.02)**	2.15 (0.53 to 11.02)	
N-acetylcysteine versus placebo (oral)[b]			**2.19 (0.53 to 10.84)**
N-acetylcysteine versus placebo (triple)[b,c]			**15.09 (1.83 to 379.97)**

CrI = credible interval; DIC = deviance information criterion; OR = odds ratio; TRD = total residual deviance.
[a] Loveman et al. 2015 did not report this comparison.
[b] Ofev CDR 2015 did not report this comparison.
[c] Study terminated early for safety reasons.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Respiratory mortality

A separate indirect comparison for respiratory mortality was conducted by Loveman et al. across eight studies.

Table 36 shows results of the analysis for respiratory mortality. Loveman et al. did not comment on the fit of the model. Results were consistent with the analysis for all-cause mortality. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant. Pirfenidone showed a lower odds of death than placebo (odds ratio 0.30 [95% CrI, 0.12 to 0.68]).

TABLE 36: RESULTS OF INDIRECT COMPARISONS OF RESPIRATORY MORTALITY (ODDS RATIOS)

	Loveman et al. 2015
	Fixed Effects OR (95% CrI)
Nintedanib versus placebo	**0.62 (0.36 to 1.08)**
Nintedanib versus pirfenidone	**2.10 (0.77 to 6.17)**
Nintedanib versus N-acetylcysteine (all)[a]	
Pirfenidone versus placebo	**0.30 (0.12 to 0.68)**
N-acetylcysteine versus placebo (all)[a]	
N-acetylcysteine versus placebo (oral)	**1.79 (0.41 to 9.09)**
N-acetylcysteine versus placebo (triple)[b]	**13.04 (1.53 to 278.11)**

CrI = credible interval; OR = odds ratio.
[a] Loveman et al. 2015 did not report this comparison.
[b] Study terminated early for safety reasons.
Source: Loveman et al. 2015.[37]

Acute exacerbations

The Ofev CDR included an indirect comparison for acute exacerbations over seven studies, while Loveman et al. presented only direct comparisons due to uncertainty about the data. Data from Noble 2011 (CAPACITY) and Homma 2012 were not included in Loveman et al., while data from Raghu 2012 was not included in Ofev CDR. Data in Richeldi 2011 were converted from events per person-years to proportions.

Table 37 shows results of the analysis for acute exacerbations. The total residual deviance of the fixed effects model indicated a poor fit. According to the DIC, the random effects model was a better fit. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant, but the wide credible interval suggests uncertainty. Pirfenidone was not statistically different from placebo. The pairwise meta-analyses showed heterogeneity in the pirfenidone studies (I^2 = 60.0%), which was explored in sensitivity analyses.

Exclusion of the three Japanese studies (two pirfenidone, one N-acetylcysteine) produced a good fit for the fixed effects model. The odds ratio for nintedanib versus placebo for the fixed effects model was unchanged from the analysis including all studies (0.56 [95% CrI, 0.35 to 0.89]). Exclusion of Azuma 2005 alone and Azuma 2005 and Homma 2012 also improved the fit.

TABLE 37: RESULTS OF INDIRECT COMPARISONS OF ACUTE EXACERBATIONS OF IPF

	Ofev CDR 2015		Loveman et al. 2015
	Fixed Effects OR (95% CrI) DIC 78.0, TRD 21.8	Random Effects OR (95% CrI) DIC 75.5, TRD 14.6	Fixed Effects OR (95% CrI)
Nintedanib versus placebo	0.56 (0.35 to 0.89)	**0.47 (0.01 to 15.96)**	**0.50 (0.31 to 0.79)**
Nintedanib versus pirfenidone[a]	0.96 (0.36 to 2.58)	**1.22 (0.02 to 257)**	
Nintedanib versus N-acetylcysteine (all)[a]	1.04 (0.27 to 4.51)	**0.98 (0.01 to 189.3)**	
Pirfenidone versus placebo	0.59 (0.24 to 1.35)	**0.37 (0.01 to 4.81)**	**0.43 (0.14 to 1.26)**
N-acetylcysteine versus placebo (all)[a]	0.54 (0.13 to 1.91)	**0.47 (0.01 to 17.32)**	
N-acetylcysteine versus placebo (oral)[b]			**0.99 (0.17 to 6.00)**
N-acetylcysteine versus placebo (triple)[b]			**32.12 (1.47 to 9,036.71)**

CrI = credible interval; DIC = deviance information criterion; IPF = idiopathic pulmonary fibrosis; OR = odds ratio; TRD = total residual deviance.
[a] Loveman et al. 2015 did not report this comparison.
[b] Ofev CDR 2015 did not report this comparison.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Loss of lung function (change from baseline FVC)

Loveman et al. reported a network meta-analysis of change from baseline FVC, based on standardized data. Eleven studies contributed data. Ofev CDR did not report this end point.

Table 38 shows results of the analysis of for loss of lung function as measured in change of baseline FVC. The authors indicated that results for the random effects model were not statistically significant and that observed study heterogeneity validated the use of fixed effects models; otherwise they did not report the numerical results or their assessment of fit. The comparison between nintedanib versus pirfenidone (odds ratio 0.67 [95% CrI, 0.51 to 0.88]) and nintedanib versus placebo (odds ratio 0.41 [95% CrI, 0.34 to 0.51]) are both statistically significant. Pirfenidone versus placebo (odds ratio 0.62 [95% CrI, 0.54 to 0.74]) was also statistically significant. Similar results were obtained in sensitivity analyses when baseline FVC was controlled for and FVC was measured in litres.

TABLE 38: RESULTS OF INDIRECT COMPARISONS OF CHANGE FROM BASELINE FORCED VITAL CAPACITY

	Loveman et al. 2015
	Fixed Effects OR (95% CrI)
Nintedanib versus placebo	**0.41 (0.34 to 0.51)**
Nintedanib versus pirfenidone	**0.67 (0.51 to 0.88)**
Pirfenidone versus placebo	**0.62 (0.54 to 0.74)**
N-acetylcysteine versus placebo (oral)	**0.91 (0.59 to 1.41)**
N-acetylcysteine versus placebo (inhaled)	**0.65 (0.29 to 1.47)**
N-acetylcysteine versus placebo (triple)	**1.05 (0.59 to 1.85)**

CrI = credible interval; OR = odds ratio.
Source: Loveman et al. 2015.[37]

Loss in lung function (dichotomized)

Five studies reported a dichotomized end point around 10% change in FVC per cent predicted by the end of study follow-up. One study included death in its end point (King 2014) and the handling of death in another is unclear (Martinez 2014). Data from Homma 2012 and Raghu 2012 were not included in Ofev CDR, and data from Martinez 2014 were not included in Loveman et al.

Table 39 shows the results of the analysis of loss of lung function. In Ofev CDR, the total residual deviance approximated the number of data points, suggesting a good fit for both models, and the lower DIC suggested that the fixed effects model provided a better fit. Loveman et al. reported only the fixed effects analysis as giving a better fit by DIC. The fixed effects analyses of nintedanib versus placebo were statistically significant in both indirect comparisons: odds ratio 0.54 (95% CrI, 0.42 to 0.69) in Ofev CDR and odds ratio 0.61 (95% CrI, 0.48 to 0.78) in Loveman et al. There is no statistically significant difference between nintedanib and pirfenidone in either study or any analysis.

Similar results were produced by sensitivity analyses excluding King 2014 for heterogeneity, excluding King 2014 and Richeldi 2011 as potentially having different patient populations, and recalculating loss of lung function to include deaths with and without Noble 2011.

Loveman et al. excluded King 2014 on account of its comparatively larger effect size, and, since they did not have access to the patient-level data, they used statistical methods to adjust the numbers with a decline in FVC per cent predicted ≥ 10% to exclude deaths. The results were unchanged.

For the nintedanib studies, the numbers of patients meeting the criteria for loss of function differed between Ofev CDR and Loveman et al. In Richeldi 2011, the observed events in Ofev CDR were 25 and 34 in nintedanib and placebo, respectively, and in Loveman et al. the corresponding events were 20 and 37. For Richeldi 2014, the observed events in the Ofev CDR were 148 and 153 for nintedanib and placebo, respectively, and in Loveman et al. the corresponding events were 191 and 167. Ofev CDR used unpublished data while Loveman et al. used published data. The differences in point estimate between Ofev CDR and Loveman et al. are consistent with the differences in data.

TABLE 39: RESULTS OF INDIRECT COMPARISONS OF LOSS OF LUNG FUNCTION (FORCED VITAL CAPACITY PERCENTAGE DECREASED BY 10% OR GREATER)

	Ofev CDR 2015		Loveman et al. 2015
	Fixed Effects OR (95% CrI) DIC 74.7, TRD 10.3	Random Effects OR (95% CrI) DIC 75.2, TRD 9.3	Fixed Effects OR (95% CrI)
Nintedanib versus placebo	**0.54 (0.42 to 0.69)**	0.54 (0.11 to 2.70)	**0.61 (0.48 to 0.78)**
Nintedanib versus pirfenidone	**0.98 (0.67 to 1.42)**	0.99 (0.10 to 9.98)	**1.21 (0.86 to 1.72)**
Nintedanib versus N-acetylcysteine (all)[a]	**0.53 (0.29 to 0.96)**	0.53 (0.03 to 8.90)	
Pirfenidone versus placebo	**0.55 (0.41 to 0.72)**	0.54 (0.11 to 2.69)	**0.50 (0.39 to 0.65)**
N-acetylcysteine versus placebo (all)[a]	**1.02 (0.59 to 1.76)**	1.02 (0.10 to 9.73)	

CrI = credible interval; DIC = deviance information criterion; OR = odds ratio; TRD = total residual deviance.
[a] Loveman et al. 2015 did not report this comparison.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

Progression-free survival
For progression-free survival, only Ofev CDR reported pairwise comparisons based on two trials.

The estimated hazard ratio of nintedanib versus pirfenidone was 1.00 (95% CrI, 0.71 to 1.39).

Six-minute walk test
For the 6MWT, only Ofev CDR reported analyses.[26] Its network comprised three studies, one for each comparison, and therefore only direct comparisons were reported. Weighted mean difference change from baseline was statistically better for pirfenidone versus placebo (odds ratio 23.7 [95% CrI, 4.1 to 43.4]) on a fixed effects analysis. It was not statistically significantly different for nintedanib versus placebo (odds ratio 6.2 [95% CrI, –26.5 to 38.8]).

Withdrawals due to adverse events
Ofev CDR reported an indirect comparison of withdrawals due to adverse events that included seven studies.

Table 40 shows the results of the analysis for withdrawals due to adverse events. The total residual deviance of the fixed effects model indicated a poor fit, and according to the DIC, the fixed effects model fit better. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant for the random effects model. Pirfenidone was associated with withdrawal due to an adverse event compared with placebo (odds ratio 1.78 [95% CrI, 1.09 to 3.37]). There was no statistically significant difference between nintedanib and pirfenidone or between nintedanib and N-acetylcysteine, although comparisons with N-acetylcysteine have broad credible intervals, since only two withdrawals due to adverse events were included.

Exclusion of studies on the basis of baseline patient characteristics also produced a poor fit for the fixed effects model, but selecting for phase 3 data and excluding studies in Japanese populations improved the model fit. In all cases there was no significant change in the point estimates.

TABLE 40: RESULTS OF INDIRECT COMPARISONS OF WITHDRAWALS DUE TO ADVERSE EVENTS

	Ofev CDR 2015	
	Fixed Effects OR (95% CrI) DIC 89.9, TRD 11.4	Random Effects OR (95% CrI) DIC 91.7, TRD 11.8
Nintedanib versus placebo	**1.53 (1.13 to 2.08)**	1.50 (0.72 to 2.93)
Nintedanib versus pirfenidone	**0.88 (0.57 to 1.37)**	0.84 (0.31 to 1.84)
Nintedanib versus N-acetylcysteine (all)	**2.94 (0.36 to 21.86)**	2.42 (0.26 to 27.47)
Pirfenidone versus placebo	**1.73 (1.27 to 2.38)**	1.78 (1.09 to 3.37)
N-acetylcysteine versus placebo (all)	**0.61 (0.07 to 4.13)**	0.62 (0.06 to 4.98)

CrI = credible interval; DIC = deviance information criterion; OR = odds ratio; TRD = total residual deviance.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015.[26]

Serious cardiac adverse events
Ofev CDR reported an indirect comparison for serious cardiac events (including thromboembolic events and myocardial infarction) that included four studies, two for nintedanib and one for each of the comparators.

Table 41 shows the results of the analysis for serious cardiac adverse events. The total residual deviance of the fixed effects model indicated a poor fit. According to the DIC, the random effects model was a better fit, but the wide confidence intervals suggest high uncertainty. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant for the random effects model. Pirfenidone was not statistically different from placebo.

Exclusion of Richeldi 2011 and thus restricting the analysis to phase 3 studies improved the fit without significantly changing the results.

TABLE 41: RESULTS OF INDIRECT COMPARISONS OF SERIOUS CARDIAC ADVERSE EVENTS

	Ofev CDR 2015	
	Fixed Effects OR (95% CrI) DIC 50.3, TRD 10.7	Random Effects OR (95% CrI) DIC 48.1, TRD 7.6
Nintedanib versus placebo	0.76 (0.45 to 1.27)	**0.42 (0 to 21.16)**
Nintedanib versus pirfenidone	0.60 (0.26 to 1.39)	**0.34 (0 to 340.6)**
Nintedanib versus N-acetylcysteine (all)	0.14 (0.02 to 0.65)	**0.07 (0 to 75.98)**
Pirfenidone versus placebo	1.26 (0.65 to 2.49)	**1.26 (0 to 459.98)**
N-acetylcysteine versus placebo (all)	5.40 (1.27 to 41.00)	**5.64 (0.01 to 2,610.28)**

CrI = credible interval; DIC = deviance information criterion; OR = odds ratio; TRD = total residual deviance.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015.[26]

Serious gastrointestinal adverse events

Ofev CDR reported an indirect comparison for serious gastrointestinal adverse events (including liver enzyme and bilirubin elevations, gastrointestinal disorders, and gastrointestinal perforation) that included four studies, two for nintedanib and one for each of the comparators.

Table 42 shows the results of the analysis for serious gastrointestinal adverse events. The total residual deviance approximated the number of data points for both models, suggesting a good fit, and the lower DIC suggested that the fixed effects model provided a better fit. Neither the comparison of nintedanib versus pirfenidone nor the comparison of nintedanib versus placebo was statistically significant for the random effects model. Pirfenidone was not statistically different from placebo.

Similar results were produced by excluding Richeldi 2011 to restrict the analysis to phase 3 studies.

TABLE 42: RESULTS OF INDIRECT COMPARISONS OF SERIOUS GASTROINTESTINAL ADVERSE EVENTS

	Ofev CDR 2015	
	Fixed Effects OR (95% CrI) DIC 42.4, TRD 8.2	Random Effects OR (95% CrI) DIC 42.5, TRD 7.7
Nintedanib versus placebo	**2.25 (1.05 to 5.88)**	3.52 (0.08 to 429.92)
Nintedanib versus pirfenidone	**3.96 (1.18 to 14.51)**	5.94 (0.01 to 12570)
Nintedanib versus N-acetylcysteine (all)	**70.77 (4.39 to 32340)**	134.2 (0.01 to 1466000)
Pirfenidone versus placebo	**0.60 (0.23 to 1.45)**	0.59 (0 to 178.99)
N-acetylcysteine versus placebo (all)	**0.03 (0.00 to 0.46)**	0.03 (0 to 14.94)

CrI = credible interval; DIC = deviance information criterion; OR = odds ratio; TRD = total residual deviance.
Note: Best-fitting model in bold.
Source: Ofev CDR 2015.[26]

Critical appraisal

Ofev 2015 did not provide details for study selection. Ten of 11 identified studies were also found by Loveman et al., for which study selection was well described and well conducted.

The analyses for the indirect comparisons were appropriately conducted. Both indirect comparisons used Bayesian network meta-analyses with fixed and random effects models, using DIC to select for models and reporting odds ratios for dichotomous outcomes and change from baseline for continuous. Analyses were conducted using WinBUGS 1.4.3, and they both described standard methods for initialization, burn-in, and thinning. Ofev CDR reported estimates for both fixed and random effects, whereas Loveman et al. reported fixed effects and referenced random effects for confirmation. For well-fitting models, fixed effects gave equivalent or lower DIC, and for most outcomes, the estimates obtained were consistent between fixed and random effects models.

The network size ranged from 11 studies (decline in FVC, Loveman et al.) to three to four studies (safety end points). Neither indirect comparison excluded studies on the basis of quality, which is generally considered appropriate practice; Ofev CDR explored the effect of high risk of bias in sensitivity analyses. None of the studies involved a direct comparison between the active treatments, therefore all calculated comparisons were indirect, with the loss of power that entails. The model for studies of N-acetylcysteine differed between the two indirect comparisons: Loveman et al. used separate nodes for N-acetylcysteine given as oral, inhaled, or in triple therapy, while Ofev CDR combined the three into a single node. This may not have been appropriate, however results of the two analyses were very similar, and the findings of the PANTHER trial suggest that N-acetylcysteine is not different from placebo.

The small number of studies meant limited ability to detect statistical heterogeneity or assess the effects of removing studies during sensitivity analyses. There was clinical heterogeneity between studies in the form of variations in race or ethnicity of the patients, inclusion and exclusion criteria, mean disease duration before entry and baseline measurements in lung function, dosing, length of follow-up, definition of outcomes, and timing of end point assessment. These aspects are expanded on below. In addition, some analyses involved small numbers of events, especially for mortality, exacerbations, and safety end points, producing broad credible intervals and sensitivity to data sources.

Reviewers have commented elsewhere on the potential clinical impact of the differing inclusion criteria between the nintedanib and pirfenidone pivotal trials.[33] The two INPULSIS trials did not exclude people

with normal lung function, while the ASCEND trial (King 2014) imposed an upper limit on FVC. This resulted in a clinically meaningful difference in baseline FVC per cent predicted between the two trials, although the mean duration of disease was similar, suggesting that patients recruited to ASCEND may have had more advanced disease. This would violate the assumption of similarity of the control groups.

Only patients who received the licensed dose of nintedanib were included in the indirect comparisons. Two doses were allowed for pirfenidone (2,403 mg and 1,800 mg), the lower dose having been received almost exclusively by Japanese patients and rationalized on the basis of lower body weight. Results were unchanged when trials using the lower dose were excluded on the basis of patient characteristics.

Duration of follow-up varied from 32 weeks to 72 weeks: 52 weeks for nintedanib, 39 to 72 weeks for pirfenidone, and 32 to 60 weeks for N-acetylcysteine. Where end points were measured at multiple time points, the last available (usually end of follow-up) was used. Given the three-year to five-year survival after diagnosis, these differences in duration of observation were potentially significant for the accrual of events and changes in lung function required to detect clinically meaningful differences. No sensitivity analysis looked at the impact on duration of follow-up.

Ofev CDR used pre-specified sensitivity analyses, removing outlying studies or recalculating end points, to explore the effects of some of these variations. Loveman et al. used pre-specified meta-regression to explore influence of baseline FVC on decline FVC. However, the size of the network and the small number of events made the effect of sensitivity analyses difficult to detect.

The patients recruited into the studies reflected a mild to moderate IPF population, without pulmonary comorbidities, who were treated for up to 18 months. Thus, the effect on severe or advanced disease, on patients with other lung diseases, or with longer term treatment is unknown. Important comparators were included: Pirfenidone has been licensed for IPF, and N-acetylcysteine has been historically used. There is little consensus for the standard of care.[63] Both indirect comparison analyses included important outcomes of overall survival, exacerbations, and loss of lung function. Neither was able to include the patient-important outcomes of quality of life and health care utilization outcomes, since these were inconsistently reported or not reported across studies.

Discussion

The two indirect comparisons, Ofev CDR and Loveman et al., had an almost identical pool of studies (10 of 11 in common). They used the same analytic approach of a Bayesian framework with vague priors, calculating fixed and random effects, and using DIC for model selection. They differed in their handling of the dose variation in N-acetylcysteine: Ofev 2015 included all N-acetylcysteine trials in the same node of the network, while Loveman et al. separated them into three nodes. This is unlikely to have influenced the findings, as the results of the PANTHER trial suggest that N-acetylcysteine is no better than placebo.

The available studies did not include any head-to-head comparisons of nintedanib with its active comparators pirfenidone or N-acetylcysteine. Both found that the fixed effects models tended to fit the data and that statistical heterogeneity was in most cases low. Estimates of effects were similar, and discrepancies in results were potentially attributable to the use of different data sources: Ofev CDR used unpublished data for nintedanib from the manufacturer and data from regulatory documents, whereas Loveman et al. used published data.

Conclusion

Two indirect comparisons were reviewed, one supplied by the manufacturer (Ofev CDR) and one published (Loveman et al.). The two indirect comparisons used similar search and analysis methods and identified 11 studies each, with 10 studies in common, including pivotal trials of nintedanib and pirfenidone. The underlying network models were similar, with differences in the modelling for N-acetylcysteine.

Table 43 summarizes the comparison between nintedanib and pirfenidone in the form of the calculated odds ratios for the model with best fit (unless otherwise indicated, this was the fixed effects model).

Compared with pirfenidone, nintedanib decreased loss of lung function by one of the measures (rate of change from baseline FVC) but not the other (dichotomized around ≥ 10% change in FVC per cent predicted). Both nintedanib and pirfenidone were better than placebo. There was no difference between nintedanib and pirfenidone in mortality, respiratory mortality, and exacerbations. In Loveman et al., pirfenidone was statistically superior to placebo for all-cause and respiratory mortality. There was overall consistency of findings between the two indirect comparisons, with differences attributable to the use of unpublished data versus published data for mortality and exacerbations.

Compared with pirfenidone and placebo, nintedanib was associated with more serious gastrointestinal adverse events. Neither comparison with placebo showed a difference for serious adverse cardiac events, while pirfenidone was associated with more withdrawals due to adverse events of all types.

TABLE 43: SUMMARY OF COMPARISON FOR NINTEDANIB AND PIRFENIDONE FOR TREATMENT OF IDIOPATHIC PULMONARY FIBROSIS

	Odds ratio (95% CI)	
	Ofev CDR 2015	Loveman et al. 2015
All-cause mortality	1.00 (0.55 to 1.85)	1.39 (0.70 to 2.82)
Respiratory mortality[a]		2.10 (0.77 to 6.17)
Exacerbations[b]	1.22 (0.02 to 257)[c]	
Change from baseline FVC[a]		0.67 (0.51 to 0.88)
≥ 10% change FVC % predicted	0.98 (0.67 to 1.42)	1.21 (0.86 to 1.72)
Withdrawals due to adverse events[d]	0.84 (0.31 to 1.84)[c]	
Serious cardiac adverse events[d]	0.34 (0 to 340.6)[c]	
Serious gastrointestinal adverse events[d]	3.96 (1.18 to 14.51)	

CI = confidence interval; FVC = forced vital capacity.
[a] End point reported by Loveman et al. 2015 only.
[b] Loveman et al. 2015 did not report this comparison.
[c] Results from random effects analysis.
[d] Loveman et al. 2015 did not report this end point.
Source: Ofev CDR 2015,[26] Loveman et al. 2015.[37]

The evidence network was small and contained no direct comparisons between active treatments (i.e., the comparisons of nintedanib and pirfenidone in both analyses are indirect). The underlying assumption of similarity between the placebo groups may be invalid because of differences in study population (particularly in baseline FVC and potential disease severity) between the nintedanib and pirfenidone trials. Length of follow-up for accrual of events and variation in time points of assessments was another significant potential source of clinical heterogeneity. Although no statistical heterogeneity

within trials of individual treatments was detected, the established tests for heterogeneity tend to be insensitive, particularly for small numbers of trials. The data did not allow analysis of end points identified by patients such as quality of life or for health care utilization. The trials had been powered to detect changes in the surrogate end point of lung function, and for important end points such as all-cause mortality, disease exacerbations, and adverse events, the number of events was small and the estimates imprecise. This was particularly concerning for the safety end points of withdrawal due to adverse events and serious cardiac and gastrointestinal events, as it precluded detection of potentially clinically significant differences. Credible intervals suggested substantial uncertainty in the estimates, and the fragility of the findings was indicated by the effect of the use of published versus unpublished data for pirfenidone in the mortality analysis, where the results for pirfenidone versus placebo differed between the two indirect comparisons.

Overall, there remains a high degree of uncertainty in the comparative results for nintedanib versus pirfenidone.

REFERENCES

1. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011 Mar 15;183(6):788-824.
2. Wollin L, Wex E, Pautsch A, Schnapp G, Hostettler KE, Stowasser S, et al. Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. Eur Respir J [Internet]. 2015 May [cited 2015 May 22];45(5):1434-45. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4416110/pdf/ERJ-01749-2014.pdf
3. Spagnolo P, Sverzellati N, Rossi G, Cavazza A, Tzouvelekis A, Crestani B, et al. Idiopathic pulmonary fibrosis: an update. Ann Med. 2015 Feb;47(1):15-27.
4. Mazzei ME, Richeldi L, Collard HR. Nintedanib in the treatment of idiopathic pulmonary fibrosis. Ther Adv Respir Dis. 2015 Jun;9(3):121-9.
5. ^{Pr}Ofev^{TM} (nintedanib capsules): 100 mg and 150 mg nintedanib (as nintedanib esilate) [product monograph]. Burlington (ON): Boehringer Ingelheim (Canada) Ltd.; 2015 Jun 24.
6. Karimi-Shah BA, Chowdhury BA. Forced vital capacity in idiopathic pulmonary fibrosis--FDA review of pirfenidone and nintedanib. N Engl J Med. 2015 Mar 26;372(13):1189-91.
7. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med. 2011 Dec 15;184(12):1382-9.
8. Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, et al. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. Eur Respir J. 2010 Apr;35(4):830-6.
9. Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. Am J Respir Crit Care Med. 2003 Sep 1;168(5):543-8.
10. Collard HR, King TE, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2003 Sep 1;168(5):538-42.
11. Jones PW, Forde Y. St George's Respiratory Questionnaire manual [Internet]. Version 2.3. London (UK): St George's University of London; 2009 Jun. [cited 2015 Jul 7]. Available from: http://www.healthstatus.sgul.ac.uk/SGRQ_download/SGRQ%20Manual%20June%202009.pdf
12. Leidy NK, Wilcox TK, Jones PW, Murray L, Winnette R, Howard K, et al. Development of the EXAcerbations of Chronic Obstructive Pulmonary Disease Tool (EXACT): a patient-reported outcome (PRO) measure. Value Health. 2010 Dec;13(8):965-75.
13. Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, et al. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. Respir Med [Internet]. 2010 Feb [cited 2015 Jun 29];104(2):296-304. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2856332/pdf/nihms-188878.pdf
14. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014 May 29;370(22):2071-82.
15. Suissa S, Ernst P. The INPULSIS enigma: exacerbations in idiopathic pulmonary fibrosis. Thorax. 2015 May;70(5):508-10.

16. Raghu G, Collard HR, Anstrom KJ, Flaherty KR, Fleming TR, King TE Jr, et al. Idiopathic pulmonary fibrosis: clinically meaningful primary endpoints in phase 3 clinical trials. Am J Respir Crit Care Med. 2012 May 15;185(10):1044-8.

17. Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, et al. The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. Respir Med [Internet]. 2012 Oct [cited 2015 Jul 16];106(10):1447-55. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3441137

18. Clinical Study Report: 1199.32 [INPULSIS-1]. A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual forced vital capacity decline, in patients with idiopathic pulmonary fibrosis (IPF) [**CONFIDENTIAL** internal manufacturer's report]. [Ingelheim am Rhein]: Boehringer Ingelheim International GmbH; 2014 Apr 8.

19. Clinical Study Report: 1199.34 [INPULSIS-2]. A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120, 150 mg twice daily, on annual forced vital capacity decline, in patients with idiopathic pulmonary fibrosis (IPF) [**CONFIDENTIAL** internal manufacturer's report]. [Ingelheim am Rhein]: Boehringer Ingelheim International GmbH; 2014 Apr 8.

20. Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. Eur Respir Rev. 2012 Dec 1;21(126):355-61.

21. Raghu G, Chen SY, Yeh WS, Maroni B, Li Q, Lee YC, et al. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. Lancet Respir Med. 2014 Jul;2(7):566-72.

22. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, et al. An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. Am J Respir Crit Care Med. 2015 Jul 15;192(2):e3-e19.

23. Idiopathic Pulmonary Fibrosis Clinical Research Network, Raghu G, Anstrom KJ, King TE Jr, Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. N Engl J Med [Internet]. 2012 May 24 [cited 2015 Jul 3];366(21):1968-77. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3422642

24. [Pr]Esbriet® (pirfenidone capsules): 267 mg [product monograph]. Mississauga (ON): Hoffmann-La Roche Limited; 2014 Dec 10.

25. Common Drug Review. Pirfenidone resubmission (Esbriet - Hoffmann-La Roche Limited). Indication: idiopathic pulmonary fibrosis [Internet]. Ottawa: CADTH; 2015. [cited 2015 Jul 3]. (CDEC final recommendation). Available from: https://www.cadth.ca/sites/default/files/cdr/complete/cdr_compelte_SR0393_Esbriet_Apr-17-15.pdf

26. CDR submission: Ofev™ (nintedanib), 100 mg and 150 mg capsules. Company: Boehringer Ingelheim (Canada) Ltd. [**CONFIDENTIAL** manufacturer's submission]. Burlington (ON): Boehringer Ingelheim (Canada) Ltd.; 2015 Apr 23.

27. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Medical review(s) [Internet]. In: (Ofev) nintedanib capsules. Company: Boehringer Ingelheim Pharmaceuticals, Inc. Application No.: 205832 Orig1. Approval date: 10/15/2014. Rockville (MD): FDA; 2014 Oct 15 [cited 2015 May 5]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/205832Orig1s000TOC.cfm.

28. Center for Drug Evaluation and Research, U.S. Food and Drug Administration. Statistical review(s) [Internet]. In: (Ofev) nintedanib capsules. Company: Boehringer Ingelheim Pharmaceuticals, Inc. Application No.: 205832 Orig1. Approval date: 10/15/2014. Rockville (MD): FDA; 2015 Oct 15 [cited 2015 May 5]. (FDA drug approval package). Available from: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2014/205832Orig1s000TOC.cfm.

29. Richeldi L, Costabel U, Selman M, Kim DS, Hansell DM, Nicholson AG, et al. Efficacy of a tyrosine kinase inhibitor in idiopathic pulmonary fibrosis. N Engl J Med. 2011 Sep 22;365(12):1079-87.

30. Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. Chest. 1998 Mar;113(3):619-24.

31. Crawford B, Monz B, Hohlfeld J, Roche N, Rubin B, Magnussen H, et al. Development and validation of a cough and sputum assessment questionnaire. Respir Med. 2008 Nov;102(11):1545-55.

32. Ogura T, Taniguchi H, Azuma A, Inoue Y, Kondoh Y, Hasegawa Y, et al. Safety and pharmacokinetics of nintedanib and pirfenidone in idiopathic pulmonary fibrosis. Eur Respir J. 2015 May;45(5):1382-92.

33. Hunninghake GM. A new hope for idiopathic pulmonary fibrosis. N Engl J Med. 2014 May 29;370(22):2142-3.

34. Moodley Y, Corte T, Richeldi L, King TE Jr. Do all patients with idiopathic pulmonary fibrosis warrant a trial of therapeutic intervention? A pro-con perspective. Respirology [Internet]. 2015 Apr [cited 2015 May 22];20(3):389-94. Available from: http://onlinelibrary.wiley.com/doi/10.1111/resp.12484/epdf

35. Collard HR, Moore BB, Flaherty KR, Brown KK, Kaner RJ, King TE Jr, et al. Acute exacerbations of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med [Internet]. 2007 Oct 1 [cited 2015 Jul 3];176(7):636-43. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2094133

36. Common Drug Review. Pirfenidone (Esbriet - InterMune International AG). Indication: idiopathic pulmonary fibrosis [Internet]. Ottawa: CADTH; 2013. [cited 2015 Jul 3]. (CDEC final recommendation). Available from: https://www.cadth.ca/sites/default/files/cdr/complete/cdr_complete_Esbriet_April-24-13.pdf

37. Loveman E, Copley VR, Scott DA, Colquitt JL, Clegg AJ, O'Reilly KM. Comparing new treatments for idiopathic pulmonary fibrosis - a network meta-analysis. BMC Pulm Med. 2015 Apr 18;15(1):37.

38. Supplement to: Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med [Internet]. 2014 May 29 [cited 2015 Jul 3];370(22):2071-82. Available from: http://www.nejm.org/doi/suppl/10.1056/NEJMoa1402584/suppl_file/nejmoa1402584_appendix.pdf

39. Swigris JJ, Wamboldt FS, Behr J, du Bois RM, King TE, Raghu G, et al. The 6 minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. Thorax. 2010 Feb;65(2):173-7.

40. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med. 2011 May 1;183(9):1231-7.

41. Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. Respir Med. 2009 Oct;103(10):1430-5.

42. Swigris JJ, Esser D, Conoscenti CS, Brown KK. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. Health Qual Life Outcomes [Internet]. 2014 [cited 2015 Jun 29];12:124. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4148554/pdf/12955_2014_Article_124.pdf

43. Richeldi L, Ryerson CJ, Lee JS, Wolters PJ, Koth LL, Ley B, et al. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. Thorax [Internet]. 2012 May [cited 2015 Jun 29];67(5):407-11. Available from: http://thorax.bmj.com/content/67/5/407.full.pdf+html

44. Patel AS, Siegert RJ, Keir GJ, Bajwah S, Barker RD, Maher TM, et al. The minimal important difference of the King's Brief Interstitial Lung Disease Questionnaire (K-BILD) and forced vital capacity in interstitial lung disease. Respir Med. 2013 Sep;107(9):1438-43.

45. Eaton T, Young P, Milne D, Wells AU. Six-minute walk, maximal exercise tests: reproducibility in fibrotic interstitial pneumonia. Am J Respir Crit Care Med [Internet]. 2005 May 15 [cited 2015 Jun 29];171(10):1150-7. Available from: http://ajrccm.atsjournals.org/content/171/10/1150.full.pdf+html

46. Lederer DJ, Arcasoy SM, Wilt JS, D'Ovidio F, Sonett JR, Kawut SM. Six-minute-walk distance predicts waiting list survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med [Internet]. 2006 [cited 2015 Jun 29];174(6):659-64. Available from: http://ajrccm.atsjournals.org/content/174/6/659.full.pdf+html

47. du Bois RM, Albera C, Bradford WZ, Costabel U, Leff JA, Noble PW, et al. 6-Minute walk distance is an independent predictor of mortality in patients with idiopathic pulmonary fibrosis. Eur Respir J. 2014 May;43(5):1421-9.

48. Garin MC, Highland KB, Silver RM, Strange C. Limitations to the 6-minute walk test in interstitial lung disease and pulmonary hypertension in scleroderma. J Rheumatol. 2009 Feb;36(2):330-6.

49. Belkin A, Swigris JJ. Health-related quality of life in idiopathic pulmonary fibrosis: where are we now? Curr Opin Pulm Med. 2013;19(5):474-9.

50. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis. 1992 Jun;145(6):1321-7.

51. Bourbeau J, Maltais F, Rouleau M, Guimont C. French-Canadian version of the Chronic Respiratory and St George's Respiratory questionnaires: an assessment of their psychometric properties in patients with chronic obstructive pulmonary disease. Can Respir J. 2004 Oct;11(7):480-6.

52. Jones PW. St. George's Respiratory Questionnaire original English version [Internet]. London: St. George's University of London; 2003. [cited 2015 Jul 7]. Available from: http://www.healthstatus.sgul.ac.uk/SGRQ_download/Original%20English%20version.pdf

53. Nishiyama O, Taniguchi H, Kondoh Y, Kimura T, Kataoka K, Nishimura K, et al. Health-related quality of life does not predict mortality in idiopathic pulmonary fibrosis. Sarcoidosis Vasc Diffuse Lung Dis. 2012 Oct;29(2):113-8.

54. Maly M, Vondra V. Generic versus disease-specific instruments in quality-of-life assessment of chronic obstructive pulmonary disease. Methods Inf Med. 2006;45(2):211-5.

55. Ong KC, Lu SJ, Soh CS. Does the multidimensional grading system (BODE) correspond to differences in health status of patients with COPD? Int J Chron Obstruct Pulmon Dis [Internet]. 2006 [cited 2015 Jul 7];1(1):91-6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2706601

56. Meguro M, Barley EA, Spencer S, Jones PW. Development and validation of an improved, COPD-specific version of the St. George Respiratory Questionnaire. Chest [Internet]. 2007 Aug [cited 2015 Jul 7];132(2):456-63. Available from: http://journal.publications.chestnet.org/data/Journals/CHEST/22059/456.pdf

57. Shukla V, Husereau D, Boucher M, Mensinkai S, Dales R. Long-acting b_2-agonists for maintenance therapy of stable chronic obstructive pulmonary disease: a systematic review [Internet]. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 2002. [cited 2015 Jul 7]. (Technology report no.27). Available from: http://www.cadth.ca/media/pdf/124_laba_tr_e.pdf

58. Peng S, Li Z, Kang J, Hou X. Cross-sectional and longitudinal construct validity of the Saint George's Respiratory Questionnaire in patients with IPF. Respirology. 2008 Nov;13(6):871-9.

59. Loveman E, Copley VR, Colquitt J, Scott DA, Clegg A, Jones J, et al. The clinical effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation. Health Technol Assess [Internet]. 2015 Mar [cited 2015 May 21];19(20):i-xxiv. Available from: http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0011/139547/FullReport-hta19200.pdf

60. Loveman E, Copley VR, Colquitt JL, Scott DA, Clegg AJ, Jones J, et al. The effectiveness and cost-effectiveness of treatments for idiopathic pulmonary fibrosis: systematic review, network meta-analysis and health economic evaluation. BMC Pharmacol Toxicol [Internet]. 2014 [cited 2015 May 21];15:63. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4247619/pdf/40360_2014_Article_347.pdf

61. Loveman E, Colquitt J, Copley V, Scott D, Clegg A, Richeldi L, et al. Treatments for idiopathic pulmonary fibrosis: an updated systematic review and network meta-analysis [protocol]. PROSPERO [Internet]. 2014 [cited 2015 Jun 17];CRD42014009688. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42014009688#.VYK6cmfbKvs

62. Loveman E, Colquitt J, Jones J, Clegg A, Bausewein C, Singh S, et al. Treatments for idiopathic pulmonary fibrosis: a systematic review and economic evaluation [protocol]. PROSPERO [Internet]. 2012 [cited 2015 Jun 17];CRD42012002116. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42012002116#.VYK5R2fbKvs

63. Covvey JR, Mancl EE. Recent evidence for pharmacological treatment of idiopathic pulmonary fibrosis. Ann Pharmacother. 2014 Dec;48(12):1611-9.