

Nutzenbewertung



Gemeinsamer
Bundesausschuss

**von Arzneimitteln mit neuen Wirkstoffen nach
§ 35a SGB V**

**Bewertung von Arzneimitteln für seltene Leiden
nach § 35a Absatz 1 Satz 10 i.V.m.
5. Kapitel § 12 Nr. 1 Satz 2 Verfo**

Wirkstoff: Nintedanib

Datum der Veröffentlichung: 15. Juni 2015

Inhaltverzeichnis

Inhaltverzeichnis.....	2
Tabellenverzeichnis.....	3
Abbildungsverzeichnis.....	5
Abkürzungsverzeichnis.....	6
Hintergrund	8
1. Einführung	10
2. Nutzenbewertung.....	12
2.1 Fragestellung	12
2.2 Zulassungsbegründende Studien.....	12
2.3 Liste der verwendeten Quellen.....	13
2.4 Studiencharakteristika.....	14
2.5 Endpunkte und Erhebungsinstrumente.....	23
2.5.1 In der Nutzenbewertung berücksichtige Endpunkte	23
2.5.2 Verzerrungspotential auf Studien- und Endpunktebene	24
2.5.3 Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz	25
2.5.4 Statistische Methoden.....	38
2.6 Ergebnisse zum Zusatznutzen	41
2.7 Einschätzung zum Ausmaß des Zusatznutzens durch den pU.....	66
3. Methodische Anmerkungen und Bewertung der Unterlagen	68
3.1 Zulassungsstatus und Zulassungspopulation von Nintedanib.....	68
3.2 Design und Methodik der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)	68
3.3 Wirksamkeit.....	69
3.3.1 Zusammenfassende Einschätzung zur Wirksamkeit	71
3.4 Lebensqualität.....	72
3.4.1 Zusammenfassende Einschätzung zur Lebensqualität	73
3.5 Sicherheit.....	74
3.5.1 Zusammenfassende Einschätzung zur Sicherheit.....	75
4. Anforderungen an eine qualitätsgesicherte Anwendung	76
5. Zusammenfassung der Bewertung	77
Referenzen.....	79

Tabellenverzeichnis

Tabelle 1: Charakterisierung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)	14
Tabelle 2: Charakterisierung der Intervention in den Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34).....	18
Tabelle 3: Charakterisierung der Studienpopulation in INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) zum Baseline.....	21
Tabelle 4: Vom pU im Dossier als patientenrelevant bezeichnete und in der Nutzenbewertung berücksichtigte Endpunkte der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ...	23
Tabelle 5: Verzerrungspotential der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) auf Studienebene	24
Tabelle 6: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte in den Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34).....	24
Tabelle 7: Beschreibung der Endpunkte der Kategorie Mortalität	26
Tabelle 8: Beschreibung der Endpunkte der Kategorie Morbidität	29
Tabelle 9: Beschreibung der Endpunkte der Kategorie Lebensqualität.....	32
Tabelle 10: Beschreibung der Endpunkte der Kategorie Sicherheit.....	36
Tabelle 11: Allgemeine Angaben in den Studien INPULSIS-1 und INPULSIS-2.....	41
Tabelle 12: Ergebnisse zur Mortalität – Zeit bis zum Tod	43
Tabelle 13: jährliche Abnahme der FVC	44
Tabelle 14: FVC-Responderanalyse.....	45
Tabelle 15: (adjudizierte) akute Exazerbation.....	46
Tabelle 16: PGI-C-Responderanalyse	48
Tabelle 17: Veränderung der EQ-5D-VAS.....	48
Tabelle 18: Ergebnisse zur Lebensqualität – Veränderung von Baseline bis Woche 52	49
Tabelle 19: Ergebnisse zur Lebensqualität – SGRQ-Responder und SGRQ-I-Responder	52
Tabelle 20: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase	53
Tabelle 21: UE mit Inzidenz ≥ 10 % in den Studien INPULSIS-1 oder INPULSIS-2.....	56

Tabelle 22: Subgruppenanalyse FVC-Responder.....	59
Tabelle 23: Subgruppenanalyse Zeit bis zur ersten akuten Exazerbation.....	60
Tabelle 24: Subgruppenanalyse SGRQ-(\leq -4 Punkte)-Responder.....	61
Tabelle 25: Subgruppenanalyse Veränderung SGRQ-I-Gesamtscore.....	62
Tabelle 26: Subgruppenanalyse SGRQ-I-(\leq -4 Punkte)-Responder.....	62
Tabelle 27: Subgruppenanalyse unerwünschte Ereignisse.....	64
Tabelle 28: Subgruppenanalyse unerwünschte Ereignisse von besonderem Interesse.....	65
Tabelle 1: Zusammenfassende Darstellung der Studienergebnisse	77

Abbildungsverzeichnis

Abbildung 1: Mittlere beobachtete FVC-Veränderung (standard error of the mean, SEM) gegenüber dem Ausgangswert (ml) im Zeitverlauf, Studien INPULSIS-1 und INPULSIS-2 gepoolt	45
Abbildung 3: Kaplan-Meier-Kurven für Endpunkt Zeit bis zur ersten akuten Exazerbation für die gepoolte Analyse der INPULSIS Studien	47

Abkürzungsverzeichnis

ALAT	Latin American Thoracic Society
ALT	Alanin-Aminotransferase
AST	Aspartat-Aminotransferase
ATS	American Thoracic Society
bid	bis in die (zweimal täglich)
CASA-Q	Cough and Sputum Assessment Questionnaire
CHMP	Committee for Medicinal Products for Human Use
CRF	Case Report Form
CTCAE	Common Terminology Criteria for Adverse Events
DL _{CO}	Diffusing Capacity of the Lung for Carbon Monoxide (Kohlenmonoxid-Diffusionskapazität)
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EQ-5D	EuroQoL 5-Dimensional Quality of Life Questionnaire
ERS	European Respiratory Society
FEV ₁	Forced Expiratory Volume in one Second (forciertes exploratorisches Volumen in 1 Sekunde)
FVC	Forced Vital Capacity (Forcierte Vitalkapazität)
G-BA	Gemeinsamer Bundesausschuss
GFR	Glomeruläre Filtrationsrate
GKV	Gesetzliche Krankenversicherung
HRCT	High Resolution Computed Tomography (Hochauflösende Computer-Tomographie)
HRQoL	Health-Related Quality of Life (Gesundheitsbezogene Lebensqualität)
IAC	Independent Adjudication Committee (Unabhängiges Bewertungskomitee)
IIP	idiopathische interstitielle Pneumonien
INR	International Normalized Ratio (internationale normalisierte Ratio)
IPF	Idiopathic Pulmonary Fibrosis (Idiopathische Lungenfibrose)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
JRS	Japanese Respiratory Society
KI	Konfidenzintervall

MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligramm
MID	Minimal Important Difference
MMRM	Mixed-Effects-Modell für wiederholte Messungen
MW	Mittelwert
N	Anzahl
PGI-C	Patient's Global Impression of Change
PT	Preferred Term
PTT	Partial Thromboplastin Time (Partielle Thromboplastinzeit)
pU	Pharmazeutischer Unternehmer
qd	omne in die (einmal täglich)
SD	Standardabweichung
SGB	Sozialgesetzbuch
SGRQ	St. George's Respiratory Questionnaire
SGRQ-I	St. George's Respiratory Questionnaire IPF-specific version
SOBQ	Shortness of Breath Questionnaire
SOC	Systemorganklasse
SpO ₂	Blutsauerstoffsättigung
SUE	Schwerwiegende/s unerwünschte/s Ereignis/se
UE	Unerwünschte/s Ereignis/se
ULN	Upper Limits of Normal (Oberer Normgrenzwert)
VAS	Visual Analogue Scale
VerfO	Verfahrensordnung
ZN	Zusatznutzen
ZNS	Zentralnervensystem

Hintergrund

Nintedanib ist zugelassen als Arzneimittel zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. Gemäß § 35a Absatz 1 Satz 10 SGB V gilt der medizinische Zusatznutzen durch die Zulassung als belegt.

Gemäß § 35a Absatz 2 SGB V entscheidet der Gemeinsame Bundesausschuss (G-BA), ob er die Nutzenbewertung selbst durchführt oder das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) beauftragt. Ausgehend von der gesetzlichen Vorgabe in § 35a Absatz 1 Satz 10 SGB V, dass der Zusatznutzen eines Orphan Drug durch die Zulassung als belegt gilt, hat der G-BA in seiner Sitzung vom 15. März 2012 das Verfahren der Nutzenbewertung von Orphan Drugs dahingehend modifiziert, dass bei Orphan Drugs zunächst keine eigenständige Festlegung einer zweckmäßigen Vergleichstherapie mehr durch den G-BA als Grundlage der insoweit allein rechtlich zulässigen Bewertung des Ausmaßes eines gesetzlich zu unterstellenden Zusatznutzens erfolgt. Vielmehr wird ausschließlich auf der Grundlage der Zulassungsstudien das Ausmaß des Zusatznutzens durch den G-BA bewertet.

Dementsprechend hat der G-BA in seiner Sitzung am 15. März 2012 den mit Beschluss vom 1. August 2011 erteilten Auftrag an das IQWiG zur Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen gemäß § 35a Absatz 2 SGB V in der Weise abgeändert, dass bei Orphan Drugs eine Beauftragung des IQWiG mit der Durchführung einer Nutzenbewertung bei zuvor festgelegter Vergleichstherapie erst dann erfolgt, wenn der Umsatz des betreffenden Arzneimittels die gesetzliche Grenze von 50 Millionen Euro überschritten hat und damit einer uneingeschränkten Nutzenbewertung unterliegt (vgl. § 35a Absatz 1 Satz 11 SGB V).

Der G-BA bestimmt gemäß 5. Kapitel § 12 Absatz 1 Nummer 1 Satz 2 der Verfahrensordnung des G-BA (VerfO) das Ausmaß des Zusatznutzens für die Anzahl der Patienten und Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Nutzenbewertung). Zur Bestimmung des Ausmaßes des Zusatznutzens von Nintedanib zieht der G-BA die Zulassung und Unterlagen zur Zulassung (insbesondere den EPAR), die Studien, auf deren Grundlage die Zulassung des Arzneimittels beruht, sowie das Dossier des pharmazeutischen Unternehmers heran und bewertet die darin enthaltenen Daten nach Maßgabe der im 5. Kapitel § 5 Absatz 7 Nr. 1 bis 4 VerfO festgelegten Kriterien im Hinblick auf ihre therapeutische Relevanz.

Der Unterausschuss Arzneimittel hat die Nutzenbewertung zum Wirkstoff Nintedanib in seiner Sitzung am 9. Juni 2015 zur Kenntnis genommen. Darüber hinaus wurde mit Schreiben vom 16. März 2015 das IQWiG beauftragt, ergänzend eine Bewertung der Angaben des pharmazeutischen Unternehmers in Modul 3 zu folgenden Gesichtspunkten durchzuführen:

- Kosten der Therapie für die gesetzliche Krankenversicherung
- Anzahl der GKV-Patienten in der Zielpopulation

Die Nutzenbewertung wird am 15. Juni 2015 zusammen mit der Bewertung der Therapiekosten und Patientenzahlen des IQWiG auf der Internetseite des G-BA (<http://www.g-ba.de>) veröffentlicht und damit das schriftliche Stellungnahmeverfahren eingeleitet. Es wird darüber hinaus eine mündliche Anhörung durchgeführt. Über das Ausmaß des Zusatznutzens beschließt der G-BA innerhalb von drei Monaten nach Veröffentlichung der Nutzenbewertung.

1. Einführung

idiopathische Lungenfibrose (IPF)

Die idiopathische Lungenfibrose (IPF) ist eine chronische, progredient verlaufende, fibrosierende interstitielle Pneumonie unklarer Genese, die zu den interstitiellen Lungenerkrankungen (ILD) gezählt wird und primär bei älteren Erwachsenen auftritt [4,36,34].

Pathologisches Hauptcharakteristikum einer IPF ist eine Fibrosierung des Lungeninterstitiums, d. h. eine vermehrte Bildung des Stütz- oder Bindegewebes zwischen den Lungenbläschen (Alveolen). Als Folge einer solchen Fibrosierung kommt es zu einem Verlust an gesundem Lungengewebe. Diese pathologischen Vorgänge reduzieren die Dehnbarkeit der Lunge und begünstigen Störungen beim Gasaustausch, was schließlich mit fortschreitendem Funktionsverlust zu einer Schrumpfung der Lunge führt. Klinisch manifestiert sich eine IPF daher auch in einer zunehmend eintretenden Luftnot [4,27].

Die IPF wird in der Literatur als die häufigste oder zweithäufigste Form innerhalb des Formenkreises der idiopathischen interstitiellen Pneumonien (IIP) beschrieben, mit einer breiten Spanne von 17 % bis 86 % der IIP-Patienten [33]. Die Inzidenz nimmt mit höherem Alter zu, wobei sich die Erkrankung typischerweise in der 6. und 7. Lebensdekade manifestiert. Das Durchschnittsalter bei Diagnosestellung ca. 66 Jahre beträgt. Familiäre Formen der IPF wurden berichtet, betreffen aber weniger als 5% aller IPF-Patienten. Insgesamt erkranken Männer häufiger als Frauen. Obgleich es sich bei der idiopathischen Lungenfibrose definitionsgemäß um eine Erkrankung unbekannter Ätiologie handelt, wurden einige mögliche Risikofaktoren wie Rauchen und die Exposition zu Metall- und Holzstäuben beschrieben [4,29].

Die Prognose der IPF ist ungünstig, so beträgt das mediane Überleben nach Diagnosestellung zwei bis vier Jahre [3].

Nintedanib

Nintedanib wurde, aufgrund seiner antiangiogenen Wirkung, initial in onkologischen Indikationen untersucht und ist inzwischen in Kombination mit Docetaxel zur Behandlung von erwachsenen Patienten mit lokal fortgeschrittenem, metastasiertem oder lokal rezidiertem nicht-kleinzelligem Lungenkarzinom (NSCLC) mit Adenokarzinom - Histologie nach Erstlinien-Chemotherapie – zugelassen.

Der Wirkstoff gehört zur ATC-Klasse der Proteinkinase-Inhibitoren. Als niedermolekularer Tyrosinkinaseinhibitor, hemmt Nintedanib insbesondere folgende Rezeptoren: von Blutplättchen abgeleiteter Wachstumsfaktor-Rezeptor (platelet-derived growth factor receptor, PDGFR) α und β , Fibroblasten-Wachstumsfaktor-Rezeptor (fibroblast growth factor receptor, FGFR) 1 – 3 und VEGFR 1 – 3. Nintedanib bindet kompetitiv an die Adenosintriphosphat (ATP)-Bindungstasche dieser Rezeptoren und blockiert so die intrazelluläre Signalübertragung. Darüber hinaus hemmt Nintedanib die Kinasen Flt-3 (Fmslike tyrosine-

protein kinase), Lck (lymphocyte- specific tyrosine-protein kinase), Lyn (tyrosine-protein kinase lyn) und Src (protooncogenetyrosine-protein kinase src).

Nintedanib hemmt die Aktivierung der FGFR- und PDGFR-Signalwege, die eine entscheidende Rolle bei der Proliferation, Migration und Differenzierung von Lungenfibroblasten / Myofibroblasten spielen, welche die Pathologie der idiopathischen Lungenfibrose kennzeichnen. Die möglichen Auswirkungen der VEGFR-Hemmung durch Nintedanib und die antiangiogene Aktivität von Nintedanib auf die IPF-Pathologie sind derzeit nicht vollständig untersucht. In präklinischen Krankheitsmodellen der Lungenfibrose entfaltet Nintedanib eine starke antifibrotische und antiinflammatorische Aktivität. Nintedanib hemmt bei humanen Lungenfibroblasten von IPF-Patienten die Proliferation, Migration und Transformation zu Myofibroblasten.

Nintedanib ist zugelassen zur Behandlung der idiopathischen Lungenfibrose (IPF) bei Erwachsenen. Die empfohlene Dosis beträgt 150 mg Nintedanib zweimal täglich, die im Abstand von ca. 12 Stunden eingenommen wird. Die empfohlene maximale Tagesdosis von 300 mg sollte nicht überschritten werden. Die Anwendung der Dosis von 100 mg zweimal täglich wird nur bei Patienten empfohlen, die die Dosis von 150 mg zweimal täglich nicht tolerieren. Falls ein Patient 100 mg zweimal täglich nicht verträgt, sollte die Behandlung mit Nintedanib abgebrochen werden.

Die Kapseln sollten zu einer Mahlzeit eingenommen werden; sie sollten im Ganzen mit Wasser geschluckt und nicht zerkaut oder zerkleinert werden.

2. Nutzenbewertung

2.1 Fragestellung

Nintedanib (Ofev®) ist ein Arzneimittel mit dem Status „Orphan Drug“. Gemäß § 35a Abs. 1 Satz 10 SGB V gilt für Orphan Drugs der medizinische Zusatznutzen bereits durch die Zulassung als belegt. Der G-BA bestimmt bei Orphan Drugs, die einen Umsatz von 50 Millionen Euro in den letzten zwölf Kalendermonaten nicht übersteigen, das Ausmaß des Zusatznutzens auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. In die Bewertung werden gemäß Zulassungsstatus Patienten in folgender Indikation eingeschlossen:

- Nintedanib wird angewendet bei Erwachsenen zur Behandlung der idiopathischen Lungenfibrose (IPF).

2.2 Zulassungsbegründende Studien

Pivotale Studie

- INPULSIS-1 (1199.32): *“A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120 (= Nintedanib), 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF).”* Es handelt sich um eine 52-wöchige randomisierte, placebokontrollierte, doppelblinde Studie der Phase III mit dem Ziel der Untersuchung der Wirksamkeit sowie des Nutzen-Risiko-Verhältnisses von oral verabreichtem Nintedanib 150 mg bid im Vergleich zu Placebo bei 513 Patienten mit IPF. Als primärer Endpunkt wurde die Verringerung der Lungenfunktionsabnahme, gemessen als jährliche Abnahme der FVC, ermittelt. Patienten die an der Studie 1199.32 teilgenommen hatten, konnten nach Abschluss dieser Studie an einer open-label Extensionsstudie teilnehmen (Studie 1199.33).
- INPULSIS-2 (1199.34): *“A 52 weeks, double blind, randomized, placebo-controlled trial evaluating the effect of oral BIBF 1120 (= Nintedanib), 150 mg twice daily, on annual Forced Vital Capacity decline, in patients with Idiopathic Pulmonary Fibrosis (IPF).”* Es handelt sich um eine 52-wöchige randomisierte, placebokontrollierte, doppelblinde Studie der Phase III mit dem Ziel der Untersuchung der Wirksamkeit sowie des Nutzen-Risiko-Verhältnisses von oral verabreichtem Nintedanib 150 mg bid im Vergleich zu Placebo bei 548 Patienten mit IPF. Als primärer Endpunkt wurde die Verringerung der Lungenfunktionsabnahme, gemessen als jährliche Abnahme der FVC, ermittelt. Patienten die an der Studie 1199.34 teilgenommen hatten, konnten nach Abschluss dieser Studie an einer open-label Extensionsstudie teilnehmen (Studie 1199.33).

Weitere Studien

- 1199.31: *“A double-blind, randomised, placebo-controlled (within a dose group) study to evaluate safety and pharmacokinetics of multiple rising doses of BIBF 1120 at 50 mg*

Dossierbewertung für Orphan Drugs

Nintedanib

b.i.d. (14 days), 100 mg b.i.d. (14 days), and 150 mg b.i.d. (28 days) po, on top of standard medical care with stratification according to pirfenidone use, in Japanese patients with idiopathic pulmonary fibrosis. Es handelt sich um eine 2- bis 4-wöchige randomisierte, placebokontrollierte, doppelblinde Studie der Phase IIa zur Untersuchung der Sicherheit und Pharmakokinetik verschiedener Dosierungen von oral verabreichtem Nintedanib in unterschiedlichen Dosierungen von 50 mg zweimal täglich (2 Wochen), 100 mg zweimal täglich (2 Wochen) und 150 mg zweimal täglich (1 Woche) bei 50 japanischen Patienten mit IPF. Die Patienten wurden nach begleitender Pirfenidon-Therapie stratifiziert. Patienten die an der Studie 1199.31 teilgenommen hatten, konnten nach Abschluss dieser Studie an einer open-label Extensionsstudie teilnehmen (Studie 1199.40).

- TOMORROW (1199.30): *“A 52 week, double blind, randomized, placebo-controlled trial evaluating the effect of BIBF 1120 (= Nintedanib), administered at oral doses of 50 mg qd, 50 mg bid, 100 mg bid and 150 mg bid on Forced Vital Capacity decline during one year, in patients with Idiopathic Pulmonary Fibrosis, with optional active treatment extension until last patient out”.* Es handelt sich um eine 52-wöchige randomisierte, placebokontrollierte, doppelblinde 5-armige Studie der Phase II, mit dem Ziel der Untersuchung der Wirksamkeit und Sicherheit von oral verabreichtem Nintedanib in unterschiedlichen Dosierungen von 50 mg einmal täglich, 50 mg zweimal täglich, 100 mg zweimal täglich und 150 mg zweimal täglich bei 428 Patienten mit IPF (Dosisfindungsstudie). Der primäre Wirksamkeitsendpunkt war die Verringerung der Lungenfunktionsabnahme innerhalb eines Jahres. Patienten die an der Studie 1199.30 teilgenommen hatten, konnten nach Abschluss dieser Studie an einer open-label Extensionsstudie teilnehmen (Studie 1199.35).
- 1199.187: *“A six month double blind randomized placebo controlled trial followed by each arm being converted to oral nintedanib 150 mg twice daily comparing the effect on high resolution computerized tomography quantitative lung fibrosis score, lung function, Six Minute Walk Test distance and St. George’s Respiratory Questionnaire after six months of treatment in patients with idiopathic pulmonary fibrosis with continued evaluations over a period of up to eighteen months.”* Es handelt sich um eine randomisierte, placebokontrollierte, doppelblinde Studie der Phase III mit dem Ziel der Untersuchung des Effekts von Nintedanib 150 mg bid im Vergleich zu Placebo bei Patienten mit IPF. Als primärer Endpunkt wird die Verringerung auf dem High Resolution Computerized Tomography Quantitative Lung Fibrosis Score im Vergleich zum Baseline gemessen. Die Studie ist noch nicht abgeschlossen, die geplante Anzahl an Patienten ist 275.

2.3 Liste der verwendeten Quellen

Zur Bestimmung des Ausmaßes des Zusatznutzens für Nintedanib wurden folgende Studien und Daten herangezogen:

- Herstellerdossier zu Nintedanib in der Indikation IPF
- Unterlagen der Zulassungsbehörde, insbesondere Committee for Medicinal Products for Human Use Assessment Report (CHMP, 2014) [15]
- Studienbericht inkl. Studienprotokoll und statistischem Analyseplan der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Die Dosisfindungsstudie TOMORROW (1199.30) weist ein ähnliches Design auf wie die INPULSIS Studien. Anders als in den INPULSIS-Studien und anders als in der Fachinformation zu Nintedanib festgelegt, war in der TOMORROW-Studie nach Dosisreduktion keine erneute Reeskalation möglich, weshalb TOMORROW nicht im Dossier und nicht in der vorliegenden Nutzenbewertung zur Bestimmung des Zusatznutzens von Nintedanib herangezogen wurden. Wesentliche Ergebnisse der TOMORROW-Studie werden in der Diskussion dieser Nutzenbewertung in Kapitel 3.3.1 beschrieben.

2.4 Studiencharakteristika

Die Unterlagen zum Nachweis des Ausmaßes des Zusatznutzens für Nintedanib basieren auf der Zulassungsstudie INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34). Die Studien, Interventionen und Baselinecharakteristika werden in den Tabellen 1 bis 3 charakterisiert.

Tabelle 1: Charakterisierung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Charakteristikum	Beschreibung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ¹
Design	Multizentrische, randomisierte, placebokontrollierte, doppelblinde, klinische Studie der Phase III im Parallelgruppendedesign (Verhältnis 3:2 Nintedanib 150 mg bid versus Placebo) zur Untersuchung der Wirksamkeit sowie des Nutzen/Risiko-Verhältnisses von Nintedanib 150 mg bid im Vergleich zu Placebo bei IPF. Visite 1: Screening; Visite 2: Therapiestart und nachfolgende Behandlungsphase über 52 Wochen; Nachbeobachtung 28 Tage nach Therapieende.
Population	<u>Wesentliche Einschlusskriterien:</u> <ul style="list-style-type: none"> • Unterschriebene Einverständniserklärung • Alter \geq 40 Jahre zu Visite 1 • Diagnose IPF nach ATS-, ERS-, JRS- und ALAT-Richtlinien (innerhalb von 5 Jahren vor Baseline) • Thorax HRCT innerhalb von 12 Monaten vor Visite 1 • Kombination des HRCT-Muster und, wenn verfügbar, chirurgisches Lungenbiopsie-Muster, übereinstimmend mit IPF Diagnose bewertet von zentralen Reviewern • DL_{CO} (korrigiert für Hb) 30 % bis 79 % • FVC \geq 50 % des Normalwertes zu Visite 2 <u>Wesentliche Ausschlusskriterien:</u> <ul style="list-style-type: none"> • Laborparameter (zu Visite 1) <ul style="list-style-type: none"> ○ AST und ALT > 1,5-fach ULN ○ Bilirubin > 1,5-fach ULN • Lungenfunktion (zu Visit 2)

Dossierbewertung für Orphan Drugs

Nintedanib

Charakteristikum	Beschreibung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ¹
	<ul style="list-style-type: none"> ○ Atemwegsobstruktion (forciertes exploratorisches Volumen in 1 Sekunde vor Einnahme eines Bronchodilators [FEV1]/FVC < 0,7) ● Lungentransplantation im Verlauf der Studie wahrscheinlich (nach Einschätzung des Prüfarztes) ● Andere Erkrankung: <ul style="list-style-type: none"> ○ Herzerkrankung <ul style="list-style-type: none"> ▪ Myokardinfarkt innerhalb von 6 Monaten vor Visite 2 ▪ Instabile Angina innerhalb von 1 Monat vor Visite 2 ○ Blutungsrisiko <ul style="list-style-type: none"> ▪ Bekannte erbliche Prädisposition für Blutungen ▪ Hämorrhagisches ZNS-Ereignis innerhalb der letzten 12 Monate ▪ Hämoptysis oder Hämaturie ▪ Aktive gastrointestinale Blutung oder Geschwüre ▪ Größere Verletzung oder Operation ▪ Koagulationsparameter: INR > 2, Verlängerung der Prothrombinzeit und PTT (sek) um > 50% ULN (Visite 1) ▪ Fibrinolyse mit im Studienprotokoll festgelegten Interventionen (siehe Tabelle 2) ○ Thromboserisiko <ul style="list-style-type: none"> ▪ Bekannte erbliche Prädisposition für Thrombose ▪ Thrombotisches Ereignis innerhalb von 12 Monaten (inkl. Schlaganfall und transiente ischämische Attacke) vor Visite 2 ○ Bekannte Hypersensitivität gegenüber Studienmedikation oder deren Bestandteilen ○ Andere Krankheiten, die das Testprozedere oder nach Ermessen des Prüfarztes die Studienteilnahme beeinflussen oder den Patienten bei Studienteilnahme in Gefahr bringen ○ Lebenserwartung aufgrund anderer Krankheiten als IPF < 2,5 Jahre (Ermessen des Prüfarztes) ● Andere Interventionen (Forschungsstudie) innerhalb von 8 Wochen vor Visite 1 ● Geplante größere Operationen während der Studienperiode ● Schwangere oder stillende Frauen oder Frauen im gebärfähigen Alter, die keine zwei adäquaten Verhütungsmethoden für mind. 1 Monat vor Behandlungsbeginn und bis 3 Monate nach Behandlungsende verwenden ● Sexuell aktive Männer, die sich nicht verpflichten, Kondome während und 3 Monate nach Ende der Behandlung zu nutzen (außer wenn Partner nicht im gebärfähigen Alter sind) ● Aktiver Alkohol- oder Drogenmissbrauch ● Patienten, die das Studienprozedere nicht verstehen oder einhalten (inkl. Ausfüllen von Fragebögen ohne Hilfe) ● Frühere Behandlung mit Nintedanib länger als 4 Wochen ● Einnahme von N-Acetylcystein (NAC), Prednison > 15 mg/Tag oder vergleichbare Arzneimittel innerhalb von 2 Wochen vor Visite 1 ● Einnahme von Pirfenidon, Azathioprin, Cyclophosphamid, Cyclosporin A innerhalb von 8 Wochen vor Visite 1

Charakteristikum	Beschreibung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ¹								
Interventionen und Zahl der Patienten	<p>INPULSIS-1</p> <table border="1" data-bbox="373 427 1059 521"> <thead> <tr> <th>Nintedanib 150 mg bid</th> <th>Placebo</th> </tr> </thead> <tbody> <tr> <td>N=309</td> <td>N=206</td> </tr> </tbody> </table> <ul style="list-style-type: none"> • Eingeschlossene N=718, davon 203 Patienten nicht randomisiert, aufgrund der Verletzung von Ein- bzw. Ausschlusskriterien der Studie • Randomisiert N= 515 • Tatsächlich die geplante Behandlung erhalten N=513 (Nintedanib n=309; Placebo: n=204) <p>INPULSIS-2</p> <table border="1" data-bbox="373 790 1059 884"> <thead> <tr> <th>Nintedanib 150 mg bid</th> <th>Placebo</th> </tr> </thead> <tbody> <tr> <td>N=331</td> <td>N=220</td> </tr> </tbody> </table> <ul style="list-style-type: none"> • Eingeschlossen N=794, davon wurden 243 Patienten nicht randomisiert aufgrund der Verletzung von Ein- bzw. Ausschlusskriterien der Studie • Randomisiert N=551 • Tatsächlich die geplante Behandlung erhalten N=548 (Nintedanib n=329 Placebo n=219) 	Nintedanib 150 mg bid	Placebo	N=309	N=206	Nintedanib 150 mg bid	Placebo	N=331	N=220
Nintedanib 150 mg bid	Placebo								
N=309	N=206								
Nintedanib 150 mg bid	Placebo								
N=331	N=220								
Ort und Zeitraum der Durchführung	<p>INPULSIS-1</p> <p><u>Studiendauer:</u> 52 Wochen + 28 Tage Nachbeobachtung <u>Zeitraum der Durchführung:</u> 05/2011-10/2013 <u>Ort der Durchführung:</u> Die Studie wurde in 98 klinischen Zentren in 13 Ländern durchgeführt: Australien [N=10], Asien (China [N=16], Indien [N=4], Japan [N=20]) Europa (Belgien [N=9], Frankreich [N=29], Deutschland [N=31], Irland [N=1], Italien [N=30] Tschechische Republik [N=6], Vereinigtes Königreich [N=13], Israel [N=10]) USA [N=27]</p> <p>INPULSIS-2</p> <p><u>Studiendauer:</u> 52 Wochen + 28 Tage Nachbeobachtung <u>Zeitraum der Durchführung:</u> 05/2011-10/2013 <u>Ort der Durchführung:</u> Die Studie wurde in 107 klinischen Zentren in 17 Ländern durchgeführt: Asien (China [N=24], Indien [N=6], Japan [N=30], Korea [N=26]) Europa (Finnland [N=5], Frankreich [N=19], Deutschland [N=14], Griechenland [N=11], Niederlande [N=9], Portugal [N=7], Spanien [N=7]), Latein- und Südamerika (Chile [N=4], Mexiko [N=0]), Nordamerika (Kanada [N=6], USA [N=35]), Russland [N=1], Türkei [N=16]</p>								
Primärer Endpunkt, sekundäre Endpunkte und explorative Endpunkte	<p><u>Primärer Endpunkt:</u></p> <ul style="list-style-type: none"> • Jährliche FVC- Abnahme (ausgedrückt in ml über alle Messwerte innerhalb von 52 Wochen) <p><u>Sekundäre Endpunkte:</u></p> <ul style="list-style-type: none"> • Weitere Analysen der FVC <ul style="list-style-type: none"> ○ Absolute und relative Veränderung des FVC (ml) zwischen Baseline-Wert und 								

Charakteristikum	Beschreibung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ¹
gemäß Studienprotokoll	<p>Wert nach 52 Wochen</p> <ul style="list-style-type: none"> ○ Absolute und relative Veränderung des FVC % vom Sollwert zwischen Baseline-Wert und Wert nach 52 Wochen ○ Rate der FVC Responder zwischen Baseline-Wert und Wert nach 52 Wochen: Abfall >5%, Anstieg >5% und Veränderungen ≤ 5% und Abfall >10%, Anstieg >10% und Veränderung ≤ 10 % <ul style="list-style-type: none"> • Zeit bis zur ersten akuten (adjudizierten) Exazerbation über 52 Wochen (in Tagen) • Veränderung des SGRQ Gesamtscores von Baseline bis Woche 52 • Anteil SGRQ-(≤ -4 Punkte)-Responder (definiert als absolute Veränderung im SGRQ-Gesamtscore um ≤ 4 Punkte) • Veränderung SGRQ-Domänen in Punkten • Veränderung SGRQ-I Gesamtscore • Veränderung SOBQ • Veränderung Hustenimpact und Hustensymptome (CASA-Q)-Score • Anteil PGI-C Responder • Veränderung EQ-5D-VAS • Zeit bis zum Tod • Zeit bis zum respiratorisch bedingten Tod • Zeit bis zum Tod unter Behandlung • Zeit bis zum Tod oder Lungentransplantation • Zeit bis zum Tod oder Lungentransplantation oder Qualifizierung für Lungentransplantation (Kriterien: FVC < 45% vom Sollwert oder DLCO < 30% vom Sollwert oder SpO₂ < 88% in Ruhe) • Veränderung SpO₂ in Ruhe • Veränderung DL_{CO} in Ruhe • Prä-Dosis Plasmakonzentration (C_{pre,ss,N}) von Nintedanib, BIBF 1202 und BIBF 1202 Glukuronid Visite 4 (4 Wochen) und Visite 7 (24 Wochen) <p><u>Post-hoc für das Dossier ausgewertete Endpunkte (nicht im Studienbericht):</u></p> <ul style="list-style-type: none"> • SGRQ-I-(≤ -4 Punkte)-Responder <p><u>Sicherheit:</u></p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse • Vitalzeichen • Körperliche Untersuchung • Gewicht • Klinische Labortests • 12-Kanal-EKG (japanische Studienzentren)
Subgruppenanalysen	<p><u>Subgruppenanalysen wurden für die folgenden Variablen geplant</u></p> <ul style="list-style-type: none"> • Geschlecht • Alter • Japanische versus nicht-japanische Patienten • Ethnie • Gewicht

Charakteristikum	Beschreibung der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) ¹
	<ul style="list-style-type: none"> • Körperoberfläche • Nierenfunktion basieren auf Kreatinin-Clearance und GFR <p>Außer für japanische versus nicht-japanische wurden alle Subgruppenanalysen nur für pharmakokinetische Endpunkte geplant und in den Studien dokumentiert.</p> <p><u>Subgruppenanalysen im Dossier (post-hoc)</u></p> <ul style="list-style-type: none"> • Geschlecht: männlich, weiblich • Alter: < 65 Jahre, ≥ 65 Jahre, < 65 Jahre, 65-74 Jahre, ≥ 75 Jahre • Ethnie: kaukasisch, asiatisch • FVC (% pred.): ≤ 70 %, > 70 % • SGRQ Baselinewert: ≤ 40 Punkte, > 40 Punkte • Raucherstatus: Nichtraucher, Ex-Raucher/Raucher • systemische Kortikosteroide: ja, nein • Bronchodilator: ja, nein • Niereninsuffizienz: ja, nein • Gewicht: < 65 kg, ≥ 65 kg • Studie: 1199.32, 1199.34 <p>Für alle im Dossier berichteten Endpunkte wurden Subgruppenanalysen post hoc durchgeführt.</p>

¹⁾ Sofern nicht eindeutig kenntlich gemacht, gelten alle Angaben sowohl für INPULSIS-1 als auch für INPULSIS-2

Tabelle 2: Charakterisierung der Intervention in den Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Nintedanib 150 mg bid	Placebo
<p><u>Studie INPULSIS-1 (1199.32)</u> Nintedanib 150 mg 2-mal täglich per os über 52 Wochen</p> <ul style="list-style-type: none"> • Einnahme mit einem Glas Wasser in einem Intervall von 12 Stunden, wenn möglich täglich zur gleichen Zeit (morgens zwischen 6.00 Uhr und 11.00 Uhr und abends zwischen 18.00 Uhr und 23.00 Uhr). Zur Vermeidung von Magenproblemen sollte die Einnahme nach dem Essen erfolgen. <p><i>Schema zur Dosisreduktion:</i></p> <ul style="list-style-type: none"> • Beim Auftreten bestimmter UE konnte eine Dosisreduktion auf 100 mg zweimal täglich oder eine Unterbrechung erwogen werden. • Im Falle einer erheblichen Toxizität oder wenn auch eine reduzierte Dosis nicht toleriert wurde, sollte die Therapie abgebrochen und nicht wieder aufgenommen werden. Falls die Dosisreduzierung toleriert wurde, war eine Reeskalation zur Startdosis innerhalb von 4 Wochen erlaubt. 	<p><u>Studie INPULSIS-1 (1199.32)</u> Placebo 2-mal täglich per os über 52 Wochen</p> <ul style="list-style-type: none"> • Schema zur Einnahme wie im Interventionsarm <p><i>Schema zur Dosisreduktion:</i></p> <ul style="list-style-type: none"> • Schema zur Dosisreduktion wie im Interventionsarm
<u>Studie INPULSIS-2 (1199.34)</u>	<u>Studie INPULSIS-2 (1199.34)</u>

Nintedanib 150 mg bid	Placebo
Homogenes Studiendesign zu INPULSIS-1	Homogenes Studiendesign zu INPULSIS-1
<p><u>Nicht erlaubte Begleitmedikation</u> Innerhalb des gesamten Behandlungszeitraumes und 8 Wochen vor Visite 1:</p> <ul style="list-style-type: none"> • Pirfenidon • Jede andere experimentelle Therapie zur Behandlung der IPF <p>Die folgenden Therapien waren während des gesamten Behandlungszeitraumes nicht erlaubt. Falls sie jedoch im Verlauf der Studie als medizinisch notwendig erachtet wurden, war eine 4-wöchige Auswaschphase der Studienmedikation vor ihrem Einsatz zu beachten:</p> <ul style="list-style-type: none"> • Fibrinolyse, Antikoagulantien in voller therapeutischer Wirkungsdosierung (z.B. Vitamine-K-Antagonisten, Dabigatran, Heparin, Hirudin etc.) oder eine hochdosierte antithrombozytären Therapie. <p><u>Erlaubte Begleitmedikation</u> Die folgende Begleitmedikation war erlaubt sofern sie für mindestens 8 Wochen vor Visite 1 in stabiler Dosierung eingenommen wurde:</p> <ul style="list-style-type: none"> • Prednison, wenn es in einer gleichbleibenden Dosierung von ≤ 15 mg/Tag (oder ≤ 30 mg alle 2 Tage) eingenommen wurde oder eine äquivalente Dosierung anderer oraler Kortikosteroid. <p>Im Falle einer akuten Exazerbation:</p> <ul style="list-style-type: none"> • Jegliche Medikamente (z.B. Prednison in hoher Dosierung, Azathioprine, Cyclophosphamide, Cyclosporin A, Acetylcystein), die für das Wohlergehen der Patienten notwendig waren, konnten gemäß Entscheidung der Prüfarzte eingesetzt werden mit Ausnahme von Pirfenidon <p>Im Falle einer Verschlechterung nach 6 Monaten:</p> <ul style="list-style-type: none"> • Jegliche Medikamente (z.B. Prednison in hoher Dosierung, Azathioprine, Cyclophosphamide, Cyclosporin A, Acetylcystein), die für das Wohlergehen der Patienten notwendig waren, konnten gemäß Entscheidung der Prüfarzte eingesetzt werden mit Ausnahme von Pirfenidon. Eine Abnahme von ≥ 10 % des FVC % vom Sollwert oder eine Abnahme von 15 % des DL_{CO} % vom Sollwert, jeweils bezogen auf den Sollwert, wurde als Verschlechterung angesehen (Fortschreiten der Erkrankung) <p>Prophylaktische Antikoagulation:</p> <ul style="list-style-type: none"> • Niedrig dosiertes Heparin oder Heparin-Spüllösungen zur Pflege von implantierten Medizinprodukten (z.B. Enoxaparin 4000 I.U. s.c. pro Tag) sowie antithrombozytäre Therapie (z.B. Acetylsalicylsäure bis zu 324 mg/Tag oder Clopidogrel 75 mg/Tag oder eine äquivalente Dosierung anderer antithrombozytärer Therapien zur Prophylaxe waren gestattet. <p><u>Eingeschränkte Begleitmedikation:</u> Die folgenden Therapien waren innerhalb der ersten 6 Monate der Behandlung außer im Fall einer akuten Exazerbation aufgrund der IPF nicht erlaubt. Nach 6 Monaten war eine Einnahme im Fall einer Verschlechterung nach Ermessen des Prüfarztes erlaubt.</p> <ul style="list-style-type: none"> • Azathioprine, Cyclophosphamide und Cyclosporin A. Diese durften nicht innerhalb von 8 Wochen nach Visite 1 eingenommen werden. • Acetylcystein und Prednison, in einer Dosierung > 15 mg/Tag (oder > 30 mg alle 2 Tage oder eine äquivalente Dosierung anderer oraler Kortikosteroid). Diese durften nicht innerhalb von 2 Wochen nach Visite 1 eingenommen werden. • Eine Auswaschphase für Bronchodilatatortherapie von 24 Stunden (für langwirksame Wirkstoffe) bzw. 8 Stunden (für Wirkstoffe mit kurzer Halbwertszeit) vor spirometrischen Untersuchungen 	

Die vom pU im Dossier dargestellten Studien INPULSIS-1 und INPULSIS-2 weisen ein homogenes Studiendesign auf. Hinsichtlich der Charakteristika der eingeschlossenen Studienpopulation sind die Studien ebenfalls grundsätzlich vergleichbar (siehe Tabelle 3). Das mittlere Alter in den Studien betrug 66,4 Jahre bis 67,1 Jahre, wobei die Verteilung nach Alterskategorien etwas unausgeglichen war, mit jeweils mehr Patienten in der mittleren Alterskategorie zwischen 65 und <75 Jahren in den Placebo-Armen der Studien und mehr Patienten im Alter zwischen 45 bis <65 Jahre und ≥ 75 Jahre in den Nintedanib-Armen der Studien. Die Zeit seit der IPF-Diagnose betrug in den Behandlungsgruppen der Studien INPULSIS-1 und INPULSIS-2 1,55 Jahre bis 1,66 Jahre. Das entspricht dem Durchschnittsalter bei Diagnosestellung von ca. 66 Jahren (siehe Abschnitt 1) der Erkrankung.

Der Anteil an männlichen Studienteilnehmern lag zwischen 77,8 % und 81,2 %. Der pU beschreibt im Dossier (siehe Modul 4, Abschnitt 4.3.1.2.1), dass die IPF auch in der deutschen Versorgungsrealität bei Männern häufiger Auftritt auf als bei Frauen. Allerdings sei der Anteil Männer laut pU in den Studien noch höher als unter IPF-Patienten in der Gesamtbevölkerung. Der Anteil der Kaukasier an der Studienpopulation unterscheidet sich leicht innerhalb der Studien: ca. 65 % in der Studie INPULSIS-1 und ca. 50 % in der Studie INPULSIS-2. Der Anteil der Patienten asiatischer Herkunft betrug ca. 21 % (INPULSIS-1) und 39 % (INPULSIS-2).

Tabelle 3: Charakterisierung der Studienpopulation in INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) zum Baseline

	INPULSIS-1 (1199.32) ¹⁾		INPULSIS-2 (1199.34) ¹⁾		INPULSIS-1 & INPULSIS-2 gepoolt ¹⁾	
	Nintedanib 150 mg bid N=309	Placebo N=204	Nintedanib 150 mg bid N=329	Placebo N=219	Nintedanib 150 mg bid N=638	Placebo N=423
<i>Alter, Jahre</i>						
Mittelwert (SD)	66,9 (8,4)	66,9 (8,2)	66,4 (7,9)	67,1 (7,5)	66,6 (8,1)	67,0 (7,9)
Median (min;max)	68,0,(42;44)	68,0 (45;87)	66,0 (42;98)	68,0 (42;88)		
<i>Altersklassen, n (%)</i>						
<65 Jahre	119 (38,5)	71 (34,8)	139 (42,2)	74 (33,8)	258 (40,44)	145 (34,28)
≥65 Jahre bis <75 Jahre	130 (42,1)	102 (50,0)	133 (40,4)	114 (52,1)	263 (41,22)	216 (51,06)
≥75 Jahre	60 (19,4)	31 (15,2)	57 (17,3)	31 (14,2)	117 (18,34)	62 (14,66)
<i>Geschlecht, n (%)</i>						
Männlich	251 (81,2)	163 (79,9)	256 (77,8)	171 (78,1)	507 (79,5)	334 (79,0)
Weiblich	58 (18,8)	41 (20,1)	73 (22,2)	48 (21,9)	131 (20,5)	89 (21,0)
<i>Ethnie (genetisch), n (%)²⁾</i>						
Kaukasier/weiß	198 (64,1)	135 (66,2)	162 (49,2)	113 (51,6)	360 (56,43)	248 (58,63)
schwarz	0	0	2 (0,6)	0	2 (0,31)	0
asiatisch	66 (21,4)	41 (20,1)	128 (38,9)	86 (39,7)	194 (30,41)	128 (30,26)
keine Information	45 (14,6)	28 (13,7)	37 (11,2)	19 (8,7)	82 (12,85)	47 (11,11)
<i>Gewicht, kg</i>						
MW (SD)	82,03 (16,80)	81,17 (16,30)	76,59 (15,93)	76,28 (16,46)	79,23 (16,57)	78,64 (16,54)
<i>BMI, kg/m²</i>						
MW (SD)	28,63 (4,52)	28,12 (4,63)	27,56 (4,55)	27,17 (4,49)	28,08 (4,57)	27,63 (4,58)
<i>Raucherstatus, n (%)</i>						
Nicht-Raucher	71 (23,0)	51 (25,0)	103 (31,1)	71 (32,4)	174 (27,3)	122 (28,8)
Ex-Raucher	217 (70,2)	144 (70,6)	218 (66,3)	139 (63,5)	435 (68,2)	283 (66,9)
Raucher	21 (6,8)	9 (4,4)	8 (2,4)	9 (4,1)	29 (4,5)	18 (4,3)

	INPULSIS-1 (1199.32) ¹⁾		INPULSIS-2 (1199.34) ¹⁾		INPULSIS-1 & INPULSIS-2 gepoolt ¹⁾	
	Nintedanib 150 mg bid N=309	Placebo N=204	Nintedanib 150 mg bid N=329	Placebo N=219	Nintedanib 150 mg bid N=638	Placebo N=423
Zeit seit IPF-Diagnose, Jahre MW (SD)	1,66 (1,37)	1,59 (1,35)	1,64 (1,34)	1,55 (1,27)	1,65 (1,36)	1,57 (1,31)
<i>FVC, MW (SD)</i>						
in ml	2.756,8 (735,1)	2.844,5 (820,1)	2.672,8 (776,0)	2.619,0 (787,3)	2.713,5 (757,0)	2.727,7 (810,2)
in % vom Sollwert	79,47 (17,03)	80,53 (17,34)	79,99 (18,08)	78,09 (18,97)	79,74 (17,57)	79,27 (18,22)
FEV ₁ / FVC (%), MW (SD)	81,54 (5,36)	80,83 (6,13)	81,77 (6,25)	82,44 (5,72)	81,66 (5,83)	81,66 (5,97)
DL _{CO} (mmol/min/kPa), MW (SD)	3,96 (1,20)	3,96 (1,11)	3,77 (1,23)	3,75 (1,32)	3,87 (1,22)	3,85 (1,23)
SpO ₂ (%), MW (SD)	95,9 (2,0)	95,9 (1,9)	95,8 (2,6)	95,7 (2,1)	95,9 (2,3)	95,8 (2,0)
Systemische Kortikosteroid-Therapie, n (%)	68 (22,0)	43 (21,1)	68 (20,7)	46 (21,0)	136 (21,3)	89 (21,0)

¹⁾ Treated Set

²⁾ Das Merkmal Ethnie wurde in Frankreich aufgrund entsprechender gesetzlicher Vorschriften nicht erfasst.

Abkürzungen: bid: zweimal täglich, BMI: Body Mass Index, MW: Mittelwert, DL_{CO}: Kohlenmonoxid-Diffusionskapazität, FEV₁: forciertes exploratorisches Volumen in 1 Sekunde, FVC: Forcierte Vitalkapazität, IPF: idiopathische Lungenfibrose, kPA: Kilopascal, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, SD: Standardabweichung, SpO₂: Blutsauerstoffsättigung

2.5 Endpunkte und Erhebungsinstrumente

2.5.1 In der Nutzenbewertung berücksichtigte Endpunkte

In der Nutzenbewertung wurden die in Tabelle 4 abgebildeten und vom pU als patientenrelevant eingestuften Endpunkte berücksichtigt:

Tabelle 4: Vom pU im Dossier als patientenrelevant bezeichnete und in der Nutzenbewertung berücksichtigte Endpunkte der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Endpunkt ¹⁾	Kategorie	INPULSIS-1 und INPULSIS-2	
		Vom pU als patientenrelevant eingestuft	In der Nutzenbewertung berücksichtigt
Zeit bis Tod, unabhängig von der Ursache	Mortalität	•	•
Zeit bis zum respiratorisch bedingtem Tod		•	•
Primärer Endpunkt der Studien: Jährliche FVC-Abnahme (ml)		•	○ ²⁾
FVC-Responder (Schwellenwert 10 % bzw. 5 % vom Sollwert)		•	○ ²⁾
Zeit bis zur ersten akuten Exazerbation	Morbidität	•	•
Zeit bis zur ersten adjudizierten akuten Exazerbation		•	•
Veränderung EQ-5D-VAS		•	•
PGI-C-Responder		•	•
Veränderung SGRQ-Gesamtscore	Lebensqualität	•	•
SGRQ-(≤ -4 Punkte)-Responder		•	•
Veränderung SGRQ-I-Gesamtscore		•	•
SGRQ-I-(≤ -4 Punkte)-Responder		•	•
Veränderung SOBQ-Gesamtscore		•	•
Veränderung CASA-Q Husten-Symptome		•	•
Veränderung CASA-Q Husten-Impact	•	•	
Gesamtzahl UE	Sicherheit	•	•
schwerwiegende UE		•	•
UE, die zum Therapieabbruch führten		•	•
UE, die zum Tod führten		•	•
UE von besonderem Interesse		•	•

¹⁾ Die Definition der Endpunkte stimmen für die Studien INPULSIS-1 und INPULSIS-2 überein.

²⁾ FVC wird als nicht ausreichend validiertes Surrogat für Mortalität angesehen. Da die jährliche Abnahme des FVC primärer Endpunkt der Studien war, wird der Endpunkt ergänzend dargestellt

• bewertungsrelevant ○ nicht bewertungsrelevant, aber primärer Endpunkt

Abkürzungen: CASA-Q: Cough and Sputum Assessment Questionnaire, EQ-5D-VAS: EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale, FVC: forcierte Vitalkapazität, IPF: idiopathische Lungenfibrose, PGI-C: Patient's Global Impression of Change, pU: pharmazeutischer Unternehmer, SGRQ: St. George's Respiratory Questionnaire, SGRQ: St. George's Respiratory Questionnaire IPF spezifische Version, SOBQ: Shortness of Breath Questionnaire, UE: unerwünschtes Ereignis

2.5.2 Verzerrungspotential auf Studien- und Endpunktebene

Tabelle 5: Verzerrungspotential der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34) auf Studienebene

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung Patient	Verblindung Behandler	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial auf Studienebene
INPULSIS-1	ja	ja	ja	ja	nein	nein	niedrig
INPULSIS-2	ja	ja	ja	ja	nein	nein	niedrig

Bei den beiden Studien INPULSIS-1 und INPULSIS-2 handelt es sich um randomisierte, doppelblinde, placebokontrollierte Studien. Die Zuteilung auf die Behandlungsgruppen Nintedanib und Placebo erfolgte durch ein interaktives Voice-Response-System bzw. Web-Response-System (IXRS) im Verhältnis 3:2. Die Randomisierung war nicht stratifiziert. Die aktive Studienmedikation und Placebo waren äußerlich nicht voneinander unterscheidbar. Das Verzerrungspotential auf Studienebene wird für beide Studien als niedrig eingeschätzt. Das Verzerrungspotential der relevanten Endpunkte wird in Tabelle 6 dargestellt.

Tabelle 6: Verzerrungspotential der in der Nutzenbewertung berücksichtigten Endpunkte in den Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Endpunkt ¹⁾	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial
Mortalität					
Zeit bis Tod, unabhängig von der Ursache	ja	ja	nein	nein	niedrig
Zeit bis respiratorisch bedingtem Tod	ja	ja	nein	nein	niedrig
Jährliche FVC-Abnahme (ml) ¹⁾	ja	ja	nein	nein	niedrig
FVC-Responder (Schwellenwert 10% bzw. 5% vom Sollwert)	ja	ja	nein	nein	niedrig
Morbidität					
Zeit bis zur ersten adjudizierten akuten Exazerbation	ja	ja	nein	nein	niedrig

Endpunkt ¹⁾	Verblindung adäquat	ITT adäquat	Hinweise auf ergebnis- gesteuerte Berichter- stattung	Sonstige das Verzerrungs- potenzial beeinflussende Punkte	Verzerrungs- potenzial
Zeit bis zur ersten akuten Exazerbation	ja	ja	nein	nein	niedrig
Veränderung EQ-5D-VAS	ja	ja	nein	nein	niedrig
PGI-C-Responder	ja	ja	nein	nein	niedrig
Lebensqualität					
Veränderung SGRQ-Gesamtscore	ja	ja	nein	nein	niedrig
SGRQ-(\leq -4 Punkte)-Responder	ja	ja	nein	nein	niedrig
Veränderung SGRQ-I-Gesamtscore	ja	ja	nein	nein	niedrig
SGRQ-I-(\leq -4 Punkte)-Responder	ja	ja	nein	nein	niedrig
Veränderung SOBQ-Gesamtscore	ja	ja	nein	nein	niedrig
Veränderung CASA-Q Husten-Symptome	ja	ja	nein	nein	niedrig
Veränderung CASA-Q Husten-Impact	ja	ja	nein	nein	niedrig
Sicherheit					
Gesamtzahl UE	ja	ja	nein	nein	niedrig
schwerwiegende UE	ja	ja	nein	nein	niedrig
UE, die zum Therapieabbruch führten	ja	ja	nein	nein	niedrig
UE, die zum Tod führten	ja	ja	nein	nein	niedrig
UE von besonderem Interesse ²⁾	ja	ja	nein	nein	niedrig

¹⁾ Die Einschätzung des Verzerrungspotentials unterscheidet sich nicht zwischen INPULSIS-1 und INPULSIS-2. Daher wird auf eine getrennte Darstellung beider Studien verzichtet.

²⁾ Als UE von besonderem Interesse wurden im Dossier dargestellt: unerwünschte gastrointestinale Ereignisse, unerwünschte Leberenzym- und Bilirubinveränderungen, unerwünschte kardiale Ereignisse, unerwünschte respiratorische Ereignisse, unerwünschte Blutungsereignisse, unerwünschte Ereignisse mit einer Häufigkeit von > 10%. Diese Definition weicht zum Teil von der Definition der Zulassungsstudien ab. Eine Begründung für die Hinzunahme weiterer UE liegt nicht vor. Zudem sind einige der vom pU post-hoc definierten UE von besonderem Interesse aus der Gruppe respiratorische Ereignisse bereits Bestandteil von Wirksamkeitsendpunkten, wie etwa Progression der IPF. UE, die bereits Bestandteil von Wirksamkeitsendpunkten waren, wurden nicht als UE von besonderem Interesse berücksichtigt.

Abkürzungen: CASA-Q: Cough and Sputum Assessment Questionnaire, EQ-5D-VAS: EuroQol 5-Dimensional Quality of Life Questionnaire Visual Analogue Scale, FVC: forcierte Vitalkapazität, IPF: idiopathische Lungenfibrose, ITT: Intention to Treat, PGI-C: Patient's Global Impression of Change, pU: pharmazeutischer Unternehmer, SGRQ: St. George's Respiratory Questionnaire, SGRQ: St. George's Respiratory Questionnaire IPF spezifische Version, SOBQ: Shortness of Breath Questionnaire, UE: unerwünschtes Ereignis

2.5.3 Charakterisierung der berücksichtigten Endpunkte und Einschätzung der Patientenrelevanz

Laut Verfo des G-BA gilt der medizinische Zusatznutzen durch die Zulassung als belegt. Nachweise gemäß § 5 Absatz 1 bis 6 sind nicht verpflichtend vorzulegen. Zur Beurteilung der Patientenrelevanz der vorgelegten Endpunkte wurden das Dossier, die zugehörigen Studien-

Dossierbewertung für Orphan Drugs

Nintedanib

berichte und die in diesen Quellen zitierte Literatur herangezogen. Ferner wurden eigene Recherchen zur Validität und Patientenrelevanz der Endpunkte durchgeführt. Das Ergebnis der Betrachtung der Endpunkte findet sich aufgeteilt nach Endpunktkategorien in den Tabelle 7 bis Tabelle 10.

Die Operationalisierungen der Endpunkte der Studien INPULSIS-1 und INPULSIS-2 war identisch.

Tabelle 7: Beschreibung der Endpunkte der Kategorie Mortalität

Mortalität	
Endpunkt	Zeit bis Tod, unabhängig von der Ursache
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u> Die Zeit bis zum Tod, unabhängig von der Ursache, wurde als Zeit von der Randomisierung bis zum Todeszeitpunkt bzw. zum letzten Patientenkontakt definiert. Die Erhebung erfolgte durch den Prüfarzt. Überlebende Studienteilnehmer wurden zum Zeitpunkt des letzten Patientenkontaktes maximal nach 372 Tagen (52 Wochen + 7 Tage) zensiert.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU stellt dar, dass die IPF eine schwerwiegende und in der Regel zum Tod führende Erkrankung ist. Die Verlängerung der Gesamtüberlebenszeit sei daher ein wichtiges Therapieziel und patientenrelevant.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u> Die Operationalisierung des Endpunkts ist nachvollziehbar.</p> <p><u>Validität und Patientenrelevanz:</u> Das Gesamtüberleben ist ein patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittelnutzenbewertungsverordnung (AM-NutzenV).</p>
Endpunkt	Zeit bis respiratorisch bedingtem Tod
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u> Die Zeit bis zum respiratorisch bedingten Tod wurde als die Zeit von Randomisierung bis zum Zeitpunkt des respiratorisch bedingten Todes bzw. dem letzten Patientenkontakt definiert. Alle Todesfälle wurden hinsichtlich ihrer Ursache durch ein verblindetes Independent Adjudication Committee (IAC, Unabhängiges Bewertungskomitee) bewertet, dem drei Experten aus dem Bereich der Pneumologie mit besonderen Fachkenntnissen in der Behandlung der IPF angehörten. Studienteilnehmer, die aufgrund einer anderen Ursache verstarben, wurden zum Zeitpunkt ihres Todes zensiert. Überlebende Studienteilnehmer wurden zum Zeitpunkt des letzten Patientenkontaktes maximal nach 372 Tagen (52 Wochen + 7 Tage) zensiert.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU stellt dar, dass die IPF eine schwerwiegende und in der Regel zum Tod führende Erkrankung ist. Die Verlängerung der Gesamtüberlebenszeit sei daher ein wichtiges Therapieziel und patientenrelevant.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u> Die Bewertung des Endpunktes Zeit bis zum respiratorisch bedingten Tod basierend auf der Bewertung des IAC wurde a priori definiert. Kriterien für die Einstufung des Todes als respiratorisch bedingt werden in den vom pU eingereichten Unterlagen nicht genannt. Ebenso fehlen in den Unterlagen weiterführende Angaben zur Qualifikation der Experten des IAC.</p>

Mortalität	<u>Validität und Patientenrelevanz:</u> Die Zeit bis zum respiratorisch bedingten Tod ist grundsätzlich ein patientenrelevanter Endpunkt. Aufgrund der unzureichenden Darstellung der Kriterien für die Einstufung des Todes als respiratorisch bedingt, ist die Validität dieses Endpunktes jedoch fraglich.
Endpunkt	Jährliche FVC-Abnahme (ml) und FVC-Responder (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert)
Operationalisierung und Validierung: Begründung seitens des pU	<u>Operationalisierung:</u> Die jährliche FVC-Abnahme in ml wurde mittels eines Random-Coefficient-Modells mit Geschlecht, Alter und Körpergröße als Kovariate ermittelt. Die Bestimmung der FVC erfolgte bei Studienbeginn sowie zu Woche 2, 4, 6, 12, 24, 36 und 52 mittels Spirometrie. FVC Werte zu allen Zeitpunkten bis einschließlich 52 Wochen gingen in das Modell ein. Der Test erfolgte durch den Prüfarzt und wurde dreifach in sitzender Position jeweils zur gleichen Tageszeit (± 60 min.) und ohne größere körperliche Belastung innerhalb von 12 Std. vor der Erhebung durchgeführt. Die Studienteilnehmer wurden gebeten am Tag der Erhebung nicht zu rauchen und kalte Temperaturen, Staub, Rauch und unangenehme Gerüche (z.B. Parfüm) zu vermeiden. Kaffee, Tee, Schokolade, Cola und andere koffeinhaltige Getränke, alkoholhaltige Getränke oder Nahrungsmittel und eisgekühlte Getränke waren am Morgen der Erhebung nicht erlaubt. Die Auswaschphase für eine Bronchodilatatortherapie betrug 24 Stunden (für langwirksame Wirkstoffe) bzw. 8 Stunden (für Wirkstoffe mit kurzer Halbwertszeit) vor der Testdurchführung. Als FVC-Responder (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert) war ein Patient definiert, dessen FVC in Woche 52 gegenüber dem individuellen Messwert zu Studienbeginn um nicht mehr als 10 % bzw. 5 % (bezogen auf den Sollwert) abfiel. Die Erhebung erfolgte durch den Prüfarzt mittels Spirometer. Falls zu Woche 52 kein Messwert vorlag, wurde der entsprechende Studienteilnehmer konservativ als Non-Responder bewertet. <u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der Endpunkt jährliche FVC-Abnahme wird vom pU als Mortalitätsendpunkt dargestellt. Er argumentiert, dass der Effekt der Behandlung mit Nintedanib auf die Mortalität indirekt durch das nach seinen Angaben validierte Surrogat jährliche FVC-Abnahme gemessen in ml erfasst werde. Der pU argumentiert, dass es aufgrund der zunehmenden Vernarbung der Lunge zu einer mit der Zeit fortschreitenden Abnahme der FVC bei der IPF komme, sodass die Belüftung der Lunge zunehmend eingeschränkt wird bzw. eine erhebliche Atemanstrengung des Patienten erfordert. Da die Vernarbungsprozesse gegenwärtig weder rückgängig gemacht noch gestoppt werden können, zielt die Therapie der IPF darauf ab, den Abfall der FVC und damit die Krankheitsprogression zu verlangsamen und aus Sicht des pU das Mortalitätsrisiko zu verringern. Die FVC ist laut pU der am häufigsten verwendete physiologische Messparameter in klinischen Studien in der Pneumologie. Laut der Darstellung des pU komme angesichts des heterogenen und nicht absehbaren Verlaufs einer IPF prognostischen Faktoren eine besondere Bedeutung zu. Der pU stellt dar, dass zahlreiche Studien zeigten, dass eine Verschlechterung der FVC um mindestens zehn Prozentpunkte in einem Follow-up Zeitraum von sechs Monaten bzw. zwölf Monaten als relevanter Prädiktor für Mortalität anzusehen sei. Weiterhin stellt er dar, dass beim Eintreten einer derartigen Verschlechterung der FVC die Mortalität um bis zu einem Faktor von acht ansteige, wobei diese Zahl zwischen einzelnen Studien variere, sich der Zusammenhang jedoch als robust über unterschiedliche Untersuchungsdauern und

Mortalität	
	Patientenpopulationen erweise [8,11,42,16,18,31,37].
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u> Die forcierte Vitalkapazität (FVC) ist eine spirometrische Kenngröße der Lungenfunktion. Die FVC stellt das Atemvolumen dar, das nach maximaler Inspiration schnell und heftig (mit maximaler Geschwindigkeit) ausgeatmet werden kann. Die jährliche Abnahme der FVC wurde mit einem statistischen Modell unter der Berücksichtigung von Geschlecht, Alter und Körpergröße berechnet, um die Vergleichbarkeit zwischen den Patienten zu gewährleisten. Die Operationalisierung des Endpunkts ist nachvollziehbar.</p> <p><u>Validität und Patientenrelevanz:</u> Die Reliabilität der FVC als Prozentwert vom Sollwert mit den Kovariaten Alter, Geschlecht und Körpergröße wurde in mehr als 100 Patienten (Subgruppe einer größeren Studie mit mehr als 1000 Patienten) mit IPF untersucht. Dabei zeigten die jeweiligen wiederholten individuellen FVC-Messungen mit einem mittleren Messabstand von 18 Tagen eine starke Korrelation mit $r=0,93$ [11]. Mit fortschreitender und dem natürlichen Verlauf folgender IPF nimmt die FVC ab. Es existiert eine Vielzahl von Studien, die die Veränderung der FVC (Prozent vom Sollwert) als Prädiktor für das Überleben untersuchten. Zu den umfangreichsten Studien gehören die Arbeiten von du Bois [11,12], in denen unter mehr als 1000 Patienten mit IPF die Prädiktionskraft der FVC zum 1-Jahresüberleben untersucht wurde und konsistent als signifikanter Prädiktor des Überlebens in multivariat-adjustierten Überlebenszeitanalysen identifiziert wurde. Dabei hatten Patienten mit einer Abnahme der FVC über 24 Wochen von mehr als 10 % ein 5- bis 8-fach höheres Risiko innerhalb eines Jahres zu versterben als Patienten, die über den gleichen Zeitraum eine FVC-Abnahme von weniger als 5 % verzeichneten. Studien [35,42,32] mit kleineren Patientenpopulationen (zwischen 40 und 80 Patienten) aber teils deutlich längeren Nachbeobachtungszeiten von 2 bis 5 Jahren stützen diesen Zusammenhang zwischen der Veränderung der FVC und Mortalität. Allerdings beschränkten sich zwei Arbeiten [32,35] auf univariate Analysen, da aufgrund der limitierten Power eine Adjustierung für andere Prädiktoren wie Alter, Monooxid-Diffusionskapazität oder Schwere der Erkrankung nicht möglich war. Eine Arbeit an 84 Patienten mit IPF zeigte für Monooxid-Diffusionskapazität oder Schwere der Erkrankung adjustierte Modelle. Dabei war die Verschlechterung der FVC zw. 5-10% im Vergleich zu einem stabilen Wert der FVC über 6 Monate mit einem statistisch signifikant 2-3-fach erhöhten Risiko für Mortalität in den darauffolgenden 2 Jahren assoziiert [42].</p> <p>In den oben beschriebenen multivariaten Modellen konnte eine Korrelation zwischen FVC und der Mortalität gezeigt werden. Es liegen keine Modelle zur Validierung des Surrogats FVC vor, in denen Intervention als unabhängige Größe betrachtet wurde und der Zusammenhang zwischen Effekten auf das Surrogat Veränderung der FVC und Effekten auf den patientenrelevanten Endpunkt Mortalität gezeigt wurde. Interventionsstudien, in denen ein solcher Zusammenhang gezeigt werden konnte, liegen ebenfalls nicht vor. FVC wird als nicht ausreichend validiertes Surrogat für Mortalität angesehen.</p> <p>Die klinische Relevanzschwelle der Veränderung der FVC in Prozent vom Sollwert wurde ebenfalls von den Autoren von du Bois [11] untersucht. Die Wahrscheinlichkeit innerhalb eines Jahres zu versterben war bei Patienten, deren FVC Wert innerhalb von 24 Wochen um 5-10 % gemessen am Sollwert sank demnach um das Doppelte erhöht ($p<0,001$). Die MID beläuft sich je nach verwendeter Methode (verteilungsbasiert oder ankerbasiert) auf 2-6 %. Es gelten für die Responderanalyse die gleichen Einschränkungen hinsichtlich der Validität wie für die jährliche Abnahme der FVC.</p>

Tabelle 8: Beschreibung der Endpunkte der Kategorie Morbidität

Morbidität	
Endpunkt	Zeit bis zur ersten (adjudizierten) akuten Exazerbation
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Die Zeit bis zur ersten akuten Exazerbation war definiert als die Zeit von Randomisierung bis zur ersten akuten Exazerbation. Die Erhebung erfolgte durch den Prüfarzt gemäß präspezifizierter Definition für Exazerbationen, von Baseline bis Tag 372 (52 Wochen + 7 Tage).</p> <p>Eine akute Exazerbation war definiert als eine erhebliche akute Verschlechterung der klinischen Situation des Patienten innerhalb eines Monats, die auf keine andere Ursache als die zugrunde liegende IPF zurückgeführt werden konnte. Eine akute Exazerbation lag gemäß Studienprotokollen vor, sofern sämtliche der folgenden Kriterien erfüllt waren:</p> <ul style="list-style-type: none"> • Unerklärtes Neuauftreten oder Zunahme von Dyspnoe innerhalb der letzten 30 Tage • Neue pulmonale Infiltrate im Thorax-Röntgenbild und/oder neue parenchymale Anomalien ohne Pneumothorax oder Pleuraerguss (neue Milchglasverschattungen) in der hochauflösenden Computer-Tomographie (HRCT) seit der letzten Visite • Ausschluss von Infektionen • Ausschluss anderer möglicher Erklärungen, wie beispielsweise <ul style="list-style-type: none"> - Linksherzversagen - Lungenembolie - akute Lungenschädigung erkennbarer Ursache <p>Zu allen akuten Exazerbationen wurden klinischen Daten auf einem spezifischen Fragebogen dokumentiert. Zu den klinisch relevanten Daten gehörten:</p> <ul style="list-style-type: none"> • Eine Zusammenfassung des klinischen Verlaufs und der Anzeichen und Symptome • Ergebnisse der Laborparameter, Lungenfunktion und Bildgebung • eine Beschreibung der notwendigen Behandlung <p>Die vom Prüfarzt festgestellten Exazerbationen wurden nachträglich durch das IAC <u>adjudiziert</u>. Dazu wurden die Daten aus dem Fragebogen zu akuten Exazerbationen genutzt.</p> <p>Jedes aufgetretene Ereignis wurde durch das IAC einer der drei folgenden Kategorien zugeordnet:</p> <ul style="list-style-type: none"> • bestätigte Exazerbation • vermutete Exazerbation oder • keine Exazerbation <p>Bestätigte oder vermutete akute Exazerbationen wurden dem Endpunkt adjudizierte akute Exazerbationen als Ereignis zugeordnet. Die Exazerbation galt als bestätigt, sofern alle obengenannten Kriterien der Exazerbation erfüllt waren.</p> <p>Vermutete Exazerbation war definiert als klinische Verschlechterungen der IPF unbekannter Ursache, die nicht vollständig die oben genannten Kriterien der akuten Exazerbation der IPF erfüllten.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u></p> <p>Der pU führt aus, dass die IPF mit einer erheblichen Beeinträchtigung des</p>

Morbidität	
	<p>Gesundheitszustands einhergehe, die aus seiner Sicht im Wesentlichen auf zwei Komponenten beruhe: die alltäglich vom Patienten verspürten Symptome sowie das Erleben krisenhafter Verschlechterungen des Gesundheitszustands durch akute Exazerbationen.</p> <p>Der pU stützt seine Aussage darauf, dass akute Exazerbationen lebensbedrohliche Ereignisse seien, die bei Patienten mit IPF in jedem Erkrankungsstadium, jedoch vermehrt bei höherem Schweregrad der IPF, mit einer Inzidenz von 5-15 % pro Jahr aufträten [2,9,28,38]. Zur Ätiologie beschreibt er, dass die akute Exazerbation als akute, intrinsische Akzeleration des IPF Krankheitsfortschreitens zu bewerten sei, die klinisch unmittelbar zu einer rapiden Verschlechterung des Gesundheitszustands führe und in den meisten Fällen mit einer Hospitalisierung einhergehe [19,7]. Weiterhin stellt der pU dar, dass es aufgrund des Alveolarschadens und der Zunahme der Fibrosierung zu einer Verschlechterung der Lungenfunktionsparameter komme [19]. Er beschreibt, dass akute Exazerbationen mit einer erhöhten Mortalitätsrate assoziiert seien: In einer retrospektiven Studie lag laut pU das mediane Überleben von Patienten mit einer akuten Exazerbation bei 2,2 Monaten und die Letalität bei stationärer Behandlung bei 50% [38]. Die Sterblichkeit der betroffenen Patienten innerhalb eines Jahres nach einer Exazerbation gibt der pU mit 80%-90% an [9].</p> <p>Auf Basis dieser Daten zieht der pU das Fazit, dass Schwere und Häufigkeit akuter Exazerbationen ein wichtiges prognostisches Kriterium darstellt und ein patientenrelevantes Ereignis ist, dessen Vermeidung ein wichtiges Therapieziel sei.</p> <p>Der pU führt aus, dass es sich bei akuten Exazerbationen um relativ seltene Ereignisse handele, deren Abgrenzung zu akuten Verschlechterungen des Gesundheitszustands aufgrund anderer medizinischer Ursachen bei alleiniger Bewertung durch den Prüfarzt mit Unsicherheit behaftet sei. Daher wurden laut pU die durch die Prüfarzte erhobenen Verdachtsereignisse auf akute Exazerbation durch das verblindete IAC adjudiziert.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Die im Studienprotokoll prädefinierten Kriterien einer akuten Exazerbation entsprechen den Kriterien der S2K-Leitlinie zur Diagnostik und Therapie der idiopathischen Lungenfibrose [4]. Akute Exazerbationen lagen vor, sofern die Patienten ein Neuauftreten oder eine Verschlechterung von Dyspnoe erfuhren und sich neue Auffälligkeiten in der Bildgebung zeigten. Es ist den Unterlagen des pU nicht zu entnehmen, wie das Neuauftreten oder die Verschlechterung der Dyspnoe gemessen wurden.</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Eine erhebliche Verschlechterung der Dyspnoe kann grundsätzlich als patientenrelevant betrachtet werden, jedoch ist die Validität des Endpunkts aufgrund der unzureichenden Darstellung der Messung und Definition der Verschlechterung der Dyspnoe unklar.</p> <p>Die vom Prüfarzt festgestellten Exazerbationen wurden nachträglich durch das IAC adjudiziert und entweder wurde die Diagnose verworfen, bestätigt oder das Ereignis als vermutete Exazerbation eingestuft. Dies lässt vermuten, dass offenbar nicht alle Kriterien der akuten Exazerbation erfüllt sein mussten, damit ein vom Prüfarzt dokumentiertes Ereignis der akuten Exazerbation vorlag. Für die vom IAC adjudizierten akuten Exazerbationen mussten ebenfalls nicht alle Kriterien der akuten Exazerbation erfüllt gewesen sein, da auch vermutete Exazerbationen in den Endpunkt eingingen. Die Validität und Patientenrelevanz der Endpunkte akute und akute adjudizierte Exazerbation ist aus diesem Grunde unklar. Eine alleinige Veränderung der Bildgebung, die auf keine andere Ursache als die zugrunde liegende IPF zurückgeführt werden kann, wird ohne Nachweis einer Auswirkung auf</p>

Morbidität	die Morbidität des Patienten nicht als patientenrelevant betrachtet.
Endpunkt	PGI-C-Responder
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u> Mit Hilfe der Patient's Global Impression of Change (PGI-C) gibt der Patient seinen subjektiven Eindruck zur Verbesserung oder Verschlechterung seiner respiratorischen Kondition im Vergleich zum Referenzzeitpunkt bei Studienbeginn an. Das Instrument besteht aus einer Frage: "Check the one number that best describes how your respiratory condition is now, compared with how it was before you began taking medication in this study." Die Einschätzung erfolgte durch die Studienteilnehmer auf einer 7-Punkte-Likert-Skala von eins (sehr viel besser) über vier (unverändert) bis sieben (sehr viel schlechter) von Baseline bis Woche 52. Als PGI-C-Responder wurde a priori ein Patient definiert, der bei der Erhebung zu Woche 52 verglichen mit seinem subjektiv erinnerten Befinden bei Studienbeginn keine Kategorie der Verschlechterung seines Zustands angegeben hatte.</p> <p>Die Bestimmung des PGI-C erfolgte zu Woche 12, 24 und 52. Falls zu Woche 52 kein Messwert vorlag, wurde der entsprechende Patient konservativ als Non-Responder bewertet. Der Endpunkt Anteil PGI-C-Responder wurde mittels logistischer Regression ausgewertet.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Der pU begründet die Patientenrelevanz des Endpunktes PGI-C-Responder damit, dass bei einer chronisch progredienten Erkrankung wie der IPF ohne eine adäquate Therapie von einer stetigen Verschlechterung über die Zeit auszugehen sei. Weiterhin führt der pU aus, dass ein Ausbleiben einer solchen subjektiven Verschlechterung einen Therapieerfolg und somit einen patientenrelevanten Nutzen darstelle.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u> Die Operationalisierung des Endpunkts ist nachvollziehbar.</p> <p><u>Validität und Patientenrelevanz:</u> Der pU macht keine Angaben zur Validität des Messinstrumentes. Generell ist das Instrument nachvollziehbar, jedoch ist eine Einschätzung der Krankheitsschwere mit einem Zeithorizont von 52 Wochen in die Vergangenheit als nicht robust einzuschätzen. Das Responderkriterium „keine Verschlechterung“ ist nicht validiert.</p>
Endpunkt	Veränderung EQ-5D-VAS
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u> Die Veränderung des EQ-5D und des Wertes auf der visuellen Analogskala des EQ-5D (EQ-5D VAS) zwischen Studienbeginn und Woche 52 wurden bestimmt. Der pU stellt im Dossier lediglich die Veränderung der EQ-5D-VAS vom Baseline bis Woche 52 dar. Eine statistische Auswertung wurde nicht durchgeführt. Auf dem EQ-5D VAS kann der Patient seinen globalen Gesundheitszustand zwischen 0 (schlecht möglicher Gesundheitszustand) und 100 (bestmöglicher Gesundheitszustand) angeben.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u> Nach Angaben des pU stellt der EQ-5D ein generisches Fragebogeninstrument zur Erfassung der gesundheitsbezogenen Lebensqualität dar.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u> Die Operationalisierung des Endpunkts ist nachvollziehbar.</p>

Morbidität	
	<p><u>Validität und Patientenrelevanz:</u></p> <p>Der pU zieht die EQ-5D-VAS zur Bewertung der Veränderung des Gesundheitszustands des Patienten heran. Die EQ-5D-VAS wird als valide eingestuft und wird für die vorliegende Nutzenbewertung herangezogen. Es liegt keine Responderanalyse für den EQ-5D VAS vor, so dass die klinische Relevanz der Ergebnisse nicht ausreichend eingeschätzt werden kann. Da der EQ-5D VAS den allgemeinen Gesundheitszustand abbildet, wird er der Endpunktkategorie Morbidität zugeordnet.</p> <p>Der EQ-5D ist ein valider, generischer Fragebogen bestehend aus 5 Fragen, anhand derer ein Gesundheitszustand erfasst wird. Die 5 Fragen beinhalten auch Aspekte der gesundheitsbezogenen Lebensqualität. Die ermittelten Gesundheitszustände werden durch eine Gewichtung in einen Nutzwert transformiert. Wenn die in der Studie ermittelten Nutzwerte auf einer Bewertung der Gesundheitszustände durch Patienten in der jeweiligen Indikation beruhen, dann eignet sich der EQ-5D zur Bestimmung eines patientenrelevanten Zusatznutzens. Gegebenenfalls kann die Bewertung der Gesundheitszustände durch Patienten in einer vergleichbaren Indikation erfolgen. Die Übertragbarkeit einer solchen Patientengruppe muss jedoch hinreichend begründet sein. Sofern eine Bewertung von Gesundheitszuständen durch Patienten mit der jeweiligen (oder ggf. einer geeigneten übertragbaren) Erkrankung nicht vorliegt, ist der EQ-5D für eine Bewertung eines Zusatznutzens in dieser Indikation nicht geeignet. Da der Nutzwert nicht anhand einer geeigneten Patientengruppe festgelegt und kein minimal klinisch relevanter Unterschied bestimmt wurde, wird dem Vorgehen des pU gefolgt und die Ergebnisse des EQ-5D werden nicht zur Bewertung des Zusatznutzens herangezogen.</p>

Tabelle 9: Beschreibung der Endpunkte der Kategorie Lebensqualität

Lebensqualität	
Endpunkt	Veränderung SGRQ-Gesamtscore
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Operationalisiert war die Veränderung des SGRQ-Gesamtscores als Veränderung über die Zeit von Baseline bis Woche 52. Der Gesamtscore setzt sich zusammen aus den drei Domänen Symptome (Leiden aufgrund respiratorischer Symptome), Aktivität (Beeinträchtigung der Mobilität und körperlichen Aktivität) und Impact (Auswirkungen psychosozialer Einflüsse der Erkrankung auf das tägliche Leben und das Wohlbefinden). Aus den Antworten wird ein Grad der Beeinträchtigung der krankheitsspezifischen Lebensqualität in einem Bereich von 0 (keine Beeinträchtigung/gute Lebensqualität) bis 100 (maximale Beeinträchtigung/schlechte Lebensqualität) ermittelt.</p> <p>Als SGRQ-(≤ -4 Punkte)-Responder wurde a priori ein Patient definiert, dessen SGRQ-Gesamtscore in Woche 52 gegenüber dem individuellen Messwert zu Studienbeginn um mindestens 4 Punkte abfiel. Falls zu Woche 52 kein Messwert vorlag, wurde der entsprechende Patient konservativ als Non-Responder bewertet.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u></p> <p>Der pU gibt an, dass es sich beim SGRQ um ein mehrdimensionales Erhebungsinstrument zur Dokumentation der krankheitsspezifischen Lebensqualität speziell für Patienten mit einer respiratorischen Erkrankung handele, wobei dieser validierte Fragebogen ursprünglich für die chronisch obstruktive Lungenerkrankung (COPD) entwickelt worden sei [25].</p>

Lebensqualität	
	Der pU stellt dar, dass eine Verminderung des SGRQ-Gesamtscores um vier Punkte im Vergleich zum Ausgangswert allgemein als MID der subjektiv empfundenen Lebensqualität akzeptiert würde [22,23,39].
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Das mehrdimensionale Erhebungsinstrument St. George's Respiratory Questionnaire (SGRQ) wurde zur Abschätzung der Beeinträchtigung der gesundheitsbezogenen Lebensqualität bei erwachsenen Patienten mit chronischen Erkrankungen der Atemwege entwickelt und umfasst 50 Fragen [24]. Neben der englischen Originalversion liegen validierte Übersetzungen des SGRQ in Französisch, Spanisch, Schwedisch, US-amerikanisch und Hong-Kong-chinesisch vor. Die deutschsprachige validierte Version steht seit 1998 zur Verfügung [26]. Für die einzelnen Fragen bestehen zwischen 2 und 5 Antwortmöglichkeiten. Die Antworten werden gewichtet und es werden drei Subskalen gebildet: „Symptome“ (Frequenz und Ausprägung klinischer Symptome), „Aktivität“ (Beeinträchtigungen in Alltag und Funktion) und „Belastung“ (Einstellung gegenüber der Erkrankung, subjektive Belastung bei Alltagsaktivitäten). Zusätzlich wird ein SGRQ-Gesamtwert ermittelt. Um die Vergleichbarkeit zu gewährleisten, werden die Werte aller Skalen auf eine 0–100-Skala transformiert, wobei höhere Werte einer stärkeren Einschränkung der Lebensqualität entsprechen.</p> <p>Sofern keine der 3 Subskalen als fehlend gewertet wurden, kann die Bildung des Gesamtscores erfolgen. In der Studie von [26] wurde der Schwellenwert des SGRQ an Asthma- und COPD-Patienten und die Änderungssensitivität auf der Basis von 777 Patientendaten untersucht. Insgesamt lagen dieser Analyse die Daten von 544 Patienten die an Asthma und 233 Patienten die an COPD erkrankt waren zugrunde. Die Reliabilitätschätzung nach Cronbach liegt bei der Gesamtskala in einem Bereich von $\alpha = 0,762$ bis $\alpha = 0,847$. Die interne Homogenität und Reliabilität der Skalen des SGRQ konnte bei allen 3 Skalen positiv bewertet werden. Sowohl bei der Gesamtstichprobe als auch bezüglich der unterschiedlichen Diagnosegruppe konnten vergleichbare Indizes der Reliabilität erzielt werden. Anhand der Ergebnisse der Varianzanalyse kann eine Veränderungs sensitivität des SGRQ bei der Erfassung der Beeinträchtigung der krankheitsspezifischen Lebensqualität bei Asthma- und COPD-Patienten abgeleitet werden. Allerdings zeigte sich bei der Skala „Symptome“ in beiden Diagnosegruppen keine signifikante Veränderung. Gemäß der Publikationen von Jones [20,21] beträgt die minimally important difference (MID) bei Asthmapatienten 4 Punkte.</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Beim SGRQ handelt es sich um ein gut untersuchtes Instrument zur Messung der Lebensqualität bei Patienten mit chronischen Erkrankungen der Lunge. Es ist davon auszugehen, dass eine Veränderung um 4 Punkte oder mehr ein klinische und somit relevante Veränderung der gesundheitsbezogenen Lebensqualität des Patienten darstellt. Allerdings liegen keine Daten zur Eignung des Instruments bei Patienten mit IPF vor.</p>
Endpunkt	Veränderung SGRQ-I-Gesamtscore
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Die Veränderung des SGRQ-I Gesamtscores war definiert als Veränderung über die Zeit von Baseline bis Woche 52. Die Erhebung erfolgte durch Studienteilnehmer. Beim SGRQ-I Fragebogen handelt sich um eine IPF-spezifische SGRQ-Fragebogenversion. Der SGRQ-I besteht aus den drei Domänen Symptome, Aktivität und Impact. Aus den Antworten wird – analog dem SGRQ – ein Grad der Beeinträchtigung der krankheitsspezifischen Lebensqualität in einem Bereich von 0 (keine Beeinträchtigung / gute Lebensqualität) bis 100 (maximale Beeinträchtigung /</p>

Lebensqualität	
	<p>schlechte Lebensqualität) ermittelt. Der SGRQ-I-Gesamtscore wurde, wie im Studienprotokoll präspezifiziert, aus den Antworten im SGRQ-Fragebogen nach einem von den Entwicklern des SGRQ-I definierten Algorithmus ermittelt.</p> <p>Für die Indikation IPF gibt es derzeit noch keine allgemein akzeptierte MID. Um eine Responder-Analyse durchführen zu können, wurde in Anlehnung an die MID für einen klinisch relevanten Effekt bzgl. des SGRQ-Gesamtscores auch für den SGRQ-I ein Schwellenwert von -4 Punkten festgelegt. Die Responderanalyse des SGRQ-I erfolgte post-hoc.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u></p> <p>Der pU führt aus, dass mit dem SGRQ-I eine IPF-spezifische Form des SGRQ entwickelt wurde, dessen Ziel es sei, die spezifischen Einschränkungen der HRQoL durch die IPF abzubilden [41]. Dieser für IPF-Patienten angepasste Fragebogen wurde laut pU hinsichtlich seiner Reliabilität im Vergleich zum SGRQ und im Hinblick auf seine Validität mit Messungen der Atemphysiologie, der funktionalen Kapazität, der Dyspnoe und des Gesundheitsstatus evaluiert [41,39].</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Der IPF-spezifische SGRQ basiert auf dem oben beschriebenen und validierten SGRQ. Der ursprünglich 50 Items umfassende SGRQ wurde um 6 Items reduziert, die in einer IPF-Patientenpopulation von n=158 mehr als 25% der Patienten nicht beantwortet wurden [41]. In einem zweiten Schritt erfolgte die weitere Reduktion mittels einer Rasch-Analyse, bei durch einen iterativen Prozess durch Weglassen des Items mit der schlechtesten Anpassung die Anpassungsgüte der Domänen verbessert wird. Durch dieses Verfahren wurden 10 weitere Items entfernt, so dass der IPF-spezifische SGRQ (SGRQ-I) aus 34 relevanten Items in den bekannten Domänen Symptome (n=6 Items), Aktivität (n=10 Items) und Impact (n= 18 Items) besteht. Darüber hinaus wurden im Rasch-Modell die Antwortmöglichkeiten adaptiert, was bei einigen Fragen zu einer Reduktion der Antwortkategorien führte. Der pU berichtete nicht die Ergebnisse aus diesem 26 Items umfassenden SGRQ-I Fragebogen, sondern wandte einen von den Entwicklern des SGRQ-I entwickelten Algorithmus an, der die Berechnung des SGRQ-I Scores aus dem SGRQ-Score ermöglicht. Zur Ermittlung dieses Algorithmus wurden beide Instrumente, der ursprüngliche SGRQ und der neu entwickelte SGRQ-I-Fragebogen, in 158 IPF-Patienten eingesetzt. Beide Scores (SGRQ und SGRQ-I) wurden gegeneinander aufgetragen und eine Transformationsgleichung mittels linearer Regression entwickelt. Dieser Algorithmus (Transformationsgleichung) zeigte eine hohe Modellgüte, ausgedrückt als Maß der erklärten Variation des SGRQ-I Scores, für alle Einzeldomänen als auch für den Gesamtscore.</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Die Lebensqualität ist ein patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittelnutzenbewertungsverordnung (AM-NutzenV). Der SGRQ-I wurde in der relevanten Patientenpopulation mittels Rasch-Analyse aus dem SGRQ entwickelt [41]. Das Instrument zeigt eine gute interne Konsistenz mit Cronbachs alpha für die Domänen Symptome $\alpha=0,62$, Aktivität $\alpha=0,80$ und Impact $\alpha=0,85$. Die Überführung des generischen SGRQ-Scores in den SGRQ-I ist mittels linearer Regression ermittelter Gleichungen möglich. Allerdings liegen keine weiterführenden vergleichenden Untersuchungen zum SGRQ-I Score, ermittelt anhand des neuen Fragebogens und des SGRQ-I Scores berechnet mittels Transformationsgleichung aus SGRQ-Score, vor. Die Konstruktvalidität des SGRQ-I konnte durch eine Reihe von signifikanten Korrelationen zu Lungenfunktionsmessungen und anderen PRO-Instrumenten demonstriert werden. Außerdem liegen keine Informationen zur Test-Retest-Reliabilität, Veränderungssensitivität und zur klinischen Relevanz vor. Es ist fraglich, ob die</p>

Lebensqualität	
	<p>Übertragung der MID vom SGRQ auf den SGRQ-I valide ist.</p> <p>Zusammenfassend ist die Validität des transformierten SGRQ-I Scores aus dem ursprünglichen SGRQ-Scores kritisch zu betrachten, da nicht bekannt ist, inwiefern Aussagen zum SGRQ-I Score ermittelt durch den IPF-spezifischen Fragebogen auf die transformierten SGRQ-I Score übertragbar sind.</p>
Endpunkt	Veränderung SOBQ-Gesamtscore
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Die Veränderung des SOBQ-Gesamtscores war definiert als die Veränderung von Baseline bis Woche 52. Der SOBQ Fragebogen erfasst die Kurzatmigkeit der Befragten an einem durchschnittlichen Tag innerhalb der letzten Woche bei verschiedenen Tätigkeiten (z.B. beim Laufen, in Ruhe, beim Zähneputzen, beim Abwaschen etc.). Der SOBQ-Gesamtscore setzt sich zusammen aus 24 Items, die jeweils auf einer Skala von null bis fünf bewertet werden. Die Summe ergibt den SOBQ-Gesamtscore (Wertebereich 0-120), wobei ein niedrigerer Score eine geringere Beeinträchtigung der Lebensqualität durch Kurzatmigkeit, ein hoher Gesamtscore hingegen eine schwere Beeinträchtigung durch Atemnot reflektiert.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u></p> <p>Laut der Angaben des pU handelt es sich beim SOBQ um ein validiertes Fragebogeninstrument, das die Selbsteinschätzung der Beeinträchtigung der Aktivitäten des täglichen Lebens durch Dyspnoe widerspiegelt [1]. Eine kürzlich veröffentlichte Arbeit unterstütze laut pU die Validität dieses ursprünglich für COPD-Patienten entwickelten Messinstruments für IPF-Patienten [17].</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Die Operationalisierung des Endpunkts ist nachvollziehbar.</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Bei dem SOBQ handelt es sich um ein gut untersuchtes und valides Messinstrument bei Patienten mit chronischen Lungenerkrankungen [13,14,30]. Die Eignung des Instruments wurde zudem in Patienten mit IPF untersucht. Die Konstruktvalidität, die Verständlichkeit des Instruments und die Relevanz der Items des SOBQ können als geeignet bei Patienten mit IPF eingeschätzt werden [17,40].</p>
Endpunkt	Veränderung CASA-Q Husten-Symptome und Husten-Impact
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Die Veränderung des Fragebogens CASA-Q war definiert als Veränderung vom Baseline bis Woche 52. Der CASA-Q besteht ursprünglich aus 20 Items in 4 Domänen: jeweils 3 in den Symptomdomänen Husten und Sputum und 8 und 6 in den Impact-Domänen Husten und Sputum. In den INPULSIS Studien wurden ausschließlich die 11 Fragen der beiden Husten-Domänen (Husten-Symptome und Husten-Impact) berücksichtigt. Die Patienten werden dazu angehalten, auf einer 5-Punkte Skala die Häufigkeit und Schwere, so wie die Beeinträchtigung durch Husten innerhalb der letzten 7 Tage zwischen „gar nicht“ und „sehr stark“ zu bewerten. Die Domänenscores des CASA-Q werden jeweils so umgerechnet, dass die Ergebnisse der Domänen auf einer 100-Punkte-Skala liegen, wobei ein geringerer Wert eine stärkere Intensität bzw. Beeinträchtigung durch die genannten Symptome anzeigt.</p> <p><u>Begründung des pU gemäß der im Dossier vorhandenen Angaben:</u></p> <p>Der pU beschreibt, dass der CASA-Q ein in der Indikation COPD und chronische Bronchitis validiertes Fragebogeninstrument sei und der Evaluierung der Intensität der klinischen Symptome Husten und Sputum (Husten-Symptome, Sputum-</p>

Lebensqualität	
	<p>Symptome) sowie deren Auswirkungen auf die HRQoL (Husten-Impact, Sputum-Impact) diene [10]. Die Validität dieses ursprünglich für COPD-Patienten entwickelten Messinstruments sei laut pU kürzlich für IPF-Patienten bestätigt worden [17]. Die ausschließliche Erhebung der beiden Husten-Domänen (Husten-Symptome und Husten-Impact) des CASA-Q begründet der pU damit, dass Sputum als Symptom bei der IPF – im Gegensatz zur COPD – eine untergeordnete Rolle spiele.</p>
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Der Cough and Sputum Questionnaire (CASA-Q) ist ein Instrument zur Erfassung der Beeinträchtigung der Aktivitäten des täglichen Lebens durch Husten und Sputum bei chronischen inflammatorischen Atemwegserkrankungen. Das Instrument wurde anhand von Patienten mit COPD oder chronischer Bronchitis in mehreren Sprachen/Ländern entwickelt und in Deutschland, Frankreich und den USA validiert [10]. Die Inhalts- und Konstruktvalidität bzw. die Reliabilität ist in der Patientenpopulation mit COPD und chronischer Bronchitis untersucht und zeigt hinsichtlich der Korrelation mit anderen objektiven Messungen (Hustenmonitoring via Mikrofon, Plethysmographie) sowie anderer PRO-Instrumente geringe bis moderate Korrelationen. Es fehlen Angaben zur Veränderungssensitivität und klinischen Relevanz.</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Der CASA-Q verfügt über vier Domänen, wobei in den vorliegenden Studien nur die Domänen zur Hustensymptomatik und Husten-Impact herangezogen wurden. Die Reliabilität der Domäne der Hustensymptomatik wurde in Patienten mit Lungenerkrankungen bestätigt [10,17] (Cronbachs alpha 0,85; Intraclass-Korrelation: 0,74-0,77). Eine Validierung durch Plethysmographie und akustischer Aufzeichnung der Hustenfrequenz zeigte innerhalb eines kleinen Patientenkollektivs mit COPD und chronischer Bronchitis (n=32) keine nennenswerte Korrelation zu den CASA-Q Hustendomänen ($r < 0,22$).</p> <p>Eine Untersuchung innerhalb einer kleinen Patientenpopulation mit IPF zeigte eine gute Verständlichkeit und Relevanz der abgefragten Elemente der Hustendomänen. Der Bezugszeitraum als auch die Antwortoptionen konnten ebenfalls von einem Großteil der IPF-Patienten korrekt wiedergegeben werden [17].</p> <p>Es liegen kaum Informationen zur Validität des CASA-Q Instruments in Patienten mit IPF vor. Der Einsatz des Instruments in den Studien und die daraus resultierenden Ergebnisse sind daher kritisch zu bewerten.</p>

Tabelle 10: Beschreibung der Endpunkte der Kategorie Sicherheit

Sicherheit	
Endpunkt	Gesamtzahl Unerwünschte Ereignisse
Operationalisierung und Validierung: Begründung seitens des pU	<p><u>Operationalisierung:</u></p> <p>Ein UE war definiert als jegliches unerwünschtes medizinisches Vorkommnis (einschließlich einer Verschlechterung des bestehenden Krankheitszustandes) bei einem Studienteilnehmer, der die Studienmedikation erhielt. Dabei musste das Ereignis nicht unbedingt in einem kausalen Zusammenhang mit dem angewandten Arzneimittel stehen. Ereignisse wurden von Gabe der ersten Dosis der Studienmedikation bis 28 Tage nach der letzten Einnahme der Studienmedikation erfasst.</p> <p>UEs wurden laut Protokoll entsprechend MedDRA nach Systemorganklassen</p>

Sicherheit	
	<p>(SOC) und bevorzugter Bezeichnung (PT) kodiert.</p> <p>Ein SUE war definiert als jegliches UE, welches entweder zum Tode oder zu einer unmittelbar lebensbedrohlichen Situation führte, eine dauerhafte oder erhebliche Behinderung / Invalidität hervorrief, einen Krankenhausaufenthalt erforderlich machte oder verlängerte, eine angeborene Anomalie / einen Geburtsdefekt darstellte oder aus einem anderen Grund für schwerwiegend erachtet wurde oder einen medizinischen oder chirurgischen Eingriff erforderlich machte, um eine der o.g. Konsequenzen zu verhindern. Ereignisse wurden von Gabe der ersten Dosis der Studienmedikation bis 28 Tage nach der letzten Einnahme der Studienmedikation erfasst.</p> <p>Zusätzlich wurden in den Studien UE berichtet, die zum Therapieabbruch führten (keine Re-Exposition in voller oder reduzierter Dosierung), sowie UE, die zum Tod führten (für diese wurden Ereignisse auch dann berücksichtigt, wenn der Tod länger als 28 Tage nach der letzten Medikamenteneinnahme eintrat, der Beginn des als ursächlich angesehenen UE jedoch während der Medikamenteneinnahme oder innerhalb von 28 Tagen nach der letzten Medikamenteneinnahme lag).</p> <p>UE von besonderem Interesse waren entsprechend der Studienprotokolle: Diarrhoe, Übelkeit, Erbrechen, Dehydrierung, Gewichtsverlust und Appetitverlust. Zusätzlich definierte der pU im Dossier folgende UE als UE von besonderem Interesse:</p> <ul style="list-style-type: none"> • gastrointestinales Perforation • Leber- und Gallenerkrankungen (Leberenzymhöhung, Bilirubinhöhung), • kardiale Ereignisse (darunter Myokardinfarkt), • respiratorische Ereignisse (Progression der IPF, Husten, Dyspnoe, Bronchitis, Nasopharyngitis), • Blutungsereignisse • arterielle thromboembolische Ereignisse • alle weiteren Störungen, die in mindestens einem Studienarm mit einer Häufigkeit von $\geq 10\%$ der Studienteilnehmer auftraten
Einschätzung zur Qualität u. Patientenrelevanz	<p><u>Definition und Erhebung des Endpunkts:</u></p> <p>Die Operationalisierung der Sicherheitsendpunkte ist nachvollziehbar.</p> <p>Die im Dossier des pU definierten und dargestellten UE von besonderem Interesse weichen teilweise von den im statistischen Analyseplan und Studienbericht aufgeführten Ereignissen von besonderem Interesse ab. Vor allem die respiratorischen Ereignisse, darunter im Besonderen Progression der IPF, unterliegen zum Teil einer Doppelerfassung, da dieses Ereignis auch Teil von Wirksamkeitsendpunkten ist (Exazerbation der IPF).</p> <p><u>Validität und Patientenrelevanz:</u></p> <p>Unerwünschte Ereignisse sind ein patientenrelevanter Endpunkt entsprechend § 2 Satz 3 der Arzneimittelnutzenbewertungsverordnung (AM-NutzenV).</p>

2.5.4 Statistische Methoden

Statistische Modelle und Tests

Ziel der INPULSIS-Studien war die Untersuchung der Überlegenheit von Nintedanib 150mg zweimal täglich im Vergleich zu Placebo bezogen auf die Rate der jährlichen Abnahme der FVC bei Patienten mit IPF. In beiden Studien wurden die gleichen Endpunkte auf gleiche Weise ausgewertet. Die Null-Hypothese lautete: kein Unterschied bei der jährlichen Abnahme der FVC gemessen von Baseline bis Woche 52 zwischen Nintedanib und Placebo. Die Alternativhypothese lautete: geringere jährlichen Abnahme der FVC gemessen von Baseline bis Woche 52 bei Patienten in der Nintedanib-Gruppe im Vergleich zur Placebo-Gruppe.

Der primäre Endpunkt der beiden INPULSIS-Studien war die Abnahme der FVC von Baseline bis Woche 52, ausgewertet mittels Random-Coefficient-Modell mit Geschlecht, Alter und Größe als Kovariate (Term). Die zwei zentralen sekundären Endpunkte waren Zeit bis zur ersten akuten Exazerbation und Veränderung vom Ausgangswert im SGRQ Gesamtscore nach 52 Wochen. Die Zeit bis zur ersten akuten Exazerbation der IPF wurde unter Verwendung des Cox Proportional-Hazards-Modells, adjustiert für Geschlecht, Größe und Alter, und Log-Rank-Test analysiert. Die Veränderung vom Ausgangswert im SGRQ Gesamtscore nach 52 Wochen wurde mittels Mixed-Effects-Modell für wiederholte Messungen (MMRM) ausgewertet, wobei Behandlung und Visite als fixe Effekte eingingen und der Baseline-Gesamtscore als Kovariate.

Es wurde eine hierarchische Testprozedur gewählt um die Überlegenheit von Nintedanib über Placebo in Hinblick auf den primären und die beiden zentralen sekundären Endpunkte zu testen. Die nachfolgenden Schritte in der Testhierarchie wurden nur dann durchgeführt, sofern sich im vorherigen Schritt eine statistische Signifikanz auf einem einseitigen Signifikanzniveau von 2,5 % zu Gunsten von Nintedanib zeigte. In der für Europa geltenden Testhierarchie wurde zunächst der primäre Endpunkt, dann die Veränderung vom Ausgangswert im SGRQ Gesamtscore nach 52 Wochen und dann die Zeit bis zur ersten akuten Exazerbation untersucht. In der für Amerika geltenden Testhierarchie wurden die beiden zentralen sekundären Endpunkte in umgekehrter Reihenfolge (zunächst Exazerbation, dann SGRQ) getestet.

Alle kontinuierlichen sekundären Endpunkte (PROs) wurden über 52 Wochen mit MMRM analog zur Auswertung für den SGRQ analysiert. PGI-C-Responder nach 52 Wochen und SGRQ Responder nach 52 Wochen wurden mit einer logistischen Regression analysiert, mit Behandlung und Baseline als Kovariate und FVC Responder mit Behandlung, Alter, Geschlecht, Größe und Baselinewert als Kovariate. Odds Ratio p-Werte basierend auf einem 95 % Konfidenzintervall wurden berechnet. Für die Überlebenszeitanalysen (time-to-event-Analysen) wurde das Cox Proportional-Hazards-Modell angewendet, adjustiert für Geschlecht, Größe und Alter. Patienten, für die bis Tag 372 kein Ereignis dokumentiert wurde, wurden zu Tag 373 oder zum Tag der letzten Visite zensiert. Patienten, die 28 Tage nach der geplanten letzten Einnahme der Studienmedikation am Leben waren, wurden am Tag 28 nach Einnahme der letzten Studienmedikation zensiert.

Alle Wirksamkeitsanalysen wurden innerhalb der ersten 52 Wochen durchgeführt (372 Tage: 365 Tage + 7 Tage-Toleranzzeitraum). Nach Lungentransplantation wurde keine weitere Analyse der Wirksamkeit vorgenommen. Time-to-Event-Endpunkte der Wirksamkeit wurden ab der Randomisierung berechnet. Time-to-Event-Analysen der Sicherheit wurden ab dem Zeitpunkt der ersten Einnahme der Studienmedikation berechnet.

Im Dossier wurden Anzahl und Anteil der Studienteilnehmer mit einem UE der betreffenden Kategorie tabellarisch dargestellt. Der Gruppenunterschied wurde post hoc als Risk Ratio (RR), OR und absolute Risikodifferenz (RD), jeweils mit 95 %-Konfidenzintervall und p-Wert, basierend auf einem Chi-Quadrat-Test, angegeben. In der Nutzenbewertung werden ausschließlich OR herangezogen.

Auswertungspopulation

In den INPULSIS-Studien wurden Wirksamkeit und Sicherheit auf Basis des Treated Set (TS) ausgewertet. Das TS bestand aus den Studienteilnehmern, für die die Einnahme von mindestens einer Dosis der Studienmedikation dokumentiert wurde. In INPULSIS-1 hatten 2 von insgesamt 515 randomisierten Studienteilnehmern keine Dosis der Studienmedikation erhalten, in INPULSIS-2 waren dies 3 von 551 Patienten.

Sensitivitätsanalysen

Es wurden verschiedene Sensitivitätsanalysen für den primären und für die zentralen sekundären Endpunkte durchgeführt:

- Testen der Linearitätsannahme für die Abnahme des FVC und Sensitivitätsanalysen mit alternativen Modellen (quadratisch, exponentiell).
- Primäres Modell mit on-treatment Evaluation (bis zur letzten Einnahme der Studienmedikation + 1 Tag).
- Primäres Modell einschließlich der Daten nach Lungentransplantation
- Multiple Imputationsansätze für Patienten, für die keine Werte zu Woche 52 vorliegen.

Imputation fehlender Daten

Die primäre Analyse des primären Endpunkts und der zentralen sekundären Endpunkte basierte auf den beobachteten Ereignissen ohne Imputation fehlender Werte. In den Responderanalysen zu SGRQ, FVC und PGI-C wurden Studienteilnehmer, für die zu Woche 52 keine Messwerte vorlagen, konservativ als Non-Responder bewertet. Für den SOBQ und den EQ-5D wurden keine Imputationen vorgenommen und die Scores wurden als fehlend angesehen, wenn nicht alle Fragen beantwortet waren. Die CASA-Q Symptom- und Impact-Domänen umfassen jeweils 3 und 8 Items. Im CASA-Q-Benutzerhandbuch ist definiert, dass mindestens 2 bzw. 4 Items beantwortet werden müssen. Im Falle fehlender Items, wurden diese durch die durchschnittliche Punktzahl beantworteter Items in der jeweiligen Skala ersetzt, sofern die minimale Anzahl an Items erreicht war.

Subgruppenanalysen

Subgruppenanalysen wurden für die in Tabelle 1 genannten Subgruppen durchgeführt. P-Werte für Interaktionen wurden berechnet, wobei ein $p < 0,2$ ein Hinweis für eine Effektmodifikation und ein $p < 0,05$ einen Beleg für eine Modifikation des Behandlungseffekts determiniert. Ergebnisse zu Subgruppenanalysen wurden nur dargestellt, sofern sich mindestens ein Hinweis auf Effektmodifikation ergab.

Metaanalysen

Die individuellen Patientendaten (IPD) der beiden INPULSIS-Studien wurden für das Dossier gepoolt und als Metaanalyse quantitativ ausgewertet. Der pU verzichtete dabei auf die Darstellung von Forest-Plots, sondern berechnete die statistischen Modelle wie in den Einzelstudien und stellte die dazugehörigen Effektmaße dar. Für die Metaanalysen wurden Studien als Faktor in die Modelle aufgenommen. Der pU gibt an, dass aufgrund der klinischen und statistischen Homogenität der INPULSIS-Studien ein Fixed-Effect Modell gewählt wurde.

Im Dossier und in der vorliegenden Nutzenbewertung werden sowohl die Ergebnisse der INPULSIS-Einzelstudien dargestellt, als auch die Ergebnisse der Metaanalyse.

2.6 Ergebnisse zum Zusatznutzen

Allgemeine Angaben

Tabelle 11: Allgemeine Angaben in den Studien INPULSIS-1 und INPULSIS-2

	Nintedanib	Placebo
Randomisiert, <i>INPULSIS-1</i>	309	206
<i>INPULSIS-2</i>	331	220
<i>Gesamt</i>	640	426
Tatsächlich die geplante Behandlung erhalten (Treated Set) ¹⁾		
<i>INPULSIS-1</i>	309	204
<i>INPULSIS-2</i>	329	219
<i>Gesamt</i>	638	423
Abbruch der Einnahme der Studienmedikation, <i>n</i> (%)		
<i>INPULSIS-1</i>	78 (25,2)	36 (17,6)
<i>INPULSIS-2</i>	78 (23,7)	44 (20,1)
<i>Gesamt</i>	156 (24,5)	80 (18,9)
aufgrund von UE, <i>n</i> (%)		
<i>INPULSIS-1</i>	65 (21,0)	24 (11,8)
<i>INPULSIS-2</i>	62 (18,8)	35 (16,0)
unerwartete Verschlechterung der IPF, <i>n</i> (%)		
<i>INPULSIS-1</i>	9 (2,9)	9 (4,4)
<i>INPULSIS-2</i>	10 (3,0)	18 (8,2)
unerwartete Verschlechterung einer anderen bereits vorhandenen Krankheit, <i>n</i> (%)		
<i>INPULSIS-1</i>	2 (0,6)	2 (1,0)
<i>INPULSIS-2</i>	3 (0,9)	2 (0,9)
andere unerwünschte Ereignisse, <i>n</i> (%)		
<i>INPULSIS-1</i>	54 (17,5)	13 (6,4)
<i>INPULSIS-2</i>	49 (14,9)	15 (6,8)
Lost to Follow-up, <i>n</i> (%)		
<i>INPULSIS-1</i>	0	0
<i>INPULSIS-2</i>	0	1 (0,5)
Verweigerung der Medikation, <i>n</i> (%)		
<i>INPULSIS-1</i>	9 (2,9)	7 (3,4)
<i>INPULSIS-2</i>	11(3,3)	6 (2,7)
Protokollverletzung, <i>n</i> (%)		
<i>INPULSIS-1</i>	2 (0,6)	3 (1,5)
<i>INPULSIS 2</i>	2 (0,6)	1 (0,5)
Andere, <i>n</i> (%)		
<i>INPULSIS-1</i>	2 (0,6)	2 (1,0)
<i>INPULSIS 2</i>	3 (0,9)	1 (0,5)
Mediane Behandlungsdauer, Monate (<i>min</i> ; <i>max</i>)		
<i>INPULSIS-1</i>	11,93 (0,0;12,5)	11,93 (0,5;13,0)
<i>INPULSIS-2</i>	11,93 (0,0;12,7)	11,93 (0,0;13,1)

	Nintedanib	Placebo
Expositionsdauer zur tatsächlich erhaltenen Dosierung		
Dauer der Behandlung mit 150mg Nintedanib 2-mal täglich, <i>N</i> (Monate [SD])		
<i>INPULSIS-1</i>	309 (8,91 [4,09])	204 (10,62 [3,02])
<i>INPULSIS-2</i>	329 (8,82 [4,14])	219 (10,63 [2,93])
Dauer der Behandlung mit 100mg Nintedanib 2-mal täglich, <i>N</i> (Monate [SD])		
<i>INPULSIS-1</i>	82 (4,28 [3,34])	10 (3,65 [3,86])
<i>INPULSIS-2</i>	96 (4,47 [3,35])	6 (2,42 [4,13])

¹⁾ Im Folgenden beziehen sich alle Angaben auf die Patienten, die die geplante Behandlung tatsächlich erhalten haben

Abkürzungen: IPF: idiopathische Lungenfibrose, UE: unerwünschte Ereignisse, SD: Standardabweichung

Der Anteil der Patienten, die die Studienmedikation frühzeitig abbrechen, war in der INPULSIS-1 Studie in der Nintedanib-Gruppe höher als in der Placebo-Gruppe. Der häufigste Abbruchgrund war UE (21,0 % der Patienten in der Nintedanib-Gruppe und 11,8 % der Patienten in der Placebo-Gruppe) und darunter in beiden Armen jeweils am häufigsten UE, die nicht unmittelbar mit einer Verschlechterung der IPF in Zusammenhang standen (17,5 vs. 6,4 % in INPULSIS-1 bzw. 14,9 vs. 6,8% in INPULSIS-2). In INPULSIS-2 war der Anteil der Patienten, die die Studienmedikation frühzeitig abbrechen, im Nintedanib-Arm etwas höher als im Placebo-Arm (23,7 % vs. 20,1 %). Wie in INPULSIS-1 war auch in INPULSIS-2 UE der häufigste Grund für den Abbruch der Studienmedikation, wobei im Nintedanib-Arm am häufigsten UE, die nicht unmittelbar mit einer Verschlechterung der IPF in Zusammenhang standen (14,9 %) zu dem Abbruch der Studienmedikation führten und im Placebo-Arm am häufigsten UE, die auf eine unerwartete Verschlechterung der IPF zurückzuführen waren (8,2 %).

In INPULSIS-1 erfüllten 84,1 % der Patienten im Nintedanib-Arm und 85,3 % der Patienten im Placebo-Arm die geplante Beobachtungszeit mit allen geplanten Visiten bis Woche 52. Dies schließt auch Patienten ein, die die Studienmedikation vorzeitig abbrechen. In INPULSIS-2 lag der Anteil der Patienten, die die geplante Beobachtungszeit erfüllten bei 82,7 % bzw. 81,7 %.

Trat ein UE auf, konnte die Dosis von 150mg auf 100mg zweimal täglich reduziert werden. Eine erneute Aufdosierung auf 150mg zweimal täglich war innerhalb von 4 Wochen möglich, sofern das UE beseitigt werden konnte. Unterbrechungen der Dosis waren für 4 Wochen möglich, sofern UE auftraten, die mit der Studienmedikation in Verbindung gebracht wurden. Für nicht mit der Studienmedikation in Verbindung stehende UEs und bei akuten Exazerbationen der IPF konnte die Behandlung für bis zu 8 Wochen unterbrochen werden.

In INPULSIS-1 hatten 26,5 % der Patienten in der Nintedanib-Gruppe mindestens 1 Dosisreduktion von 150mg auf 100mg; in der Placebogruppe hatten 4,9 % mindestens 1 Dosisreduktion. In der Nintedanib-Gruppe reeskalierten 6,5 % der Patienten die Nintedanib-Dosis auf von 100 auf 150mg und 77,7 % der Patienten erhielten als letzte Dosis 150mg Nintedanib bid. In der Nintedanib-Gruppe lag die mittlere Dauer der Dosisreduktion bei 4,3 Monaten und im Placebo-Arm bei 3,7 Monaten. In der INPULSIS-1 Studie unterbrachen

25,6 % der Patienten im Nintedanib-Arm die Behandlung für im Schnitt 3,9 Wochen und 11,8 % im Placebo-Arm für 3,1 Wochen.

In INPULSIS-2 hatten 29,2 % der Patienten in der Nintedanib-Gruppe mindestens 1 Dosisreduktion von 150mg auf 100mg; in der Placebogruppe hatten 2,7 % mindestens 1 Dosisreduktion. In der Nintedanib-Gruppe reeskalierten 6,1 % der Patienten die Nintedanib-Dosis auf von 100 auf 150mg und 75,1 % der Patienten erhielten als letzte Dosis 150mg Nintedanib bid. In der Nintedanib-Gruppe lag die mittlere Dauer der Dosisreduktion bei 4,5 Monaten und im Placebo-Arm bei 2,4 Monaten. In der INPULSIS-2 Studie unterbrachen 21,9 % der Patienten im Nintedanib-Arm die Behandlung für im Schnitt 3,2 Wochen und 8,2 % im Placebo-Arm für 4,4 Wochen.

Mortalität

Tabelle 12: Ergebnisse zur Mortalität – Zeit bis zum Tod

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo
	N ¹⁾	Patienten mit Ereignis bis Woche 52 n (%) ²⁾	N ¹⁾	Patienten mit Ereignis bis Woche 52 n (%) ²⁾	HR (95% KI); p-Wert ³⁾
Zeit bis Tod, unabhängig von der Ursache					
INPULSIS-1	309	13 (4,2)	204	13 (6,4)	0,63 (0,29; 1,36); 0,2880
INPULSIS-2	329	22 (6,7)	219	20 (9,1)	0,74 (0,40; 1,35); 0,2995
<i>Gesamt</i>	638	35 (5,5)	423	33 (7,8)	0,70 (0,43; 1,12); 0,2995
Zeit bis respiratorisch bedingtem Tod					
INPULSIS-1	309	10 (3,2)	204	10 (4,9)	0,61 (0,25; 1,47); 0,3515
INPULSIS-2	329	14 (4,3)	219	11 (5,0)	0,86 (0,39; 1,90); 0,6654
<i>Gesamt</i>	638	24 (3,8)	423	21 (5,0)	0,74 (0,41; 1,34); 0,3435

¹⁾ Analyse auf Basis des Treated Set (TS)

²⁾ Der Median wurde nicht erreicht. Im Studienbericht und Dossier wurden an dieser Stelle die Anteile der Patienten mit Ereignis bis Woche 52 angegeben

³⁾ HR [95 % KI] INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels Cox-Regression mit Behandlung, Geschlecht, Alter und Größe als Kovariate. In die gepoolte Analyse wurde zusätzlich Studie als Faktor aufgenommen. p-Wert INPULSIS-1, INPULSIS-2, INPULSIS (gepoolt): berechnet mittels Log-Rank Test

Abkürzungen: HR: Hazard Ratio, KI: Konfidenzintervall, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis

Für die Zeit bis zum Tod und für die Zeit bis zum respiratorisch bedingten Tod innerhalb von 52 Wochen zeigte sich weder in den Einzelstudien noch in der gepoolten Analyse ein statistisch signifikanter Effekt zwischen den Behandlungsgruppen, wobei die Anzahl an Ereignissen in beiden Studien und dort jeweils in beiden Behandlungsgruppen relativ gering war. Numerisch zeigt sich in beiden Studien ein leichter Vorteil für die Nintedanib-Gruppe.

Dossierbewertung für Orphan Drugs

Nintedanib

In INPULSIS-1 verstarben 26 Patienten innerhalb der ersten 52 Wochen und insgesamt 33 Patienten und in INPULSIS-2 verstarben 42 Patienten innerhalb der ersten 52 Wochen und insgesamt 54 Patienten.

Tabelle 13: jährliche Abnahme der FVC

Studie	Behandlung	N ¹⁾	Baseline [ml] MW ²⁾ (SD)	Woche 52 [ml] MW ²⁾ (SD)	jährliche Abnahme [ml] MWD ³⁾ (SE)	Behandlungseffekt Nintedanib vs. Placebo [ml] MWD ³⁾ (SE) [95 %-KI]; p-Wert
jährliche Abnahme der FVC						
INPULSIS-1	Nintedanib	309	2.756,8 (735,12)	2.669,0 (772,04)	114,65 (15,327)	125,26 (24,209) [77,68; 172,84]; < 0,0001
	Placebo	204	2.844,5 (820,11)	2.664,4 (834,01)	239,91 (18,709)	
INPULSIS-2	Nintedanib	329	2.672,8 (775,96)	2.637,3 (811,80)	113,59 (15,726)	93,73 (24,907) [44,78; 142,68]; 0,0002
	Placebo	219	2.619,0 (787,35)	2.512,5 (821,44)	207,32 (19,309)	
Gesamt	Nintedanib	638	2.713,5 (757,03)	2.652,6 (792,29)	113,59 (10,984)	109,94 (17,368) [75,85; 144,03]; < 0,0001
	Placebo	423	2.727,7 (810,25)	2.585,1 (829,76)	223,53 (13,448)	

¹⁾Analyse auf Basis des Treated Set (TS)

²⁾ Nicht adjustierter Mittelwert

³⁾ MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels Random-Coefficient-Model mit Behandlung, Geschlecht, Alter und Größe als Kovariate. In die gepoolte Analyse wurde zusätzlich Studie als Faktor aufgenommen

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, MW: Mittelwert, MWD: Mittelwertdifferenz, N: Gesamtanzahl Treated Set, SD: Standardabweichung, SE: Standardfehler

Der primäre Endpunkt, jährliche Abnahme der FVC zeigte in beiden INPULSIS-Studien und in der gepoolten Auswertung statistisch signifikante Effekte zugunsten der mit Nintedanib behandelten Patienten (Tabelle 13). Über 52 Wochen nahm die FVC in der Nintedanib-Gruppe kontinuierlich weniger ab als in der Placebo-Gruppe (Abbildung 1).

Der Anteil der Patienten, für die keine Werte zu Studienende vorlagen, war in beiden Studien in jeweils beiden Behandlungsarmen vergleichbar (jeweils ca. 15 %).

Die Ergebnisse der Sensitivitätsanalysen bestätigten die Ergebnisse der primären Analysen.

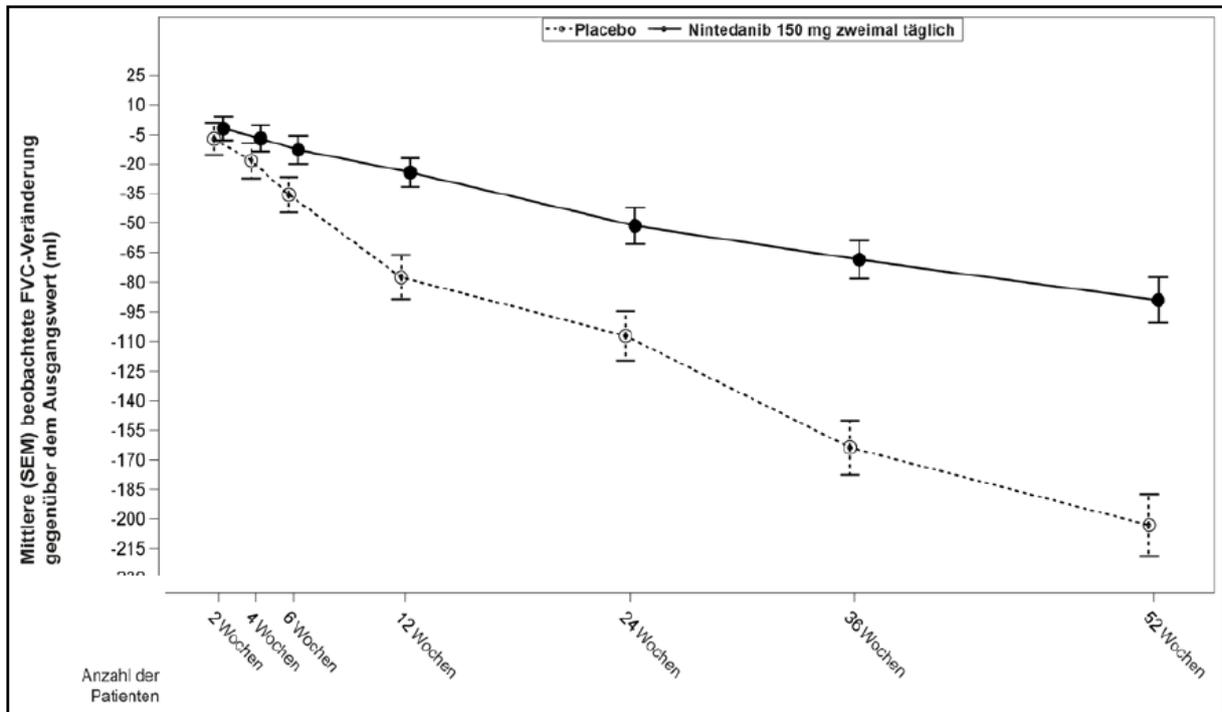


Abbildung 1: Mittlere beobachtete FVC-Veränderung (standard error of the mean, SEM) gegenüber dem Ausgangswert (ml) im Zeitverlauf, Studien INPULSIS-1 und INPULSIS-2 gepoolt

Zusätzlich zu der Auswertung des primären Endpunkts jährliche Abnahme der FVC, wurde als sekundärer Endpunkt die Rate an FVC-Respondern untersucht, definiert als Anteil der Patienten, deren FVC in Woche 52 gegenüber dem individuellen Messwert zu Studienbeginn um nicht mehr als 10 % bzw. 5 % (bezogen auf den Sollwert) abfiel (Tabelle 14).

Tabelle 14: FVC-Responderanalyse

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ²⁾
	N ¹⁾	Patienten mit Ereignis n (%)	N	Patienten mit Ereignis n (%)	OR (95 %-KI); p-Wert
FVC-Responder (Schwellenwert 10 % vom Sollwert)					
INPULSIS-1	309	218 (70,6)	204	116 (56,9)	1,91 (1,32; 2,79); 0,0007
INPULSIS-2	329	229 (69,6)	219	140 (63,9)	1,29 (0,89; 1,86); 0,1833
<i>Gesamt</i>	638	447 (70,1)	423	256 (60,5)	1,58 (1,21; 2,05); 0,0007
FVC-Responder (Schwellenwert 5 % vom Sollwert)					

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ²⁾
	N ¹⁾	Patienten mit Ereignis n (%)	N	Patienten mit Ereignis n (%)	OR (95 %-KI); p-Wert
INPULSIS-1	309	163 (52,8)	204	78 (38,2)	1,85 (1,28; 2,66); 0,0010
INPULSIS-2	329	175 (53,2)	219	86 (39,3)	1,79 (1,26; 2,55); 0,0011
<i>Gesamt</i>	638	338 (53,0)	423	164 (38,8)	1,84 (1,43; 2,36); < 0,0001

¹⁾ Analyse auf Basis des Treated Set (TS)

²⁾ OR (95 % KI) und p-Wert in INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Behandlung, Alter, Geschlecht, Größe und Baseline-FVC (%) als Kovariate; bei INPULSIS (gepoolt) wurde zusätzlich Studie als Faktor in das Modell aufgenommen

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, OR: Odds Ratio

Für alle Responderanalysen ergaben sich in den Studien und der gepoolten Analyse statistisch signifikante Effekte zugunsten der mit Nintedanib behandelten Patienten, bis auf die Responderanalyse mit einer Veränderung von > 10 % in INPULSIS-2, wo kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen gezeigt werden konnte.

Morbidität

Tabelle 15: (adjudizierte) akute Exazerbation

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo
	N ¹⁾	Patienten mit Ereignis bis Woche 52 n (%) ²⁾	N	Patienten mit Ereignis bis Woche 52 n (%) ²⁾	HR (95 %-KI); p-Wert ³⁾
Zeit bis zur ersten akuten Exazerbation⁴⁾					
INPULSIS-1	309	19 (6,1)	204	11 (5,4)	1,15 (0,54; 2,42); 0,6728
INPULSIS-2	329	12 (3,6)	219	21 (9,6)	0,38 (0,19; 0,77); 0,0050
<i>Gesamt</i>	638	31 (4,9)	423	32 (7,6)	0,64 (0,39; 1,05); 0,0823
Zeit bis zur ersten adjudizierten akuten Exazerbation⁵⁾					
INPULSIS-1	309	7 (2,3)	204	8 (3,9)	0,55 (0,20; 1,54); 0,2551
INPULSIS-2	329	5 (1,5)	219	16 (7,3)	0,20 (0,07; 0,56); 0,0020
<i>Gesamt</i>	638	12 (1,9)	423	24 (5,7)	0,32 (0,16; 0,65); 0,0010

¹⁾Analyse auf Basis des Treated Set (TS)

²⁾ Der Median wurde nicht erreicht. Im Studienbericht und Dossier wurden an dieser Stelle die Anteile der Patienten mit Ereignis bis Woche 52 angegeben

³⁾ HR [95 % KI] INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels Cox-Regression mit Behandlung, Geschlecht, Alter und Größe als Kovariate. In die gepoolte Analyse wurde zusätzlich Studie als Faktor aufgenommen. p-Wert INPULSIS-1, INPULSIS-2, INPULSIS (gepoolt): berechnet mittels Log-Rank Test

⁴⁾ Vom Prüfarzt diagnostiziert

⁵⁾ Durch das IAC bestätigte oder vermutete akute Exazerbationen

Abkürzungen: HR: Hazard Ratio, KI: Konfidenzintervall, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis

Für die Zeit bis zur ersten akuten Exazerbation und die Zeit bis zur ersten adjudizierten akuten Exazerbation zeigte sich für INPULSIS-1 jeweils kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen. Für den Endpunkt Zeit bis zur ersten adjudizierten akuten Exazerbation konnte in INPULSIS-2 und der gepoolten Analyse ein statistisch signifikanter Unterschied zugunsten von Nintedanib gezeigt werden, während dieser Effekt für den Endpunkt Zeit bis zur ersten akuten Exazerbation nur in der INPULSIS-2 Studie, jedoch nicht in INPULSIS-1 und nicht in der gepoolten Analyse nachweisbar war.

Die Ergebnisse der Sensitivitätsanalysen bestätigten die Ergebnisse der primären Analysen.

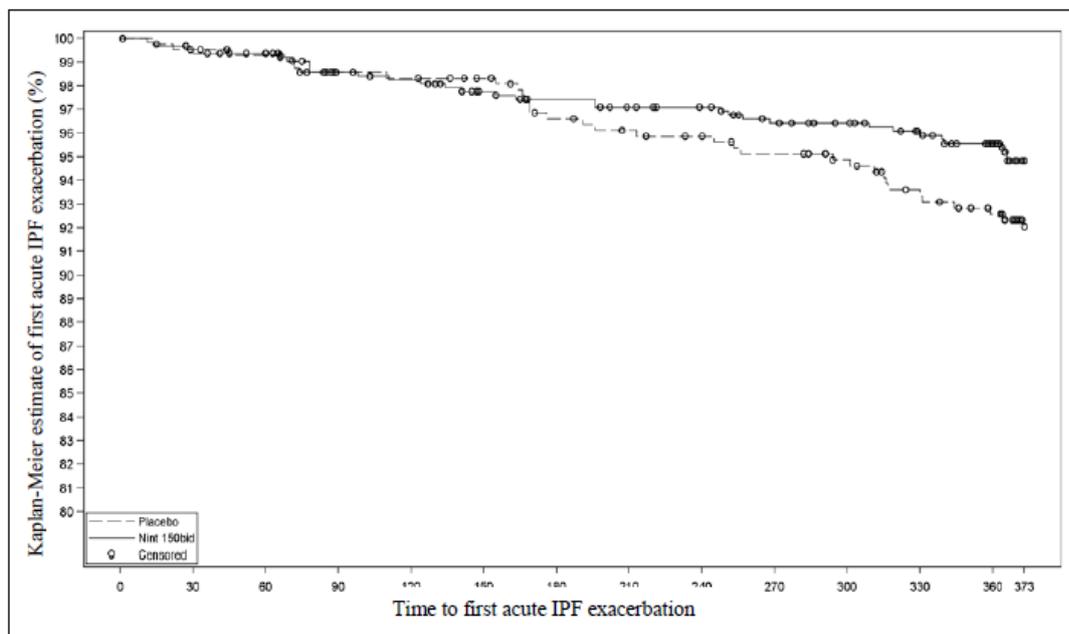


Abbildung 2: Kaplan-Meier-Kurven für Endpunkt Zeit bis zur ersten akuten Exazerbation für die gepoolte Analyse der INPULSIS Studien

Tabelle 16: PGI-C-Responderanalyse

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo
	N ¹⁾	Patienten mit Ereignis n (%)	N ¹⁾	Patienten mit Ereignis n (%)	OR (95 %-KI); p-Wert ²⁾
PGI-C-Responder³⁾					
INPULSIS-1	309	188 (60,8)	204	112 (54,9)	1,28 [0,89; 1,83]; 0,1818
INPULSIS-2	329	203 (61,7)	219	118 (53,9)	1,38 [0,98; 1,95]; 0,0690
<i>Gesamt</i>	638	391 (61,3)	423	230 (54,4)	1,33 [1,04; 1,70]; 0,0254

¹⁾Analyse auf Basis des Treated Set (TS)

²⁾ INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Behandlung als Kovariate. In der gepoolten Analyse wurde zusätzlich Studie als Faktor aufgenommen.

³⁾ Als PGI-C-Responder wurde a priori ein Patient definiert, der bei der Erhebung zu Woche 52 verglichen mit seinem subjektiv erinnerten Befinden bei Studienbeginn keine Kategorie der Verschlechterung seines Zustands angegeben hatte

Abkürzungen: KI: Konfidenzintervall, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, OR: Odds Ratio, PGI-C: Patient's Global Impression of Change

Als PGI-C-Responder wurde a priori ein Patient definiert, der bei der Erhebung zu Woche 52 verglichen mit seinem subjektiv erinnerten Befinden bei Studienbeginn keine Kategorie der Verschlechterung seines Zustands angegeben hatte. In den Studien INPULSIS-1 und 2 konnte ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen nicht nachgewiesen werden, für die gepoolte Analyse zeigte sich ein statistisch signifikanter Effekt zugunsten von Nintedanib.

Tabelle 17: Veränderung der EQ-5D-VAS

Studie	Behandlung	Baseline		Woche 52		Veränderung	
		N ¹⁾	MW (SD)	N	MW (SD)	N	MWD (SD)
Veränderung der EQ-5D-VAS							
INPULSIS-1	Nintedanib	306	66,71 (17,42)	249	65,04 (17,65)	247	-2,46 (18,92)
	Placebo	203	68,02 (16,34)	161	62,12 (20,11)	160	-5,88 (19,17)
INPULSIS-2	Nintedanib	324	69,77 (18,84)	268	68,06 (18,66)	265	-2,52 (16,95)
	Placebo	218	67,75 (16,47)	179	63,04 (18,48)	178	-5,60 (17,67)
<i>Gesamt</i>	Nintedanib	630	68,28 (18,22)	517	66,61 (18,23)	512	-2,49 (17,91)
	Placebo	421	67,88 (16,39)	340	62,61 (19,25)	338	-5,73 (18,37)

¹⁾ Analyse auf Basis des Treated Set (TS)

Abkürzungen: EQ-5D-VAS: EuroQol 5-Dimensional Quality of Life Questionnaire visuelle Analogskala, MW: Mittelwert, MWD: Mittelwertdifferenz, N: Gesamtanzahl Treated Set

Die Veränderung von Baseline bis Woche 52 auf dem EQ-5D-VAS war in den Studien INPULSIS-1, INPULSIS-2 und in der gepoolten Analyse in der Nintedanib-Gruppe geringer als in der Placebo-Gruppe. Der Wert auf dem EQ-5D-VAS sank in beiden Studien und jeweils in beiden Behandlungsgruppen im Laufe der 52 Wochen. Die Verschlechterung des Gesundheitszustandes war im Nintedanib-Arm etwas geringer als im Placebo-Arm. Eine statistische Auswertung und Responderanalysen liegen nicht vor.

Lebensqualität

Tabelle 18: Ergebnisse zur Lebensqualität – Veränderung von Baseline bis Woche 52

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ³⁾ MWD (SE) [95 %-KI]; p-Wert
		N ¹⁾	MW ²⁾ (SD)	N	MW ²⁾ (SD)	N	MWD ³⁾ (SE)	
Veränderung SGRQ-Gesamtscore (Punkte)								
INPULSIS-1	Nintedanib	298	39,55 (17,628)	240	42,67 (20,187)	289	4,34 (0,799)	-0,05 (1,248) [-2,50; 2,40]; 0,9657
	Placebo	202	39,79 (18,478)	163	42,36 (21,391)	200	4,39 (0,960)	
INPULSIS-2	Nintedanib	326	39,46 (20,471)	267	40,33 (21,977)	320	2,80 (0,730)	-2,69 (1,151) [-4,95; -0,43]; 0,0197
	Placebo	217	39,39 (18,647)	179	42,24 (21,113)	213	5,48 (0,891)	
Gesamt	Nintedanib	624	39,51 (19,151)	507	41,44 (21,160)	609	3,53 (0,540)	-1,43 (0,848) [-3,09; 0,23]; 0,0923
	Placebo	419	39,58 (18,545)	342	42,30 (21,215)	413	4,96 (0,654)	
Veränderung SGRQ-I-Gesamtscore (Punkte)								
INPULSIS-1	Nintedanib	299	41,95 (18,377)	239	44,74 (20,894)	290	4,30 (0,824)	-0,78 (1,289) [-3,31; 1,75]; 0,5446
	Placebo	202	42,06 (19,233)	163	45,44 (21,768)	200	5,08 (0,992)	
INPULSIS-2	Nintedanib	326	41,88 (21,286)	264	42,61 (22,560)	320	2,72 (0,757)	-3,12 (1,192) [-5,46; -0,79]; 0,0089
	Placebo	217	41,83 (19,620)	179	44,75 (21,375)	213	5,84 (0,921)	
Gesamt	Nintedanib	625	41,92 (19,932)	503	43,62 (21,789)	610	3,48 (0,559)	-1,99 (0,876) [-3,71; -0,27]; 0,0230

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ³⁾ MWD (SE) [95 %-KI]; p-Wert
		N ¹⁾	MW ²⁾ (SD)	N	MW ²⁾ (SD)	N	MWD ³⁾ (SE)	
	Placebo	217	41,94 (19,412)	342	45,08 (21,534)	413	5,47 (0,676)	
Veränderung SOBQ-Gesamtscore (Punkte)								
INPULSIS-1	Nintedanib	278	32,58 (22,978)	226	37,50 (26,631)	267	6,73 (1,113)	-0,88 (1,770) [-4,35;2,60]; 0,6203
	Placebo	182	32,24 (23,349)	150	38,85 (26,878)	178	7,61 (1,376)	
INPULSIS-2	Nintedanib	309	33,10 (25,700)	255	37,53 (28,087)	302	6,69 (1,073)	-2,38 (1,685) [-5,68;0,93]; 0,1587
	Placebo	210	33,53 (24,078)	172	39,84 (25,972)	204	9,07 (1,300)	
Gesamt	Nintedanib	587	32,85 (24,429)	481	37,51 (27,384)	569	6,71 (0,773)	-1,69 (1,220) [-4,09;0,70]; 0,1648
	Placebo	392	32,93 (23,721)	322	39,38 (26,361)	382	8,40 (0,945)	
Veränderung CASA-Q Husten-Symptom (Punkte)								
INPULSIS-1	Nintedanib	309	58,63 (23,592)	251	58,50 (24,455)	302	-0,76 (1,136)	-0,24 (1,803) [-3,78;3,30]; 0,8942
	Placebo	204	56,29 (22,861)	164	57,22 (24,829)	202	-0,52 (1,400)	
INPULSIS-2	Nintedanib	329	61,60 (23,858)	268	62,62 (25,197)	323	-0,33 (1,087)	2,05 (1,713) [-1,31;5,41]; 0,2326
	Placebo	219	62,52 (21,423)	181	61,23 (21,562)	215	-2,38 (1,325)	
Gesamt	Nintedanib	638	60,16 (23,757)	519	60,63 (24,902)	625	-0,57 (0,785)	0,95 (1,241) [-1,49;3,38]; 0,4450
	Placebo	423	59,52 (22,320)	345	59,32 (23,226)	417	-1,52 (0,962)	
Veränderung CASA-Q Husten-Impact (Punkte)								
INPULSIS-1	Nintedanib	309	74,22 (22,836)	251	73,82 (23,077)	302	-2,36 (1,006)	1,64 (1,596) [-1,49;4,77]; 0,3042
	Placebo	204	74,18 (22,340)	164	71,91 (25,920)	202	-4,00 (1,240)	
INPULSIS-2	Nintedanib	328	75,55 (24,116)	269	75,14 (25,252)	322	-2,58 (0,991)	1,81 (1,564) [-1,26;4,88]; 0,2475
	Placebo	219	77,04 (21,881)	181	74,79 (22,123)	215	-4,39 (1,209)	
Gesamt	Nintedanib	637	74,91 (23,495)	520	74,50 (24,212)	624	-2,48 (0,705)	1,73 (1,115) [-0,46;3,92]; 0,1212

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ³⁾ MWD (SE) [95 %-KI]; p-Wert
		N ¹⁾	MW ²⁾ (SD)	N	MW ²⁾ (SD)	N	MWD ³⁾ (SE)	
	Placebo	423	75,66 (22,125)	345	73,42 (24,011)	417	-4,21 (0,865)	

¹⁾ Analyse auf Basis des Treated Set (TS)

²⁾ nicht adjustierte Mittelwerte

³⁾ MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2: berechnet mittels MMRM mit festen Effekten für Behandlung, Visite, Baselinewert des Instruments, Baselinewert*Visite und zufälligem Effekt für Studienteilnehmer; MWD [95 % KI] und p-Wert INPULSIS (gepoolt): berechnet mittels MMRM mit festen Effekten für Studie, Behandlung, Behandlung Visite, Baselinewert des Instruments, Baselinewert*Visite und zufälligem Effekt für Studienteilnehmer

Abkürzungen: CASA-Q: Cough and Sputum Assessment Questionnaire, IPF: Idiopathische Lungenfibrose, KI: Konfidenzintervall, MW: Mittelwert, MWD: Mittelwertdifferenz, N: Gesamtanzahl Treated Set, SD: Standardabweichung, SE: Standardfehler, SGRQ: St. George's Respiratory Questionnaire, SGRQ-I: SGRQ IPF spezifische Version, SOBQ: Shortness of Breath Questionnaire

Die Lebensqualität sank gemessen an allen Instrumenten zwischen Baseline und Woche 52 sowohl im Placebo-Arm als auch im Nintedanib-Arm beider INPULSIS-Studien.

Die Veränderung des SGRQ-Gesamtscores war neben der Zeit bis zur akuten Exazerbation der zentrale sekundäre Endpunkt in den beiden INPULSIS-Studien. In INPULSIS-2 zeigte sich ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe (statistisch signifikant geringere Abnahme der Lebensqualität im Vergleich zu Placebo), in der INPULSIS-1 Studie und in der gepoolten Analyse war kein Unterschied zwischen den Behandlungsgruppen nachweisbar. Der Anteil der Patienten, für die keine Werte zu Studienende vorlagen, war in beiden Studien in jeweils beiden Behandlungsarmen vergleichbar (jeweils ca. 20 %). Die Ergebnisse der Sensitivitätsanalysen für die Veränderung des SGRQ-Gesamtscores bestätigten die Ergebnisse der primären Analysen.

In INPULSIS-2 und in der gepoolten Analyse zeigte sich für die IPF-spezifische Version des SGRQ ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe, in der INPULSIS-1 Studie war kein Unterschied zwischen den Behandlungsgruppen nachweisbar. Der Anteil der Patienten, für die keine Werte zu Studienende vorlagen, war in beiden Studien in jeweils beiden Behandlungsarmen vergleichbar (jeweils ca. 20 %).

Für den SOBQ-Fragebogen und die Husten-Symptom und Husten-Impact Domänen des CASA-Q Fragebogens zeigten sich weder in den INPULSIS-Studien noch in der gepoolten Analyse statistisch signifikante Unterschiede.

Tabelle 19: Ergebnisse zur Lebensqualität – SGRQ-Responder und SGRQ-I-Responder

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo
	N ¹⁾	Patienten mit Ereignis n (%)	N ¹⁾	Patienten mit Ereignis n (%)	OR [95 %-KI]; p-Wert ²⁾
SGRQ-(≤ -4 Punkte)-Responder					
INPULSIS-1	309	63 (20,4)	204	49 (24,0)	0,84 [0,55; 1,29]; 0,4298
INPULSIS-2	329	83 (25,2)	219	37 (16,9)	1,66 [1,08; 2,57]; 0,0218
<i>Gesamt</i>	638	146 (22,9)	423	86 (20,3)	1,19 [0,88; 1,61]; 0,2715
SGRQ-I-(≤ -4 Punkte)-Responder					
INPULSIS-1	309	66 (21,4)	204	49 (24,0)	0,89 [0,58; 1,36]; 0,5821
INPULSIS-2	329	84 (25,5)	219	34 (15,5)	1,88 [1,20; 2,93]; 0,0056
<i>Gesamt</i>	638	150 (23,5)	423	83 (19,6)	1,28 [0,95; 1,74]; 0,1091

¹⁾ Analyse auf Basis des Treated Set (TS)

²⁾ OR [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore als Kovariate; INPULSIS (gepoolt): berechnet mittels logistischer Regression mit Studie, Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore als Kovariate

Abkürzungen: IPF: Idiopathische Lungenfibrose, KI: Konfidenzintervall, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, OR: Odds Ratio, SGRQ: St. George's Respiratory Questionnaire, SGRQ-I: SGRQ IPF spezifische Version

Für den SGRQ- und den SGRQ-I-Fragebogen wurden Responderanalysen mit einem Responderkriterium von ≤ -4 Punkten durchgeführt. In INPULSIS-2 zeigte sich für den SGRQ und den SGRQ-I ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe. In INPULSIS-1 und der gepoolten Analyse war kein Unterschied im SGRQ und SGRQ-I zwischen den Behandlungsgruppen nachweisbar.

Sicherheit

Tabelle 20: Ergebnisse der Nebenwirkungen – Zusammenfassung der UE während der verblindeten Behandlungsphase

	INPULSIS-1					INPULSIS-2					Gesamt				
	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo
	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert
Gesamtzahl UE	309	298 (96,4)	204	181 (88,7)	3,44 (1,64; 7,23) <0,001	329	311 (94,5)	219	198 (90,4)	1,83 (0,95; 3,53) 0,066	638	609 (95,5)	423	379 (89,6)	2,44 (1,50; 3,96) <0,001
Schwere UE	309	81 (26,2)	204	37 (18,1)	k.A.	329	93 (28,3)	219	62 (28,3)	k.A.	638	k.A.	423	k.A.	k.A.
Schwerwiegende UE	309	96 (31,1)	204	55 (27,0)	1,22 (0,83; 1,81) 0,318	329	98 (29,8)	219	72 (32,9)	0,87 (0,60; 1,25) 0,444	638	194 (30,4)	423	127 (30,0)	1,02 (0,78; 1,33) 0,894
UE, die zu einer permanenten Dosisreduktion führten	309	52 (16,8)	204	2 (1,0)	k.A.	329	49 (14,9)	219	0 (0,0)	k.A.	638	101 (15,8)	423	2 (0,005)	k.A.
UE, die zu Therapieabbruch führten	309	65 (21,0)	204	22 (10,8)	2,20 (1,31; 3,71) 0,002	329	58 (17,6)	219	33 (15,1)	1,17 (0,79; 1,73) 0,430	638	123 (19,3)	423	55 (13,0)	1,48 (1,11; 1,99) 0,007
UE, das zum Tod führte	309	12 (3,9)	204	10 (4,9)	0,78 (0,33; 1,85) 0,577	329	25 (7,6)	219	21 (9,6)	0,78 (0,42; 1,42) 0,411	638	37 (5,8)	423	31 (7,3)	0,78 (0,48; 1,28) 0,319
UE von besonderem Interesse (a priori)															
<i>Diarrhoe</i>	309	190 (61,5)	204	38 (18,6)	6,97 (4,58; 10,6) < 0,001	329	208 (63,2)	219	40 (18,3)	7,69 (5,11; 11,60) < 0,001	638	398 (62,4)	423	78 (18,4)	7,33 (5,47; 9,84) < 0,001
<i>Übelkeit</i>	309	70 (22,7)	204	12 (5,9)	4,69 (2,47; 8,90) < 0,001	329	86 (26,1)	219	16 (7,3)	4,49 (2,55; 7,90) < 0,001	638	156 (24,5)	423	28 (6,6)	4,57 (2,99; 6,97) < 0,001
<i>Erbrechen</i>	309	40 (12,9)	204	4 (2,0)	7,43 (2,62; 21,1) < 0,001	329	34 (10,3)	219	7 (3,2)	3,49 (1,52; 8,02) 0,002	638	74 (11,6)	423	11 (2,6)	4,91 (2,58; 9,38) < 0,001

	INPULSIS-1					INPULSIS-2					Gesamt				
	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo
	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert
<i>Appetitlosigkeit</i>	309	26 (8,4)	204	14 (6,9)	1,25 (0,63; 2,45) 0,521	329	42 (12,8)	219	10 (4,6)	3,06 (1,50; 6,24) 0,001	638	68 (10,7)	423	24 (5,7)	1,98 (1,22; 3,21) 0,005
<i>Gewichtsverlust</i>	309	25 (8,1)	204	13 (6,4)	1,29 (0,65; 2,59) 0,467	329	37 (11,2)	219	2 (0,9)	13,7 (3,28; 57,70) < 0,001	638	62 (9,7)	423	15 (3,5)	2,93 (1,64; 5,22) < 0,001
UE von besonderem Interesse (zusätzlich, laut Angaben im Dossier) ²⁾															
<i>Gastrointestinale Perforationen</i>	309	k.A.	204	k.A.	k.A.	329	k.A.	219	k.A.	k.A.	638	2 (0,3)	423	0 (0)	n.b. 0,2491
<i>Leberenzymerrhöhung (AST, ALT) ≥ 3x U³⁾</i>	309	k.A.	204	k.A.	k.A.	329	k.A.	219	k.A.	k.A.	638	32 (5,0)	423	3 (0,7)	7,39 (2,25; 24,3) < 0,001
<i>Bilirubinerhöhung ≥ 1,5x ULN</i>	309	k.A.	204	k.A.	k.A.	329	k.A.	219	k.A.	k.A.	638	15 (2,4)	423	3 (0,7)	3,37 (0,97; 11,7) 0,043
<i>ALKP-Erhöhung ≥ 1,5x ULN</i>	309	k.A.	204	k.A.	k.A.	329	k.A.	219	k.A.	k.A.	638	37 (5,8)	423	4 (0,9)	6,45 (2,28; 18,2) < 0,001
<i>Unerwünschte kardiale Ereignisse</i>	309	30 (9,7)	204	19 (9,3)	1,05 (0,57; 1,92) 0,882	329	34 (10,3)	219	26 (11,9)	0,86 (0,50; 1,47) 0,572	638	64 (10,0)	423	45 (10,6)	0,94 (0,63; 1,40) 0,750
<i>Schwerwiegende unerwünschte kardiale Ereignisse</i>	309	14 (4,5)	204	11 (5,4)	0,83 (0,37; 1,87) 0,657	329	18 (5,5)	219	12 (5,5)	1,00 (0,47; 2,12) 0,997	638	32 (5,0)	423	23 (5,4)	0,92 (0,53; 1,59) 0,762
<i>Tödliche kardiale Ereignisse</i>	309	1 (0,3)	204	2 (1,0)	0,33 (0,03; 3,64) 0,340	329	2 (0,6)	219	4 (1,8)	0,33 (0,06; 1,81) 0,179	638	3 (0,5)	423	6 (1,4)	0,33 (0,08; 1,32) 0,099
<i>Myocardinfarkt/akuter Myokardinfarkt</i>	309	5 (1,6)	204	1 (0,5)	3,34 (0,39; 28,79) 0,2449	329	5 (1,5)	219	1 (0,5)	3,36 (0,39; 28,99) 0,2414	638	10 (1,6)	423	2 (0,5)	3,35 (0,73; 15,37) 0,0988

Dossierbewertung für Orphan Drugs

Nintedanib

	INPULSIS-1					INPULSIS-2					Gesamt				
	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo	Nintedanib		Placebo		Behandlungs- effekt Nintedanib vs. Placebo
	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert	N ¹⁾	n (%)	N ¹⁾	n (%)	OR (95 %-KI) p-Wert
<i>Arterielle thromboembolische Ereignisse</i>	309	8 (2,6)	204	1 (0,5)	5,40 (0,67; 43,47) 0,0764	329	8 (2,4)	219	2 (0,9)	2,70 (0,57; 12,86) 0,1933	638	16 (2,5)	423	3 (0,7)	3,60 (1,04; 12,44) 0,0305
<i>Bronchitis</i>	309	36 (11,7)	204	28 (13,7)	0,83 (0,49; 1,41) 0,486	329	31 (9,4)	219	17 (7,8)	1,24 (0,67; 2,29) 0,501	638	67 (10,5)	423	45 (10,6)	0,99 (0,66; 1,47) 0,943
<i>Nasopharyngitis</i>	309	39 (12,6)	204	34 (16,7)	0,72 (0,44; 1,19) 0,199	329	48 (14,6)	219	34 (15,5)	0,93 (0,58; 1,50) 0,764	638	87 (13,6)	423	68 (16,1)	0,82 (0,58; 1,16) 0,271
<i>Unerwünschte Blutungsereignisse</i>	309	29 (9,4)	204	17 (8,3)	1,14 (0,61; 2,13) 0,6832	329	37 (11,2)	219	16 (7,3)	1,61 (0,87; 2,97) 0,1264	638	66 (10,3)	423	33 (7,8)	1,36 (0,88; 2,11) 0,1631

¹⁾Analyse auf Basis des Treated Set (TS)

²⁾ Als UE von besonderem Interesse wurden im Dossier dargestellt: unerwünschte gastrointestinale Ereignisse, unerwünschte Leberenzym- und Bilirubinveränderungen, unerwünschte kardiale Ereignisse, unerwünschte respiratorische Ereignisse, unerwünschte Blutungsereignisse, unerwünschte Ereignisse mit einer Häufigkeit von > 10%. Diese Definition weicht zum Teil von der Definition der Zulassungsstudien ab. Eine Begründung für die Hinzunahme weiterer UE liegt nicht vor. Zudem sind einige der vom pU post-hoc definierten UE von besonderem Interesse aus der Gruppe respiratorische Ereignisse bereits Bestandteil von Wirksamkeitsendpunkten, wie etwa Progression der IPF, Dyspnoe und Husten, und wurden daher in dieser Tabelle nicht aufgeführt

³⁾ Laut Fachinformation wird bei einer Transaminaseerhöhungen (AST, ALT) von > 3xULN eine Dosisreduktion oder Unterbrechung der Therapie mit Nintedanib empfohlen [5] [6]

Abkürzungen: k.A. Keine Angabe, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, Anzahl ULN: Upper Limit of Normal (oberer Normgrenzwert)

Die Analyse der UE basierte auf der Anzahl der Patienten mit einem UE, das ab der ersten Dosis der Studienmedikation bis 28 Tage nach Einnahme der letzten Dosis der Studienmedikation aufgetreten war. UE die mit dem MedDRA Preferred Term (PT) IPF kodiert waren, schlossen Verschlechterung der IPF und Exazerbation der IPF ein.

Tabelle 21: UE mit Inzidenz ≥ 10 % in den Studien INPULSIS-1 oder INPULSIS-2

MedDRA Systemorganklasse, Preferred Term	Patienten mit mindestens einem Ereignis, n (%)			
	INPULSIS-1		INPULSIS-2	
	Nintedanib N=309	Placebo N=219	Nintedanib N=329	Placebo N=219
Gesamtrate	298 (96,4)	181 (88,7)	311 (94,5)	198 (90,4)
Infektionen und parasitäre Erkrankungen	161 (52,1)	110 (53,9)	198 (60,2)	118 (53,9)
Nasopharyngitis	39 (12,6)	34 (16,7)	48 (14,6)	34 (15,5)
Bronchitis	36 (11,7)	28 (13,7)	31 (9,4)	17 (7,8)
Infektionen der oberen Atemwege	28 (9,1)	18 (8,8)	30 (9,1)	24 (11,0)
Stoffwechsel- und Ernährungsstörungen	44 (14,2)	23 (11,3)	71 (21,6)	37 (16,9)
Appetit vermindert	26 (8,4)	14 (6,9)	42 (12,8)	10 (4,6)
Erkrankungen des Nervensystems	54 (17,5)	36 (17,6)	51 (15,5)	29 (13,2)
Gefäßerkrankungen	29 (9,4)	9 (4,4)	25 (7,6)	23 (10,5)
Erkrankungen der Atemwege, des Brustraums und Mediastinums	117 (37,9)	86 (42,2)	137 (41,6)	91 (41,6)
Husten	47 (15,2)	26 (12,7)	38 (11,6)	31 (14,2)
IPF	31 (10,0)	21 (10,3)	33 (10,0)	40 (18,3)
Dyspnoe	22 (7,1)	23 (11,3)	27 (8,2)	25 (11,4)
Erkrankungen des Gastrointestinaltrakts	235 (76,1)	71 (34,8)	253 (76,9)	97 (44,3)
Diarrhoe	190 (61,5)	38 (18,6)	208 (63,2)	40 (18,3)
Übelkeit	70 (22,7)	12 (5,9)	86 (26,1)	16 (7,3)
Erbrechen	40 (12,9)	4 (2,0)	34 (10,3)	7 (3,2)
Erkrankungen der Haut und des Unterhautzellgewebes	50 (16,2)	29 (14,2)	46 (14,0)	31 (14,2)
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	60 (19,4)	46 (22,5)	58 (17,6)	44 (22,4)
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	72 (23,3)	51 (25,0)	80 (24,3)	55 (25,1)
Untersuchungen	88 (28,5)	44 (21,6)	97 (29,5)	25 (11,4)
Gewichtsverlust	25 (8,1)	13 (6,4)	37 (11,2)	2 (0,9)
Verletzungen, Vergiftungen und Behandlungskomplikationen	25 (8,1)	20 (9,8)	30 (9,1)	28 (12,8)

Die Anzahl an Patienten mit mindestens einem UE war in beiden INPULSIS-Studien als auch in der gepoolten Analyse höher in der Nintedanib-Gruppe im Vergleich zur Placebo-Gruppe (95,5 % vs. 89,6 % für die gepoolte Analyse) (Tabelle 20). Der Unterschied in der gepoolten Analyse war statistisch signifikant zuungunsten von Nintedanib. Die häufigsten UE (>40 %) bezogen auf Systemorganklassen (SOC) waren in beiden Studien Erkrankungen des Gastrointestinaltrakts, Infektionen und parasitäre Erkrankungen und Erkrankungen der Atemwege, des Brustraums und Mediastinums (Tabelle 21). In beiden Studien zeigte sich eine deutlich höhere Inzidenz an Erkrankungen des Gastrointestinaltrakts im Nintedanib-Arm im

Dossierbewertung für Orphan Drugs

Nintedanib

Vergleich zum Placebo-Arm. Die PT Diarrhoe, Übelkeit und Erbrechen waren jeweils die häufigsten PT in der SOC Erkrankungen des Gastrointestinaltrakts und traten in beiden Studien jeweils mehr als dreimal so häufig im Nintedanib-Arm auf als im Placebo-Arm. Auch in der SOC Untersuchungen war die Inzidenz an UE im Nintedanib-Arm in beiden Studien höher als im Placebo-Arm (Tabelle 21). Die häufigsten PT in dieser SOC waren Gewichtsverlust und Leberwertveränderungen.

Als UE von besonderem Interesse (Tabelle 20) wurden in den Studien folgende PT betrachtet: Diarrhoe, Übelkeit, Erbrechen, Dehydrierung, Gewichtsverlust und Appetitverlust. In INPULSIS-1 erlebten 61,5 % der Patienten in der Nintedanib-Gruppe und 18,6 % der Patienten in der Placebo-Gruppe mindestens 1 UE des PT Diarrhoe (INPULSIS-2: 63,2 % vs. 18,3 %). Laut Angaben im Dossier konnten 78,6 % der Studienteilnehmer mit Diarrhoe beider gepoolter Studien die Therapie mit der Normaldosis fortsetzen, weitere 14,3 % der Studienteilnehmer mit reduzierter Dosis. Bei 7,0 % der Studienteilnehmer mit Diarrhoe (entsprechend 4,4 % aller Studienteilnehmer der Nintedanib-Gruppe) kam es aufgrund von Diarrhoe zu einem endgültigen Therapieabbruch. Diarrhoe trat in den meisten Fällen mit leichter oder moderater Intensität auf. Die Ereignis-Profile für andere UE von besonderem Interesse (Übelkeit, Erbrechen, Gewichtsabnahme und Appetitlosigkeit) waren ähnlich wie bei Diarrhoe. Der Anteil der Patienten in der Placebo und Nintedanib-Gruppe war in INPULSIS-1 jeweils wie folgt: Übelkeit (22,7 % vs. 5,9 %), Erbrechen (12,9 % vs. 2,0 %), Gewichtsabnahme (8,1 % vs. 6,4 %) und Appetitverlust (8,4 % vs. 6,9 %) und in INPULSIS-2 Übelkeit (26,1 % vs. 7,3 %), Erbrechen (10,3 % vs. 3,2 %) Gewichtsabnahme (11,2 % vs. 0,9 %) und verminderter Appetit (12,8 % vs. 4,6 %). Die meisten prädefinierten UE von besonderem Interesse waren leicht bis mittelschwer. Im Nintedanib-Arm der INPULSIS-1 Studie traten vereinzelt Fälle von SUE in den PT verminderter Appetit, Diarrhoe, Erbrechen und Gewichtsabnahme auf. In den meisten Fällen erholten sich die Patienten von diesen Ereignissen ohne dauerhafte Dosisreduktion oder vorzeitige Beendigung der Behandlung. In INPULSIS-2 trat ein SUE des PT Appetitverlust in der Placebo-Gruppe auf, alle anderen UE von besonderem Interesse waren leicht bis mittelschwer und nicht schwerwiegend. In der gepoolten Analyse waren alle a priori definierten UE von besonderem Interesse statistisch signifikant unterschiedlich zuungunsten von Nintedanib (Tabelle 20).

Zusätzlich zu den a priori definierten UE berichtete der pU im Dossier weitere UE als UE von besonderem Interesse (siehe Tabelle 10). In der gepoolten Analyse ergaben sich statistisch signifikante Unterschiede zuungunsten von Nintedanib für die PT Leberenzymhöhung, Bilirubinerhöhung, ALKP-Erhöhung sowie für die SOC arterielle thromboembolische Ereignisse. Im Nintedanib-Arm erlitt ein numerisch höherer Prozentsatz der Patienten einen Myokardinfarkt, ein arterielles thromboembolisches Ereignis oder ein Blutungsereignis als in der Placebogruppe (Tabelle 20). Die PT Progression der IPF und Dyspnoe waren statistisch signifikant unterschiedlich zugunsten von Nintedanib, wobei diese UE zumindest teilweise doppelt erfasst wurden, da diese Ereignisse auch Bestandteil von Wirksamkeitsendpunkten waren (Zeit bis zur ersten akuten Exazerbation).

UE, die zu einer dauerhaften Dosisreduktion führten, waren vor allem Erkrankungen des Gastrointestinaltrakts, darunter vor allem Diarrhoe (jeweils knapp 12 % in beiden Studien im Nintedanib-Arm).

UE, die zu einem Abbruch der Therapie führten, kamen in der INPULSIS-1 Studie im Nintedanib-Arm häufiger vor als im Placebo-Arm (21,0 % vs. 10,8 %). In INPULSIS-2 war die Inzidenz ausgeglichener (17,6 % vs. 15,1 %). Der Unterschied in der gepoolten Analyse war statistisch signifikant zuungunsten von Nintedanib. Im Nintedanib-Arm brachen die meisten Patienten die Therapie aufgrund von Erkrankungen des Gastrointestinaltrakts ab (INPULSIS-1: 8,4 %; INPULSIS-2: 6,4 %) und in dieser SOC vor allem aufgrund des PT Diarrhoe (4,5 % und 4,3 %).

UE die zum Tode führten traten in INPULSIS-1 bei 12 Patienten (3,9 %) im Nintedanib-Arm und bei 10 Patienten (4,9 %) im Placebo-Arm auf. Das häufigste zum Tode führende UE war IPF (2,3 % vs. 2,0 %). In INPULSIS-2 verstarben 24 Patienten (7,6 %) im Nintedanib-Arm und 21 Patienten (9,6 %) im Placebo-Arm. Auch in INPULSIS-2 war das häufigste zum Tode führende UE IPF (3,3 % vs. 5,5 %).

Die Inzidenz von SUE im Nintedanib-Arm war in der INPULSIS-1 Studie etwas höher und in der INPULSIS-2 Studie etwas niedriger als im Placebo-Arm. In der gepoolten Analyse zeigt sich kein Unterschied zwischen den Behandlungsgruppen (Tabelle 20). Das häufigste SUE in beiden Studien war Verschlechterung der IPF. In INPULSIS-1 traten zudem bei jeweils 5 Patienten der Nintedanib-Gruppe (1,6 %) Pneumonien, akutes respiratorisches Versagen, Lungenhochdruck und Brustschmerzen auf und bei 6 Patienten (2,9 %) der Placebo-Gruppe Lungenhochdruck und bei 5 (2,5 %) Pneumonie. In INPULSIS-2 traten bei 18 Patienten der Nintedanib-Gruppe (5,5 %) Pneumonien, und bei 6 (1,8 %) Lungenhochdruck auf und bei 11 Patienten (5,0 %) der Placebo-Gruppe Pneumonie und bei 5 (2,3 %) respiratorisches Versagen.

Subgruppenanalysen

Entsprechend der statistischen Analysepläne der INPULSIS-Studien, waren Subgruppenanalysen für verschiedene Subgruppen geplant, jedoch nur für pharmakokinetische Endpunkte und für die Subgruppe japanische vs. nicht japanische Patienten (siehe Tabelle 1). Im Dossier wurden Subgruppen-Analysen für die gepoolten Daten der INPULSIS Studien dargestellt, nicht jedoch für die Einzelstudien. Im Dossier ist angegeben, dass Ländereffekte im Rahmen des Zulassungsverfahrens analysiert und mit der EMA diskutiert wurden. Aufgrund der hieraus gewonnenen Erkenntnisse sei auf die Durchführung von weiteren Interaktionstests dieser Subgruppe verzichtet worden. Für die Endpunkte Zeit bis zum Tod, unabhängig von der Ursache, Zeit bis zum respiratorisch bedingtem Tod und Zeit bis zur ersten adjudizierten akuten Exazerbation wurden aufgrund der geringen Ereignisraten keine Subgruppenanalysen durchgeführt, für alle anderen Endpunkte wurden Subgruppenanalysen für das Dossier durchgeführt.

Im Folgenden werden die Ergebnisse der Subgruppenanalysen, für die Belege für Effektmodifikationen vorlagen, tabellarisch dargestellt. Subgruppenanalysen, bei denen sich in Interaktionstests Hinweise auf Effektmodifikation ergaben, werden im Text beschrieben.

Veränderung der FVC von Baseline bis Woche 52 und FVC-Responder

Ein Beleg auf Effektmodifikation zeigte sich in der Subgruppe Raucherstatus für FVC-Responder (5 % vom Sollwert). In der Subgruppe ex-Raucher und Raucher wurde ein statistisch signifikanter Effekt zugunsten von Nintedanib gezeigt, in der Subgruppe nicht-Raucher zeigte sich ein numerischer aber nicht statistisch signifikanter Unterschied.

Tabelle 22: Subgruppenanalyse FVC-Responder

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
FVC-Responder (5 % vom Sollwert)					
Raucherstatus p-Wert Interaktionsterm: 0,0166					
Nicht-Raucher	Nintedanib	174	77 (44,3)	1,12	0,70; 1,80
	Placebo	122	51 (41,8)		
Ex-Raucher/ Raucher	Nintedanib	464	261 (56,3)	2,22	1,64; 2,99
	Placebo	301	113 (37,5)		

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis, OR: Odds Ratio

Für den primären Endpunkt Veränderung der FVC ergaben sich Hinweise auf Effektmodifikation in den Subgruppen Geschlecht und SGRQ-Gesamtscore zum Baseline. In der Subgruppe Geschlecht war die FVC-Abnahme bis Woche 52 bei Männern höher als bei Frauen. Sowohl bei Männern als auch bei Frauen nahm die FVC im Placebo-Arm statistisch signifikant stärker ab als im Nintedanib-Arm. Auch für die FVC-Responderanalyse (10 % vom Sollwert) zeigte sich ein Hinweis auf Effektmodifikation mit vergleichbaren Effekten in der Subgruppe Geschlecht. In der Subgruppe SGRQ-Gesamtscore war die Veränderung des FVC bei Patienten mit einem höheren SGRQ-Gesamtscore zum Baseline größer als bei Patienten mit einem niedrigeren SGRQ-Score, was auf eine stärkere jährliche Abnahme der FVC bei Studienteilnehmern mit einer bei Studienbeginn subjektiv schlechter empfundenen gesundheitsbezogenen Lebensqualität hinweist. Der Unterschied zwischen den Behandlungsgruppen war sowohl bei einem SGRQ-Gesamtscore ≤ 40 Punkte als auch > 40 Punkte statistisch signifikant zugunsten von Nintedanib im Vergleich zu Placebo. Für die Analyse der FVC-Responder (10 % vom Sollwert) ergab sich darüber hinaus ein Hinweis auf eine Effektmodifikation bezüglich der Studien, wobei insgesamt in beiden Subgruppen ein höherer Anteil an Patienten im Nintedanib-Arm Responder waren, jedoch nur in INPULSIS-1 ein signifikanter Unterschied zugunsten von Nintedanib gezeigt werden konnte, nicht jedoch in INPULSIS-2.

Zeit bis zur ersten akuten Exazerbation

Für den Endpunkt Zeit bis zur ersten akuten Exazerbation liegt ein Beleg für Effektmodifikation aus dem Interaktionstest für das Subgruppenmerkmal Studie vor. In INPULSIS-2 war dieser Endpunkt statistisch signifikant zugunsten des Nintedanib-Arms, während sich in INPULSIS-1 kein Unterschied zeigte.

Tabelle 23: Subgruppenanalyse Zeit bis zur ersten akuten Exazerbation

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				HR	95 % KI
Zeit bis zur ersten akuten Exazerbation (vom Prüfarzt erfasst)					
Studie p-Wert Interaktionsterm: 0,0320					
INPULSIS-1	Nintedanib	309	19 (6,1)	1,15	0,54;2,42
	Placebo	204	11 (5,4)		
INPULSIS-2	Nintedanib	329	12 (3,6)	0,38	0,19;0,77
	Placebo	219	21 (9,6)		

Abkürzungen: HR: Hazard Ratio, KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis

Für den Endpunkt Zeit bis zur ersten akuten Exazerbation zeigten sich Hinweise auf Effektmodifikation in den Subgruppen FVC (% vom Sollwert zum Baseline ≥ 70 oder < 70) und für systemische Kortikosteroide zu Studienbeginn. Patienten im Nintedanib-Arm mit einem geringeren FVC-Wert zu Studienbeginn hatten statistisch signifikant seltener eine akute Exazerbation, während sich für Patienten mit besseren FVC-Werten zu Studienbeginn kein Unterschied zeigte. Patienten, die zu Studienbeginn systemische Kortikosteroide erhielten, hatten statistisch signifikant seltener eine akute Exazerbation, während sich für Patienten ohne systemische Kortikosteroide kein Unterschied zeigte.

PGI-C-Responder

Für die PGI-C-Responderanalyse zeigte sich ein Hinweis auf Effektmodifikation in der Subgruppe FVC (% vom Sollwert). Der Anteil der PGI-C-Responder war in beiden Gruppen (≤ 70 % zum Baseline, > 70 % zum Baseline) im Nintedanib-Arm größer als im Placebo-Arm. Der Unterschied zwischen den Behandlungsgruppen war bei schwerer erkrankten Studienteilnehmern statistisch signifikant zugunsten von Nintedanib.

SGRQ-Gesamtscore und Responder

Für die Bronchodilator-Begleittherapie, FVC-Baselinewert (% vom Sollwert), Alter und Studie ergaben sich Belege für Effektmodifikation im Interaktionstest in der Responderanalyse des SGRQ. Statistisch signifikante Unterschiede zeigten sich jeweils nur in der Subgruppe der Patienten mit einem schlechteren FVC-Ausgangswert als Prozent vom Sollwert und bei Patienten, die zu Studienbeginn eine Bronchodilator-Begleittherapie erhielten, sowie für die Studie INPULSIS-2, nicht aber für INPULSIS-1.

Tabelle 24: Subgruppenanalyse SGRQ-(≤ -4 Punkte)-Responder

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
SGRQ-(≤ -4 Punkte)-Responder					
FVC (% pred.) p-Wert Interaktionsterm: 0,0293					
≤ 70 %	Nintedanib	207	51 (24,6)	1,93	1,11; 3,38
	Placebo	154	23 (14,9)		
> 70 %	Nintedanib	431	95 (22,0)	0,92	0,63; 1,33
	Placebo	269	63 (23,4)		
Bronchodilator p-Wert Interaktionsterm: 0,0002					
Ja	Nintedanib	129	32 (24,8)	7,38	2,16; 25,17
	Placebo	72	3 (4,2)		
Nein	Nintedanib	509	114 (22,4)	0,97	0,70; 1,35
	Placebo	351	83 (23,7)		
Studie p-Wert Interaktionsterm: 0,0320					
INPULSIS-1	Nintedanib	309	63 (20,4)	0,84	0,55; 1,29
	Placebo	204	49 (24,0)		
INPULSIS-2	Nintedanib	329	83 (25,2)	1,66	1,08; 2,57
	Placebo	219	37 (16,9)		

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis, OR: Odds Ratio, SGRQ: St. George's Respiratory Questionnaire

Hinweise auf Effektmodifikation zeigten sich in den Subgruppen Geschlecht, FVC-Baselinewert (% vom Sollwert) und Bronchodilator-Begleittherapie für den SGRQ-Gesamtscore. Die Effektrichtung war für alle Subgruppen positiv für Nintedanib, jedoch nicht für alle statistisch signifikant. Statistisch signifikante Unterschiede zeigten sich dort jeweils nur in der Subgruppe der Frauen, bei Patienten mit einem schlechteren FVC-Ausgangswert als Prozent vom Sollwert und bei Patienten, die zu Studienbeginn eine Bronchodilator-Begleittherapie erhielten. Für ältere Patienten (≥ 65 Jahre) zeigte die Responderanalyse Effekte zugunsten von Nintedanib und für jüngere Patienten zuungunsten von Nintedanib, ohne statistische Signifikanz in einer der beiden Gruppen (Hinweis auf Effektmodifikation).

SGRQ-I Gesamtscore und Responder

Für den SGRQ-I-Gesamtscore zeigte sich ein Beleg für Effektmodifikation für den FVC Baselinewert (% vom Sollwert), wobei sich für Patienten mit einem schlechteren FVC Baselinewert (% vom Sollwert) eine statistisch signifikante Verbesserung auf dem SGRQ-I zeigte und Patienten mit besserem Wert zum Baseline eine geringe und nicht statistisch signifikante Verbesserung erreichten.

Tabelle 25: Subgruppenanalyse Veränderung SGRQ-I-Gesamtscore

Endpunkt/ Subgruppe	Behandlung	N ¹⁾	N analy- siert ²⁾	MWD (SE)	Behandlungseffekt Nintedanib vs. Placebo	
					MWD (SE)	95 % KI
Veränderung SGRQ-I-Gesamtscore						
FVC (% pred.) p-Wert Interaktionsterm: 0,0256						
≤ 70 %	Nintedanib	207	198	3,33 (0,996)	-3,86 (1,542)	-6,89; -0,83
	Placebo	154	149	7,19 (1,178)		
> 70 %	Nintedanib	431	412	3,57 (0,670)	-0,92 (1,059)	-3,00; 1,15
	Placebo	269	264	4,50 (0,820)		

¹⁾ Gesamtanzahl Patienten

²⁾ Anzahl analysierter Studienteilnehmer

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, MWD: Mittelwertdifferenz, SE: Standardfehler, SGRQ-I: St. George's Respiratory Questionnaire IPF spezifischer Fragebogen

Für die Bronchodilator-Begleittherapie und Studie ergaben sich Belege für Effektmodifikation im Interaktionstest in der Responderanalyse des SGRQ-I. Statistisch signifikante Unterschiede zeigten sich jeweils nur bei Patienten, die zu Studienbeginn eine Bronchodilator-Begleittherapie erhielten, sowie für die Studie INPULSIS-2, nicht aber für INPULSIS-1. Die SGRQ-Responderanalyse zeigte für Patienten mit einem FVC Baselinewert (% vom Sollwert) (≤ 70 %) Effekte zugunsten von Nintedanib, für Patienten mit besserem Ausgangswert wurde kein Unterschied zu Placebo nachgewiesen.

Tabelle 26: Subgruppenanalyse SGRQ-I(≤ -4 Punkte)-Responder

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
SGRQ-I(≤ -4 Punkte)-Responder					
Bronchodilator p-Wert Interaktionsterm: 0,0026					
Ja	Nintedanib	129	37 (28,7)	4,30	1,71; 10,84
	Placebo	72	6 (8,3)		
Nein	Nintedanib	509	113 (22,2)	1,06	0,76; 1,48
	Placebo	351	77 (21,9)		
Studie p-Wert Interaktionsterm: 0,0164					
INPULSIS-1	Nintedanib	309	66 (21,4)	0,89	0,58; 1,36
	Placebo	204	49 (24,0)		
INSULSIS-2	Nintedanib	329	84 (25,5)	1,88	1,20; 2,93
	Placebo	219	34 (15,5)		

Abkürzungen: FVC: forcierte Vitalkapazität, KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis, OR: Odds Ratio, SGRQ-I: St. George's Respiratory Questionnaire IPF spezifischer Fragebogen

Hinweise auf Effektmodifikation zeigten sich in den Subgruppen Geschlecht, Alter, FVC Baselinewert (% vom Sollwert) und Bronchodilator-Begleittherapie für den SGRQ-I-Gesamtscore. Die Effektrichtung war für alle Subgruppen positiv für Nintedanib, jedoch nicht für alle Subgruppen statistisch signifikant. Statistisch signifikante Unterschiede zeigten sich jeweils nur in der Subgruppe der Frauen, bei Patienten mit höherem Alter und bei Patienten, die zu Studienbeginn eine Bronchodilator-Begleittherapie erhielten.

CASA-Q

Für den CASA-Q Husten-Symptome zeigte sich ein Hinweis auf Effektmodifikation in Bezug auf die Einnahme systemischer Kortikosteroide, wobei Patienten, die zum Baseline keine Kortikosteroide erhielten, auf dem CASA-Q Husten-Symptome eine Verbesserung unter Nintedanib verzeichneten und Patienten, die Kortikosteroide erhielten sich verschlechterten, wobei in beiden Gruppen keine statistische Signifikanz gezeigt wurde.

Für den CASA-Q Husten-Impact zeigten sich Hinweise auf Effektmodifikation für Geschlecht, Alter und Bronchodilator-Begleittherapie. Außer in der Subgruppe der Patienten <65 Jahre zeigte sich in allen Subgruppen eine Verbesserung für Patienten, die mit Nintedanib behandelt wurden, wobei nur für Patienten ≥ 65 Jahre und für Patienten, die eine Bronchodilator-Begleittherapie erhielten eine statistische Signifikanz vorlag.

Unerwünschte Ereignisse

Für den Endpunkt Gesamtzahl UE fand sich bezüglich des Subgruppenmerkmals Gewicht ein Beleg für eine Effektmodifikation. Bei Patienten mit einem Gewicht < 65 kg war die Gesamtzahl UE unter Nintedanib gegenüber Placebo stärker erhöht als bei Patienten mit einem Gewicht ≥ 65 kg. In beiden Gruppen war der Unterschied statistisch signifikant zuungunsten von Nintedanib.

Für den Endpunkt UE die zum Therapieabbruch führen, fand sich ein Beleg für eine Effektmodifikation durch das Subgruppenmerkmal Geschlecht, wobei Frauen im Nintedanib-Arm häufiger ein solches Ereignis hatten als Männer und für die Subgruppe der Frauen der Unterschied statistisch signifikant war. Bezüglich der Subgruppenmerkmale Alter, Raucherstatus und Gewicht fand sich für UE, die zum Therapieabbruch führen, Hinweise auf Effektmodifikation. Im Nintedanib-Arm hatten Patienten ≥ 75 Jahre, Patienten die nicht rauchten oder in der Vergangenheit Raucher waren und Patienten, deren Gewicht <65 kg lag, jeweils ein statistisch signifikant erhöhtes Risiko eines UE, dass zu Therapieabbruch führte im Vergleich zu Patienten im Placebo-Arm. In den anderen Subgruppen zeigte der Effektschätzer ebenfalls ein höheres Risiko im Nintedanib-Arm an, die Effekte waren jedoch nicht statistisch signifikant.

Bezüglich der UE, die zum Tod führten, fand sich in der Subgruppenanalyse nach Alter ein Beleg für eine Effektmodifikation. In der Subgruppe der über 75-jährigen Patienten zeigte sich mit einem OR von 0,23 ein statistisch signifikant geringerer Nachteil unter einer Behandlung mit Nintedanib verglichen mit Placebo für UE, die zum Tod führten. Für Raucherstatus und Nierenfunktionseinschränkung zeigten sich Hinweise auf Effektmodifikation mit Effekten

zuungunsten der Nintedanib-Gruppe für Nicht-Raucher und Patienten ohne Nierenfunktionseinschränkung, wobei in keiner der Subgruppen der Effekt statistisch signifikant war.

Tabelle 27: Subgruppenanalyse unerwünschte Ereignisse

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Gesamtzahl UE					
Gewicht p-Wert Interaktionsterm: 0,015					
< 65 kg	Nintedanib	126	124 (98,4)	11,63	2,52;53,53
	Placebo	76	64 (84,2)		
≥ 65 kg	Nintedanib	512	485 (84,7)	1,82	1,07;3,10
	Placebo	347	315 (90,8)		
UE, die zu Therapieabbruch führten					
Geschlecht p-Wert Interaktionsterm: 0,025					
männlich	Nintedanib	507	84 (16,6)	1,28	0,86;1,89
	Placebo	334	45 (13,5)		
weiblich	Nintedanib	131	39 (29,8)	3,35	1,57; 7,14
	Placebo	89	10 (11,2)		
UE, die zum Tod führten					
Alter p-Wert Interaktionsterm: 0,031					
< 65 Jahre	Nintedanib	258	14 (5,4)	0,77	0,33; 1,79
	Placebo	145	10 (6,9)		
65-74 Jahre	Nintedanib	263	18 (6,8)	1,37	0,63; 2,97
	Placebo	216	11 (5,1)		
≥ 75 Jahre	Nintedanib	117	5 (4,3)	0,23	0,08; 0,71
	Placebo	62	10 (16,1)		

Abkürzungen: KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis, OR: Odds Ratio, UE: unerwünschtes Ereignis

UE von besonderem Interesse

Für Diarrhoe zeigte sich ein Beleg für Effektmodifikation in Bezug auf das Subgruppenmerkmal Raucherstatus mit größerem OR für Raucher und Ex-Raucher als für Nicht-Raucher. Für Geschlecht und Nierenfunktionseinschränkungen lagen Hinweise auf Effektmodifikation vor, mit statistisch signifikanten Effekten zuungunsten von Nintedanib in allen Subgruppen, wobei für Patienten im Nintedanib-Arm, die keine Nierenfunktionseinschränkungen hatten und männliche Patienten, das OR höher war als für Patienten mit Nierenfunktionseinschränkungen oder für weibliche Patienten.

Bezüglich des UE Übelkeit ergab sich ein Hinweis auf Effektmodifikation im Subgruppenmerkmal Ethnie, mit höherem OR für asiatische Patienten als für nicht-asiatische

Patienten und statistisch signifikanten Effekten zuungunsten von Nintedanib in beiden Gruppen.

Für das UE Erbrechen lag ein Hinweis auf Effektmodifikation bezüglich des Geschlechtes vor, mit höherem OR für weibliche Patienten im Vergleich zu männlichen Patienten im Nintedanib-Arm und statistischer Signifikanz in beiden Gruppen.

Für das UE Appetitverlust zeigte sich ein Beleg für Effektmodifikation in Bezug auf den Raucherstatus, mit statistisch signifikant höherer Chance ein solches UE zu erleben für Patienten im Nintedanib-Arm, die Nichtraucher waren. Geschlecht und Gewicht zeigten Hinweise auf Effektmodifikation mit statistisch signifikanten, höheren OR für weibliche Patienten im Vergleich zu männlichen Patienten und Patienten mit geringerem Gewicht im Vergleich zu höherem Gewicht im Nintedanib-Arm.

Tabelle 28: Subgruppenanalyse unerwünschte Ereignisse von besonderem Interesse

Endpunkt/ Subgruppe	Behandlung	N	Anzahl Ereignisse, n (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Diarrhoe					
Raucherstatus p-Wert Interaktionsterm: 0,008					
Nicht-Raucher	Nintedanib	174	93 (53,4)	4,04	2,40; 6,80
	Placebo	122	27 (22,1)		
Ex-Raucher/ Raucher	Nintedanib	464	305 (65,7)	9,40	6,58; 13,44
	Placebo	310	51 (16,9)		
Appetitverlust					
Raucherstatus p-Wert Interaktionsterm: 0,025					
Nicht-Raucher	Nintedanib	174	18 (10,3)	6,92	1,58; 30,42
	Placebo	122	2 (1,6)		
Ex-Raucher/ Raucher	Nintedanib	464	50 (10,8)	1,53	0,91; 2,59
	Placebo	310	22 (7,3)		
Blutungen					
Nierenfunktionseinschränkungen p-Wert Interaktionsterm: 0,038					
Ja	Nintedanib	331	38 (11,5)	2,13	1,13; 4,03
	Placebo	244	14 (5,7)		
Nein	Nintedanib	306	28 (9,2)	0,84	0,46; 1,56
	Placebo	178	19 (10,7)		

Abkürzungen: KI: Konfidenzintervall, N: Anzahl der Patienten im Treated Set, n: Anzahl der Patienten im Treated Set mit Ereignis, OR: Odds Ratio, UE: unerwünschtes Ereignis

2.7 Einschätzung zum Ausmaß des Zusatznutzens durch den pU

Der pU leitet aus den Daten der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien den Zusatznutzen für die Zielpopulation für unterschiedliche Endpunkte wie folgt ab:

Endpunkt	Effektschätzer [95 % KI] p-Wert	Beurteilung nach Arzneimittel-Nutzenbewertungsverordnung (AM-Nutzen V)	Ausmaß des Zusatznutzens
Mortalität			
Zeit bis zum Tod			
Zeit bis zum Tod unabhängig von der Ursache	HR: 0,70 [0,43; 1,12] 0,1399	Bisher nicht erreichte Verbesserung des therapielevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht quantifizierbaren Verlängerung der Lebensdauer	nicht quantifizierbar
Zeit bis zum respiratorisch bedingten Tod	HR: 0,74 [0,41; 1,34] 0,3435		
Forcierte Lungenfunktion – FVC – als Surrogat für die Mortalität			
FVC-Abnahme (in ml)	MWD: 109,94 [75,85; 144,03] < 0,0001	Bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens im Sinne einer deutlichen Verringerung in der Abnahme der FVC, als Surrogat für ein verringertes Mortalitätsrisiko.	beträchtlich Wahrscheinlichkeit: Hinweis
FVC Response Schwellenwert 10 % vom Sollwert	OR: 1,577 [1,21; 2,05] 0,0007		
FVC Response Schwellenwert 5 % vom Sollwert	OR: 1,835 [1,43; 2,36] < 0,0001		
Morbidität			
Exazerbationen			
Zeit bis zur ersten adjudizierten akuten Exazerbation	HR: 0,32 [0,16; 0,65] 0,0010	Bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens im Sinne einer bedeutsamen Abschwächung schwerwiegender Symptome.	beträchtlich Wahrscheinlichkeit: Hinweis
Zeit bis zur ersten akuten Exazerbation	HR: 0,64 [0,39; 1,05] 0,0823	Bisher nicht erreichte Verbesserung des therapielevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht quantifizierbaren Reduktion akuter Exazerbationen	nicht quantifizierbar
Gesundheitsbezogene Lebensqualität			
Krankheitsspezifische Lebensqualität			
SGRQ-Gesamtscore	MWD: -1,43 [-3,09; 0,23] 0,0923	Bisher nicht erreichte Verbesserung des therapielevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht quantifizierbaren Verbesserung der Lebensqualität	nicht quantifizierbar
SGRQ-(≤ -4 Punkte)-Responder	OR: 1,186 [0,88; 1,61] 0,2715	Bisher nicht erreichte Verbesserung des therapielevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht weiter quantifizierbaren Verbesserung bzw. geringere Abnahme der Lebensqualität	nicht quantifizierbar
SGRQ-I Gesamtscore	MWD: -1,99 [-3,71; -0,27] 0,0230		
SGRQ-I-(≤ -4 Punkte)-Responder	OR: 1,284 [0,95; 1,74] 0,1091		

Endpunkt	Effektschätzer [95 % KI] p-Wert	Beurteilung nach Arzneimittel-Nutzenbewertungsverordnung (AM-Nutzen V)	Ausmaß des Zusatznutzens
SOBQ-Gesamtscore	MWD: -1,69 [-4,09; 0,70] 0,1648		
CASA-Q Husten-Symptome	MWD: 0,95 [-1,49; 3,38] 0,4450		
CASA-Q Husten-Impact	MWD: 1,73 [-0,46; 3,92] 0,1212		
Allgemeine Lebensqualität			
PGI-C-Responder	OR: 1,328 [1,04; 1,70] 0,0254	Bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens im Sinne einer moderaten spürbaren Verbesserung der Lebensqualität	gering Wahrscheinlichkeit: Hinweis
EQ-5D-VAS	Nintedanib: MWD -2,49 (SD: 17,91) Placebo: MWD -5,73 (SD: 18,37)	Bisher nicht erreichte Verbesserung des therapielevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht weiter quantifizierbaren Verbesserung bzw. geringere Abnahme der Lebensqualität	nicht quantifizierbar
Unerwünschte Ereignisse			
Gesamtzahl UE	OR: 2,44 [1,50; 3,96] < 0,001	Heterogene klinische Relevanz und Doppelzählung von SUE, UEA und UE, die zum Tode führten	Nachteil
SUE	OR: 1,02 [0,78;1,33] 0,894	Kein nachweisbarer Effekt	Zusatznutzen nicht belegt
UE, die zum Therapieabbruch führten (UEA)	OR: 1,60 [1,13;2,26] 0,007	Zunahme eines unerwünschten Ereignisses	Nachteil
UE, die zum Tod führten	OR: 0,78 [0,48;1,28] 0,319	Kein nachweisbarer Effekt	Zusatznutzen nicht belegt

3. Methodische Anmerkungen und Bewertung der Unterlagen

3.1 Zulassungsstatus und Zulassungspopulation von Nintedanib

Nintedanib ist zugelassen zur Behandlung der idiopathischen Lungenfibrose (IPF) bei Erwachsenen. Die empfohlene Dosis beträgt 150 mg Nintedanib zweimal täglich. Die Anwendung der Dosis von 100 mg zweimal täglich wird nur bei Patienten empfohlen, die die Dosis von 150 mg zweimal täglich nicht tolerieren. Eine erneute Aufdosierung auf 150 mg nach Dosisreduktion ist laut Fachinformation möglich.

3.2 Design und Methodik der Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34)

Beide Zulassungsstudien, INPULSIS-1 und INPULSIS-2, sind 52-wöchige, multizentrische, randomisierte, placebokontrollierte, doppelblinde, klinische Studie der Phase III zur Untersuchung der Wirksamkeit sowie des Nutzen/Risiko-Verhältnisses von Nintedanib 150 mg bid im Vergleich zu Placebo bei Patienten mit IPF. Die Zulassungspopulation stimmt weitgehend mit der Studienpopulation überein, wobei nur Patienten in die Studie eingeschlossen worden waren, die FVC-Werte von $\geq 50\%$ des Normalwertes zu Studienbeginn aufwiesen. Die Baselinewerte waren zwischen beiden Studien vergleichbar und jeweils zwischen den Behandlungsarmen weitgehend ausgeglichen. Die Verteilung nach Alterskategorien war etwas unausgeglichen, mit jeweils mehr Patienten in der mittleren Alterskategorie zwischen 65 und <75 Jahren in den Placebo-Armen der Studien und mehr Patienten im Alter zwischen 45 bis <65 Jahre und ≥ 75 Jahre in den Nintedanib-Armen der Studien.

In beiden Studien wurden die Patienten den Therapiearmen im Verhältnis 3:2 zugeordnet (Nintedanib:Placebo). Im Nintedanib-Arm waren insgesamt 640 Patienten eingeschlossen (INPULSIS-1: 309 und INPULSIS-2: 331) und im Placebo-Arm 426 (206 und 220). Randomisierung und Verblindung sind im Studienprotokoll beschrieben und wurden adäquat umgesetzt. Die Zuteilung auf die Behandlungsgruppen Nintedanib und Placebo erfolgte durch ein interaktives Voice-Response-System bzw. Web-Response-System (IXRS). Die Randomisierung war nicht stratifiziert. Die aktive Studienmedikation und Placebo waren äußerlich nicht voneinander unterscheidbar. Wirksamkeit und Sicherheit wurden auf Basis des Treated Set (TS) ausgewertet. Das TS bestand aus den Studienteilnehmern, für die die Einnahme von mindestens einer Dosis der Studienmedikation dokumentiert wurde. In INPULSIS-1 hatten 2 von insgesamt 515 randomisierten Studienteilnehmern keine Dosis der Studienmedikation erhalten, in INPULSIS-2 waren dies 3 von 551 Patienten.

Das Verzerrungspotential auf Studienebene wird für die Studien INPULSIS-1 und INPULSIS-2 als gering eingeschätzt.

Aufgrund der Homogenität beider Zulassungsstudien wurden die Studienergebnisse im vom pU eingereichten Dossier sowohl für die Einzelstudien INPULSIS-1 und -2 dargestellt, als auch die gepoolten Ergebnisse aus den individuellen Patientendaten beider Studien. Für die

gepoolte Analyse wurde ein Fixed-Effect Modell gewählt. Die Ergebnisse der Studien INPULSIS-1 und INPULSIS-2 weichen trotz des homogenen Studiendesigns teilweise voneinander ab, was möglicherweise den Einsatz eines Random-Effects Modells gerechtfertigt hätte. Eine Auswertung mit Random Effects Modell liegt nicht vor.

3.3 Wirksamkeit

Mortalität

Zeit bis zum Tod und Zeit bis zum respiratorisch bedingten Tod

Für die Zeit bis zum Tod und für die Zeit bis zum respiratorisch bedingten Tod innerhalb von 52 Wochen zeigte sich weder in den Einzelstudien noch in der gepoolten Analyse ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen, wobei die Anzahl an Ereignissen in beiden Studien und dort jeweils in beiden Behandlungsgruppen relativ gering war.

Primärer Endpunkt – Jährliche FVC-Abnahme (ml)

FVC war der primäre Endpunkt beider INPULSIS-Studien. Im Dossier wird der Endpunkt als Surrogat für Mortalität herangezogen. In verschiedenen multivariaten Modellen konnte eine Korrelation zwischen FVC und Mortalität gezeigt werden [11,12,35,42,32]. Es liegen keine Modelle zur Validierung des Surrogats FVC vor, in denen Intervention als unabhängige Größe betrachtet wurde und der Zusammenhang zwischen Effekten auf das Surrogat Veränderung der FVC und Effekten auf den patientenrelevanten Endpunkt Mortalität gezeigt wurde. Interventionsstudien, in denen ein solcher Zusammenhang gezeigt werden konnte, liegen ebenfalls nicht vor. FVC wird als nicht ausreichend validiertes Surrogat für Mortalität angesehen.

Die jährliche Abnahme der FVC wurde mit einem statistischen Modell unter der Berücksichtigung von Geschlecht, Alter und Körpergröße berechnet, um die Vergleichbarkeit zwischen den Patienten zu gewährleisten. Der primäre Endpunkt, jährliche Abnahme der FVC zeigte in beiden INPULSIS-Studien und in der gepoolten Auswertung statistisch signifikante Effekte zugunsten der mit Nintedanib behandelten Patienten. Über 52 Wochen nahm die FVC in der Nintedanib-Gruppe kontinuierlich weniger ab als in der Placebo-Gruppe.

Zusätzlich zum primären Endpunkt wurden FVC-Responder analysiert. Als FVC-Responder (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert) war ein Patient definiert, dessen FVC in Woche 52 gegenüber dem individuellen Messwert zu Studienbeginn um nicht mehr als 10 % bzw. 5 % (bezogen auf den Sollwert) abfiel. Für alle Responderanalysen ergaben sich in den Studien und der gepoolten Analyse statistisch signifikante Effekte zugunsten der mit Nintedanib behandelten Patienten, bis auf die Responderanalyse mit einer Veränderung von > 10 % in INPULSIS-2, wo kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen gezeigt werden konnte.

Morbidität

Zeit bis zur ersten akuten Exazerbation

Eine akute Exazerbation war definiert als eine vom Prüfarzt festgestellte erhebliche akute Verschlechterung der klinischen Situation des Patienten innerhalb eines Monats, die auf keine andere Ursache als die zugrunde liegende IPF zurückgeführt werden konnte. Eine akute Exazerbation lag gemäß Studienprotokollen vor, sofern sämtliche der folgenden Kriterien erfüllt waren:

- Unerklärtes Neuauftreten oder Zunahme von Dyspnoe innerhalb der letzten 30 Tage
 - Neue pulmonale Infiltrate im Thorax-Röntgenbild und/oder neue parenchymale Anomalien ohne Pneumothorax oder Pleuraerguss (neue Milchglasverschattungen) in der hochauflösenden Computer-Tomographie (HRCT) seit der letzten Visite
- Ausschluss von Infektionen
- Ausschluss anderer möglicher Erklärungen, wie beispielsweise
 - Linksherzversagen
 - Lungenembolie
 - akute Lungenschädigung erkennbarer Ursache

Die vom Prüfarzt festgestellten Exazerbationen wurden nachträglich durch ein unabhängiges Bewertungskomitee (IAC) adjudiziert. Jedes vom Prüfarzt dokumentierte Ereignis wurde durch das IAC einer der Kategorien bestätigte Exazerbation, vermutete Exazerbation oder keine Exazerbation zugeordnet. Bestätigte oder vermutete akute Exazerbationen wurden dem Endpunkt adjudizierte akute Exazerbationen als Ereignis zugeordnet. Bestätigte Exazerbation war definiert sofern alle oben genannten Kriterien der Exazerbation erfüllt waren. Vermutete Exazerbation war definiert als klinische Verschlechterungen der IPF unbekannter Ursache, die nicht vollständig die oben genannten Kriterien der akuten Exazerbation der IPF erfüllten. Eine erhebliche Verschlechterung der Dyspnoe kann grundsätzlich als patientenrelevant betrachtet werden, jedoch ist die Validität des Endpunkts aufgrund der unzureichenden Darstellung der Messung und Definition der Verschlechterung der Dyspnoe unklar. Eine alleinige Veränderung der Bildgebung wird ohne Nachweis einer Auswirkung auf die Morbidität des Patienten ebenfalls nicht als patientenrelevant betrachtet.

Für die Zeit bis zur ersten akuten Exazerbation zeigte sich in INPULSIS-2 ein statistisch signifikanter Unterschied zugunsten von Nintedanib, nicht aber in INPULSIS-1 und auch nicht in der gepoolten Analyse. In INPULSIS-1 war der Effekt geringfügig nachteilig für Nintedanib, ohne statistische Signifikanz zu erreichen. Die Zeit bis zur ersten adjudizierten akuten Exazerbation war, bei einem numerischen Vorteil für Nintedanib, in INPULSIS-1 nicht statistisch signifikant unterschiedlich zwischen Nintedanib und Placebo, in INPULSIS-2 und der gepoolten Analyse konnte statistische Signifikanz nachgewiesen werden.

Patient's Global Impression of Change (PGI-C) Responderanalyse

Mit Hilfe des PGI-C gibt der Patient seinen subjektiven Eindruck zur Verbesserung oder Verschlechterung seiner respiratorischen Kondition im Vergleich zum Referenzzeitpunkt bei Studienbeginn an. Das Instrument besteht aus einer Frage: "Check the one number that best

describes how your respiratory condition is now, compared with how it was before you began taking medication in this study.” Die Einschätzung erfolgte durch die Studienteilnehmer auf einer 7-Punkte-Likert-Skala von eins (sehr viel besser) bis sieben (sehr viel schlechter) von Baseline bis Woche 52. Als PGI-C-Responder wurde a priori ein Patient definiert, der bei der Erhebung zu Woche 52 verglichen mit seinem subjektiv erinnerten Befinden bei Studienbeginn keine Kategorie der Verschlechterung seines Zustands angegeben hatte.

In den Studien INPULSIS-1 und 2 konnte ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen nicht nachgewiesen werden, wobei ein numerischer Vorteil für Nintedanib bestand. Für die gepoolte Analyse zeigte sich ein statistisch signifikanter Effekt zugunsten von Nintedanib.

Generell ist das Instrument nachvollziehbar, jedoch ist eine Einschätzung der Krankheitsschwere mit einem Zeithorizont von 52 Wochen in die Vergangenheit als nicht robust einzuschätzen. „Keine Verschlechterung“ auf dem PGI-C ist als Responderkriterium im Sinne eines klinisch relevanten Unterschiedes ist nicht validiert.

Veränderung der EQ-5D VAS zwischen Baseline und Woche 52

Die Veränderung des Wertes auf der visuellen Analogskala EQ-5D VAS zwischen Studienbeginn und Woche 52 wurde bestimmt. Der Wert auf dem EQ-5D-VAS sank in beiden Studien und jeweils in beiden Behandlungsgruppen im Laufe der 52 Wochen. Die Verschlechterung des Gesundheitszustandes war im etwas Nintedanib-Arm geringer als im Placebo-Arm. Responderanalysen liegen nicht vor.

3.3.1 Zusammenfassende Einschätzung zur Wirksamkeit

Die Ergebnisse der Studien INPULSIS-1 und INPULSIS-2 sind in Hinblick auf den primären Endpunkt beider Studien, die Veränderung der FVC, konsistent. Der zeitliche Verlauf der FVC-Abnahme war in beiden Studien vergleichbar. Der FVC Wert nahm in den Nintedanib-Armen langsamer ab als in den Placebo-Armen und der Effekt war kontinuierlich über 52 Wochen zu beobachten, so dass es sich offensichtlich nicht um einen einmaligen Effekt von Nintedanib, der über längere Zeit anhält, handelt.

Die Ergebnisse der Studien INPULSIS-1 und INPULSIS-2 weichen im zentralen sekundären Endpunkt akute Exazerbation voneinander ab. Die unterschiedlichen Ergebnisse wurden im Zuge des Zulassungsprozesses thematisiert. Die EMA forderte für beide Studien Subgruppenanalysen zu den zentralen sekundären Endpunkten, um mögliche Auslöser für die unterschiedlichen Ergebnisse zu beleuchten. Die Ergebnisse der Subgruppenanalysen waren zwischen den Studien konsistent und keiner der untersuchten Subgruppen zeigte einen nachteiligen Effekt von Nintedanib. Die EMA geht davon aus, dass das Versagen der Studie INPULSIS-1 in Bezug auf den Nachweis eines Nutzens von Nintedanib gegenüber Placebo auf die Zeit bis zur ersten akuten Exazerbation, wahrscheinlich auf die zugrundeliegende Variabilität und mangelnde Sensibilität des Endpunkts zurückzuführen sei.

Ergebnisse zur Morbidität beschränken sich in den vorliegenden Studien auf die Zeit bis zur ersten (adjudizierten) akuten Exazerbation und auf den PGI-C Fragebogen sowie die EQ-5D

VAS, die jeweils eine Frage zur Einschätzung des Gesundheitszustandes beinhalten. Ein weiterer Morbiditätsendpunkt zur körperlichen Leistungsfähigkeit der Patienten wäre wünschenswert gewesen. In der Dosisfindungsstudie zu Nintedanib TOMORROW (1199.30) wurde zusätzlich die 6MWD untersucht. Es handelte sich bei dieser Dosisfindungsstudie um eine 52-wöchige randomisierte, placebokontrollierte, doppelblinde 5-armige Studie der Phase II, mit dem Ziel der Untersuchung der Wirksamkeit und Sicherheit von oral verabreichtem Nintedanib in unterschiedlichen Dosierungen von 50 mg einmal täglich, 50 mg zweimal täglich, 100 mg zweimal täglich und 150 mg zweimal täglich bei 428 Patienten mit IPF. Anders als in den INPULSIS-Studien und anders als in der Fachinformation zu Nintedanib festgelegt, war in der TOMORROW-Studie nach Dosisreduktion keine erneute Reeskalation möglich, weshalb TOMORROW nicht im Dossier und nicht in der vorliegenden Nutzenbewertung berücksichtigt wurde. Ein- und Ausschlusskriterien waren mit den INPULSIS-Studien vergleichbar und auch die Baselinecharakteristika waren weitgehend vergleichbar, wobei mehr Patienten kaukasischer Herkunft in TOMORROW eingeschlossen waren als in INPULSIS (ca. 75 % vs. 57 %). In der Nintedanib 150 mg bid Gruppe wurden 85 Patienten behandelt und 85 Patienten erhielten Placebo. Der primäre Endpunkt in TOMORROW war ebenfalls die Veränderung der FVC und die Ergebnisse für die 150 mg Nintedanib-Gruppe im Vergleich zu Placebo waren konsistent mit den Ergebnissen der beiden INPULSIS Studien (Mittelwertdifferenz zwischen Nintedanib und Placebo von 131 mL). Für die Zeit bis zur ersten akuten Exazerbation (gleiche Definition wie in den INPULSIS-Studien) ergab sich in der TOMORROW Studie für die 150 mg Nintedanib-Gruppe ein statistisch signifikanter Vorteil zugunsten von Nintedanib (HR 0,16; 95 % KI 0,035-0,711; p=0,0161). Kein Unterschied konnte für die 6MWD gezeigt werden.

3.4 Lebensqualität

Veränderung St. George's Respiratory Questionnaire (SGRQ)-Gesamtscore zwischen Baseline und Woche 52 und SGRQ-Responderanalyse

Das mehrdimensionale Erhebungsinstrument SGRQ ist ein validiertes Instrument zur Abschätzung der Beeinträchtigung der gesundheitsbezogenen Lebensqualität bei erwachsenen Patienten mit chronischen Erkrankungen der Atemwege und umfasst 50 Fragen. Die Veränderung des SGRQ-Gesamtscores wurde als Veränderung über die Zeit von Baseline bis Woche 52 gemessen. Aus den Antworten wird ein Grad der Beeinträchtigung der krankheitsspezifischen Lebensqualität im Bereich von 0 (keine Beeinträchtigung) bis 100 (maximale Beeinträchtigung) ermittelt. Responderkriterium war eine Verbesserung von mindestens 4 Punkten.

Die Veränderung des SGRQ-Gesamtscore war neben der Zeit bis zur akuten Exazerbation der zentrale sekundäre Endpunkt in den beiden INPULSIS-Studien. In INPULSIS-2 zeigte sich ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe, in der INPULSIS-1 Studie und in der gepoolten Analyse war kein Unterschied zwischen den Behandlungsgruppen nachweisbar. In der Responderanalyse zeigte sich kein statistisch signifikanter Unterschied in INPULSIS-1, wobei die Effektrichtung leicht zuungunsten von Nintedanib ausfiel, und auch kein statistisch signifikanter Unterschied in der gepoolten Analyse. In INPULSIS-2 zeigte sich ein statistisch signifikanter Vorteil von Nintedanib im Vergleich zu Placebo.

Veränderung SGRQ-I Gesamtscore zwischen Baseline und Woche 52 und SGRQ-I-Responderanalyse

Der SGRQ-I ist ein IPF-spezifischer Fragebogen, der aus dem SGRQ Fragebogen entwickelt wurde. Der pU berichtete nicht die Ergebnisse aus dem 26 Items umfassenden SGRQ-I Fragebogen, sondern wandte einen Algorithmus an, der die Berechnung des SGRQ-I Scores aus dem SGRQ-Score ermöglicht. Die Validität des transformierten SGRQ-I Scores ist kritisch zu betrachten, da nicht hinreichend geklärt ist, inwiefern Aussagen zum SGRQ-I Score ermittelt durch den IPF-spezifischen Fragebogen auf die transformierten SGRQ-I Score übertragbar sind. Um eine Responder-Analyse durchführen zu können, wurde in Anlehnung an die MID für einen klinisch relevanten Effekt bzgl. des SGRQ-Gesamtscores auch für den SGRQ-I ein Schwellenwert von 4 Punkten festgelegt. Die Responderanalyse des SGRQ-I erfolgte post-hoc. Für das Responderkriterium liegt keine Validierung vor.

In INPULSIS-2 und in der gepoolten Analyse zeigte sich für den Gesamtscore der IPF-spezifischen Version des SGRQ ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe, in der INPULSIS-1 Studie war kein Unterschied zwischen den Behandlungsgruppen nachweisbar. In INPULSIS-2 zeigte sich in der Responderanalyse ein statistisch signifikanter Vorteil zugunsten der Nintedanib-Gruppe. In INPULSIS-1 und der gepoolten Analyse war kein Unterschied in der SGRQ-I-Responderanalyse zwischen den Behandlungsgruppen nachweisbar.

Veränderung SOBQ-Gesamtscore, Veränderung CASA-Q Husten-Symptome und Veränderung CASA-Q Husten-Impact zwischen Baseline und Woche 52

Der SOBQ Fragebogen erfasst Kurzatmigkeit an einem durchschnittlichen Tag innerhalb der letzten Woche bei verschiedenen Tätigkeiten (z.B. beim Laufen, in Ruhe, beim Zähneputzen, beim Abwaschen etc.). Der CASA-Q Fragebogen besteht ursprünglich aus 4 Domänen: Husten-Symptome, Husten-Impact, Sputum-Symptome und Sputum-Impact. In den INPULSIS Studien wurden ausschließlich die beiden Husten-Domänen berücksichtigt.

Für den SOBQ-Fragebogen und die Husten-Symptom und Husten-Impact Domänen des CASA-Q Fragebogens zeigten sich weder in den INPULSIS-Studien noch in den gepoolten Analysen statistisch signifikante Unterschiede.

3.4.1 Zusammenfassende Einschätzung zur Lebensqualität

Tendenziell zeigt sich für alle Lebensqualitätsinstrumente eine weniger starke Abnahme der Lebensqualität unter Therapie mit Nintedanib im Vergleich zu Placebo, statistische Signifikanz konnte jedoch nur in der INPULSIS-2 Studie für die Endpunkte Veränderung des SGRQ und SGRQ-I Gesamtscores und für die Responderanalysen des SGRQ und SGRQ-I sowie für die gepoolte Analyse der Veränderung des SGRQ-I Gesamtscores gezeigt werden. Für INPULSIS-1 ergab sich in keinem der Endpunkte zur Lebensqualität ein statistisch signifikanter Vorteil von Nintedanib.

In beiden INPULSIS Studien wurden der primäre Endpunkt und die beiden zentralen sekundären Endpunkte hierarchisch getestet, wobei zunächst der primäre Endpunkt, dann die

Dossierbewertung für Orphan Drugs

Nintedanib

Veränderung vom Ausgangswert im SGRQ Gesamtscore nach 52 Wochen und dann die Zeit bis zur ersten akuten Exazerbation untersucht wurden. Da in INPULSIS-1 für die Veränderung des SGRQ Gesamtscores kein statistisch signifikanter Unterschied gezeigt werden konnte, sind die Ergebnisse zum Endpunkt akute Exazerbation in der INPULSIS-1 Studie formal rein deskriptiv zu interpretieren.

Wie schon für die Zeit bis zur ersten akuten Exazerbation forderte die EMA im Zuge des Zulassungsprozesses eine Subgruppenanalyse für die Veränderung des SGRQ, um Gründe für die abweichenden Ergebnisse zwischen den beiden INPULSIS-Studien zu ermitteln. Wie auch bei der Subgruppenanalyse der Exazerbation waren die Ergebnisse zwischen den Studien konsistent und keiner der untersuchten Subgruppen zeigte einen nachteiligen Effekt von Nintedanib, so dass die Zulassungsbehörde die Unterschiede auf die zugrundeliegende Variabilität und mangelnde Sensibilität des Endpunktes zurückführte.

3.5 Sicherheit

Die Analyse der unerwünschten Ereignisse (UE) basierte auf der Anzahl der Patienten mit einem UE, das ab der ersten Dosis der Studienmedikation bis 28 Tage nach Einnahme der letzten Dosis der Studienmedikation aufgetreten war. Die Anzahl an Patienten mit mindestens einem UE war in beiden INPULSIS-Studien als auch in der gepoolten Analyse höher in der Nintedanib-Gruppe im Vergleich zur Placebo-Gruppe (95,5 % vs. 89,6 % für die gepoolte Analyse). Der Unterschied in der gepoolten Analyse war statistisch signifikant zuungunsten von Nintedanib. Die häufigsten UE (>40 %) bezogen auf Systemorganklassen (SOC) waren in beiden Studien Erkrankungen des Gastrointestinaltrakts, die jeweils im Nintedanib-Arm deutlich häufiger auftraten als im Placebo-Arm. Die Preferred Terms (PT) Diarrhoe, Übelkeit und Erbrechen waren jeweils die häufigsten PT in der SOC Erkrankungen des Gastrointestinaltrakts und traten in beiden Studien jeweils mehr als dreimal so häufig im Nintedanib-Arm auf wie im Placebo-Arm.

Als UE von besonderem Interesse wurden in den Studien Diarrhoe, Übelkeit, Erbrechen, Dehydrierung, Gewichtsverlust und Appetitverlust betrachtet. In INPULSIS-1 erlebten 61,5 % der Patienten in der Nintedanib-Gruppe und 18,6 % der Patienten in der Placebo-Gruppe mindestens 1 UE des PT Diarrhoe (INPULSIS-2: 63,2 % vs. 18,3 %). Laut Angaben im Dossier konnten 78,6 % der Studienteilnehmer mit Diarrhoe beider gepoolter Studien die Therapie mit der Normaldosis fortsetzen, weitere 14,3 % der Studienteilnehmer mit reduzierter Dosis. Bei 7,0 % der Studienteilnehmer mit Diarrhoe (entsprechend 4,4 % aller Studienteilnehmer der Nintedanib-Gruppe) kam es aufgrund von Diarrhoe zu einem endgültigen Therapieabbruch. Die Ereignis-Profile für andere UE von besonderem Interesse (Übelkeit, Erbrechen, Gewichtsabnahme und Appetitlosigkeit) waren ähnlich wie bei Diarrhoe. Die meisten prädefinierten UE von besonderem Interesse waren leicht bis mittelschwer. Im Nintedanib-Arm der INPULSIS-1 Studie traten vereinzelt Fälle von schwerwiegenden UE in den PT verminderter Appetit, Diarrhoe, Erbrechen und Gewichtsabnahme auf. In den meisten Fällen erholten sich die Patienten von diesen Ereignissen ohne dauerhafte Dosisreduktion oder vorzeitige Beendigung der Behandlung. In der gepoolten Analyse waren alle a priori

definierten UE von besonderem Interesse statistisch signifikant unterschiedlich zuungunsten von Nintedanib (Tabelle 20).

Zusätzlich zu den a priori definierten UE berichtete der pU im Dossier weitere UE als UE von besonderem Interesse. In der gepoolten Analyse ergaben sich statistisch signifikante Unterschiede zuungunsten von Nintedanib für die PT Leberenzymhöhung, Bilirubinerhöhung, ALKP-Erhöhung sowie für die SOC arterielle thromboembolische Ereignisse.

UE, die zu einem Abbruch der Therapie führten, kamen in der INPULSIS-1 Studie im Nintedanib-Arm häufiger vor als im Placebo-Arm (21,0 % vs. 10,8 %). In INPULSIS-2 war die Inzidenz ausgeglichener (17,6 % vs. 15,1 %). Der Unterschied in der gepoolten Analyse war statistisch signifikant zuungunsten von Nintedanib. Im Nintedanib-Arm brachen die meisten Patienten die Therapie aufgrund von Erkrankungen des Gastrointestinaltrakts ab (INPULSIS-1: 8,4 %; INPULSIS-2: 6,4 %) und in dieser SOC vor allem aufgrund des PT Diarrhoe (4,5 % und 4,3 %).

3.5.1 Zusammenfassende Einschätzung zur Sicherheit

Die SOC Kategorisierung legt nahe, dass durch Nintedanib Erkrankungen des Gastrointestinaltrakts hervorgerufen werden. Der höhere Anteil an Patienten mit Gewichtsverlust ist vermutlich eine Folge der Tolerabilitäts- und Sicherheitsproblematik von Nintedanib im Gastrointestinaltrakt. Insbesondere Diarrhoen wurden in den Nintedanib-Gruppen deutlich häufiger beobachtet.

Zusätzlich wurden vermehrt Transaminaseerhöhungen beobachtet, wodurch ein Potential für Lebererkrankungen besteht. Laut Fachinformation wird bei einer Transaminaseerhöhungen (AST, ALT) von $> 3 \times \text{ULN}$ eine Dosisreduktion oder Unterbrechung der Therapie mit Nintedanib empfohlen. Nach Rückgang der Transaminasen auf das Ausgangsniveau kann die Behandlung mit Nintedanib in der vollen Dosis (150 mg zweimal täglich) wieder aufgenommen oder zunächst wieder mit einer reduzierten Dosis (100 mg zweimal täglich) begonnen werden, die anschließend auf die volle Dosis erhöht werden kann. Bei jeglichen Erhöhungen von Leberwerten, die mit klinischen Anzeichen oder Symptomen einer Leberschädigung, z. B. Gelbsucht, einhergehen, sollte die Behandlung mit Nintedanib dauerhaft abgesetzt werden.

UE die mit dem MedDRA Preferred Term (PT) IPF kodiert waren, schlossen Verschlechterung der IPF und Exazerbation der IPF ein. Diese Ereignisse wurden auch in der Analyse der Wirksamkeit berücksichtigt, so dass zum Teil eine Doppelerfassung von Ereignissen vorliegt.

4. Anforderungen an eine qualitätsgesicherte Anwendung

Die Vorgaben der Fachinformation sind zu berücksichtigen.

Die Behandlung mit Nintedanib (Ofev®) sollte nur von Ärzten begonnen werden, die Erfahrung mit der Diagnose und Behandlung der idiopathischen Lungenfibrose (IPF) haben.

5. Zusammenfassung der Bewertung

Nintedanib ist zugelassen für die Behandlung der idiopathischen Lungenfibrose (IPF) bei Erwachsenen. Die Nutzenbewertung von Nintedanib basiert auf den zulassungsbegründenden Studien INPULSIS-1 (1199.32) und INPULSIS-2 (1199.34). Es handelt sich bei den INPULSIS-Studien um 52-wöchige, multizentrische, randomisierte placebokontrollierte, doppelblinde klinische Studie der Phase III zur Untersuchung der Wirksamkeit sowie des Nutzen/Risiko-Verhältnisses von Nintedanib 150 mg bid im Vergleich zu Placebo bei Patienten mit IPF. Die individuellen Patientendaten der beiden INPULSIS-Studien wurden für das Dossier gepoolt und metaanalytisch ausgewertet.

Die Ergebnisse der INPULSIS-Studien werden in der folgenden Tabelle unter Berücksichtigung des Verzerrungspotentials zusammengefasst. Es wird dargestellt, ob unter Nintedanib ein statistisch signifikant positiver Effekt bei hohem oder unklarem (↑) oder niedrigem (↑↑) Verzerrungspotential bzw. ein statistisch signifikant negativer Effekt bei hohem oder unklarem (↓) oder niedrigem (↓↓) Verzerrungspotential bzw. kein Unterschied (↔) gezeigt werden konnte. Zusätzlich werden die Ergebnisse der Meta-Analyse dargestellt.

Tabelle 29: Zusammenfassende Darstellung der Studienergebnisse

Endpunkt	INPULSIS-1 Nintedanib (N=309) vs. Placebo (N=204)	INPULSIS-2 Nintedanib (N=329) vs. Placebo (N=219)	Effekt INPULS IS-1/ INPULS IS-2	Meta-Analyse Nintedanib (N=638) vs. Placebo (N=423) ¹⁾
Mortalität				
Tod, unabhängig von der Ursache, n (%)	13 (4,2) vs. 13 (6,4)	22 (6,7) vs. 20 (9,1)	↔/↔ ¹⁵⁾	35 (5,5) vs. 33 (7,8)
Respiratorisch bedingter Tod, n (%)	10 (3,2) vs. 10 (4,9)	14 (4,3) vs. 11 (5,0)	↔/↔ ¹⁵⁾	24 (3,8) vs. 21 (5,0)
Primärer Endpunkt der Studien: Jährliche FVC-Abnahme (ml), MWD (SE) ²⁾	114,65 (15,327) vs. 239,91 (18,709)	113,59 (15,726) vs. 207,32 (19,309)	↑↑/↑↑	113,59 (10,984) vs. 223,53 (13,448)
Morbidität				
Akute Exazerbationen ³⁾ , 52 Wochen, n (%) ⁴⁾	19 (6,1) vs. 11 (5,4)	12 (3,6) vs. 21 (9,6)	↔/↑↑ ¹⁵⁾	31 (4,9) vs. 32 (7,6)
PGI-C-Responder, n (%) ⁵⁾	188 (60,8) vs. 112 (54,9)	203 (61,7) vs. 118 (53,9)	↔/↔	391 (61,3) vs. 230 (54,4)
Lebensqualität				
Veränderung SGRQ-Gesamtscore, MWD (SE) ^{6),7),8)}	4,34 (0,799) vs. 4,39 (0,960)	2,80 (0,730) vs. 5,48 (0,891)	↔/↑↑	3,53 (0,540) vs. 4,96 (0,654)
SGRQ-(≤ -4 Punkte)-Responder, n (%) ^{7),9)}	63 (20,4) vs. 49 (24,0)	83 (25,2) vs. 37 (16,9)	↔/↑↑	146 (22,9) vs. 86 (20,3)
Veränderung SOBQ-Gesamtscore, MWD (SE) ^{6),8),10)}	6,73 (1,113) vs. 7,61 (1,376)	6,69 (1,073) vs. 9,07 (1,300)	↔/↔	6,71 (0,773) vs. 8,40 (0,945)

Endpunkt	INPULSIS-1 Nintedanib (N=309) vs. Placebo (N=204)	INPULSIS-2 Nintedanib (N=329) vs. Placebo (N=219)	Effekt INPULS IS-1/ INPULS IS-2	Meta-Analyse Nintedanib (N=638) vs. Placebo (N=423) ¹⁾
Veränderung CASA-Q Husten-Symptome, MWD (SE) ^{6),11),12)}	-0,76 (1,136) vs. -0,52 (1,400)	-0,33 (1,087) vs. -2,38 (1,325)	↔/↔	-0,57 (0,785) vs. -1,52 (0,962)
Veränderung CASA-Q Husten-Impact, MWD (SE) ^{6),11),12)}	-2,36 (1,006) vs. -4,00 (1,240)	-2,58 (0,991) vs. -4,39 (1,209)	↔/↔	-2,48 (0,705) vs. -4,21 (0,865)
Sicherheit				
Gesamtzahl UE, n (%)	298 (96,4) vs. 181 (88,7)	311 (94,5) vs. 198 (90,4)	↓↓/↔	609 (95,5) vs. 379 (89,6)
UE, die zu Therapieabbruch führten, n (%)	65 (21,0) vs. 22 (10,8)	58 (17,6) vs. 33 (15,1)	↓↓/↔	123 (19,3) vs. 55 (13,0)
Diarrhoe, n (%) ¹³⁾	190 (61,5) vs. 38 (18,6)	208 (63,2) vs. 40 (18,3)	↓↓/↓↓	398 (62,4) vs. 78 (18,4)
Übelkeit, n (%) ¹³⁾	70 (22,7) vs. 12 (5,9)	86 (26,1) vs. 16 (7,3)	↓↓/↓↓	156 (24,5) vs. 28 (6,6)
Erbrechen, n (%) ¹³⁾	40 (12,9) vs. 4 (2,0)	34 (10,3) vs. 7 (3,2)	↓↓/↓↓	74 (11,6) vs. 11 (2,6)
Leberenzymerrhöhung (AST, ALT) ≥ 3x ULN, n (%) ¹⁴⁾	k.A.	k.A.	-	32 (5,0) vs. 3 (0,7)

¹⁾ **fett:** statistisch signifikant

²⁾ FVC wird als nicht ausreichend validiertes Surrogat für Mortalität angesehen. Da die jährliche Abnahme des FVC primärer Endpunkt der Studien war, wird der Endpunkt ergänzend dargestellt, jedoch nicht zur Bestimmung des Zusatznutzens herangezogen.

³⁾ Vom Prüfarzt festgestellt

⁴⁾ Akute Exazerbationen lagen vor, sofern die Patienten ein Neuaufreten oder eine Verschlechterung von Dyspnoe erfuhren und sich neue Auffälligkeiten in der Bildgebung zeigten und Infektionen sowie andere mögliche Erklärungen, wie beispielsweise Linksherzversagen, Lungenembolie oder akute Lungenschädigung erkennbarer Ursache ausgeschlossen werden konnten

⁵⁾ Mit Hilfe der Patient's Global Impression of Change (PGI-C) gibt der Patient seinen subjektiven Eindruck zur Verbesserung oder Verschlechterung seiner respiratorischen Kondition im Vergleich zum Referenzzeitpunkt bei Studienbeginn an. Als PGI-C-Responder wurde a priori ein Patient definiert, der bei der Erhebung zu Woche 52 verglichen mit seinem subjektiv erinnerten Befinden bei Studienbeginn keine Kategorie der Verschlechterung seines Zustands angegeben hatte.

⁶⁾ Veränderung zwischen Baseline und Woche 52

⁷⁾ Krankheitsspezifisches Instrument zur Abschätzung der Beeinträchtigung der gesundheitsbezogenen Lebensqualität bei erwachsenen Patienten mit chronischen Erkrankungen der Atemwege

⁸⁾ Ein höherer Wert steht für eine stärkere Beeinträchtigung der gesundheitsbezogenen Lebensqualität

⁹⁾ Als SGRQ-(≤ -4 Punkte)-Responder wurde a priori ein Patient definiert, dessen SGRQ-Gesamtscore in Woche 52 gegenüber dem individuellen Messwert zu Studienbeginn um mindestens 4 Punkte abfiel

¹⁰⁾ Der SOBQ Fragebogen erfasst die Kurzatmigkeit der Befragten an einem durchschnittlichen Tag innerhalb der letzten Woche bei verschiedenen Tätigkeiten

¹¹⁾ Instrument zur Erfassung der Symptome und Beeinträchtigung der Aktivitäten des täglichen Lebens durch Husten und Sputum. In den INPULSIS-Studien wurden nur die Husten-spezifischen Domänen betrachtet

¹²⁾ Ein niedrigerer Wert steht für eine stärkere Beeinträchtigung der gesundheitsbezogenen Lebensqualität

¹³⁾ Prädefinierte UE von besonderem Interesse

¹⁴⁾ Laut Fachinformation wird bei einer Transaminaseerhöhungen (AST, ALT) von > 3xULN eine Dosisreduktion oder Unterbrechung der Therapie mit Nintedanib empfohlen

¹⁵⁾ Überlebenszeitanalyse bis Woche 52

Abkürzungen: ALT: Alanin-Aminotransferase, AST: Aspartat-Aminotransferase CASA-Q: Cough and Sputum Assessment Questionnaire, FVC: forcierte Vitalkapazität, IPF: Idiopathische Lungenfibrose, k.A.: Keine Angabe, MWD: Mittelwertdifferenz, N: Gesamtanzahl Treated Set, n: Anzahl der Studienteilnehmer im Treated Set mit Ereignis, PGI-C: Patient's Global Impression of Change, SE: Standardfehler, SGRQ: St. George's Respiratory Questionnaire, SOBQ: Shortness of Breath Questionnaire ULN: Upper Limit of Normal (oberer Normgrenzwert)

Referenzen

1. **Archibald CJ, Guidotti TL.** Degree of objectively measured impairment and perceived shortness of breath with activities of daily living in patients with chronic obstructive pulmonary disease. *Canadian Journal of Rehabilitation* 1987; 1 (1): 45-54.
2. **Azuma A, Nukiwa T, Tsuboi E, Suga M, Abe S, Nakata K, Taguchi Y, Nagai S, Itoh H, Ohi M, Sato A, Kudoh S.** Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2005; 171 (9): 1040-7.
3. **Behr J.** The diagnosis and treatment of idiopathic pulmonary fibrosis. *Deutsches Arzteblatt international* 2013; 110 (51-52): 875-81.
4. **Behr J, Günther A, Ammenwerth W, Bittmann L, Bonnet R, Buhl R, Eickelberg O, Ewert R, Gläser S, Gottlieb J, Grohe C, Kreuter M, Kroegel C, Markart P, Neurohr C, Pfeifer M, Prasse A, Schönfeld N, Schreiber J, Sitter, H., Theegarten D, Theile A, Wilke A, Witt C, Worth H, Zabel P, Müller-Quernheim J, Costabel U.** S2K-Leitlinie zur Diagnostik und Therapie der idiopathischen Lungenfibrose. *Pneumologie* 2013; 67 (02): 81-111.
5. **Boehringer Ingelheim Pharma GmbH & Co.KG.** Fachinformation Nintedanib (Ofev) 100 mg Weichkapseln (Stand: Januar 2015). 2015;
6. **Boehringer Ingelheim Pharma GmbH & Co.KG.** Fachinformation Nintedanib (Ofev) 150 mg Weichkapseln (Stand: Januar 2015). 2015;
7. **Brown AW, Fischer CP, Shlobin OA, Buhr RG, Ahmad S, Weir NA, Nathan SD.** Outcomes after hospitalization in idiopathic pulmonary fibrosis: a cohort study. *Chest* 2015; 147 (1): 173-9.
8. **Collard HR, Loyd JE, King TE, Jr., Lancaster LH.** Current diagnosis and management of idiopathic pulmonary fibrosis: a survey of academic physicians. *Respir Med* 2007; 101 (9): 2011-6.
9. **Collard HR, Moore BB, Flaherty KR, Brown KK, Kaner RJ, King TE, Jr., Lasky JA, Loyd JE, Noth I, Olman MA, Raghu G, Roman J, Ryu JH, Zisman DA, Hunninghake GW, Colby TV, Egan JJ, Hansell DM, Johkoh T, Kaminski N, Kim DS, Kondoh Y, Lynch DA, Muller-Quernheim J, Myers JL, Nicholson AG, Selman M, Toews GB, Wells AU, Martinez FJ, Idiopathic Pulmonary Fibrosis Clinical Research Network Investigators.** Acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2007; 176 (7): 636-43.
10. **Crawford B, Monz B, Hohlfeld J, Roche N, Rubin B, Magnussen H, Nivens C, Ghafouri M, McDonald J, Tetzlaff K.** Development and validation of a cough and sputum assessment questionnaire. *Respir Med* 2008; 102 (11): 1545-55.
11. **du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, King TE, Jr., Lancaster L, Noble PW, Sahn SA, Thomeer M, Valeyre D, Wells AU.** Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and

- minimal clinically important difference. *Am J Respir Crit Care Med* 2011; 184 (12): 1382-9.
12. **du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, Lancaster L, Noble PW, Raghu G, Sahn SA, Swarcberg J, Thomeer M, Valeyre D, King TE, Jr.** Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 184 (4): 459-66.
 13. **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; 113 (3): 619-24.
 14. **Eakin EG, Sassi-Dambron DE, Ries AL, Kaplan RM.** Reliability and validity of dyspnea measures in patients with obstructive lung disease. *Int J Behav Med* 1995; 2 (2): 118-34.
 15. **European Medicines Agency (EMA).** European Public Assessment Report. 2015;
 16. **Flaherty KR, Mumford JA, Murray S, Kazerooni EA, Gross BH, Colby TV, Travis WD, Flint A, Toews GB, Lynch JP, III, Martinez FJ.** Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 2003; 168 (5): 543-8.
 17. **Gries KS, Esser D, Wiklund I.** Content validity of CASA-Q cough domains and UCSD-SOBQ for use in patients with Idiopathic Pulmonary Fibrosis. *Global journal of health science* 2013; 5 (6): 131-41.
 18. **Jegal Y, Kim DS, Shim TS, Lim CM, Do Lee S, Koh Y, Kim WS, Kim WD, Lee JS, Travis WD, Kitaichi M, Colby TV.** Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 2005; 171 (6): 639-44.
 19. **Johansson K, Collard HR.** Acute Exacerbation of Idiopathic Pulmonary Fibrosis: A Proposal. *Current respiratory care reports* 2013; 2 (4): 233-40.
 20. **Jones PW.** Interpreting thresholds for a clinically significant change in health status in asthma and COPD. *Eur Respir J* 2002; 19 (3): 398-404.
 21. **Jones PW.** St. George's Respiratory Questionnaire: MCID. *COPD* 2005; 2 (1): 75-9.
 22. **Jones PW.** Estimation and application of the minimum clinically important difference in COPD. *Lancet Respir Med* 2014; 2 167-9.
 23. **Jones PW, Donohue JF, Nedelman J, Pascoe S, Pinault G, Lassen C.** Correlating changes in lung function with patient outcomes in chronic obstructive pulmonary disease: a pooled analysis. *Respiratory research* 2011; Dec 29 (12): 161.
 24. **Jones PW, Quirk FH, Baveystock CM.** The St. George's Respiratory Questionnaire. *Respir Med* 1991; 85 (Suppl B): 25-31.

25. **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. The American review of respiratory disease 1992; 145 (6): 1321-7.
26. **Karpinski NA, Bauernschmitt K, Schultz K, Petermann F.** Der Saint George's Respiratory Questionnaire (SGRQ) im Rahmen der Qualitätssicherung: Ist die deutschsprachige Version des SGRQ bei Patienten mit Asthma- und COPD ein valides Meßinstrument der krankheitsspezifischen Lebensqualität? Prävention und Rehabilitation 2005; 17 (4): 164-76.
27. **Katzenstein AL, Myers JL.** Idiopathic pulmonary fibrosis: clinical relevance of pathologic classification. Am J Respir Crit Care Med 1998; 157 (4 Pt 1): 1301-15.
28. **King TE, Jr., Behr J, Brown KK, du Bois RM, Lancaster L, de Andrade JA, Stahler G, Leconte I, Roux S, Raghu G.** BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2008; 177 (1): 75-81.
29. **King TE, Pardo A, Selman M.** Idiopathic pulmonary fibrosis. Lancet 2011; 378 (9807): 1949-61.
30. **Kupferberg DH, Kaplan RM, Slymen DJ, Ries AL.** Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. J Cardpulm Rehabil 2005; 25 (6): 370-7.
31. **Latsi PI, du Bois RM, Nicholson AG, Colby TV, Bisirtzoglou D, Nikolakopoulou A, Veeraraghavan S, Hansell DM, Wells AU.** Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. Am J Respir Crit Care Med 2003; 168 (5): 531-7.
32. **Lee SH, Shim HS, Cho SH, Kim SY, Lee SK, Son JY, Jung JY, Kim EY, Lim JE, Lee KJ, Park BH, Kang YA, Kim YS, Kim SK, Chang J, Park MS.** Prognostic factors for idiopathic pulmonary fibrosis: clinical, physiologic, pathologic, and molecular aspects. Sarcoidosis Vasc Diffuse Lung Dis 2011; 28 (2): 102-12.
33. **Ley B, Collard HR.** Epidemiology of idiopathic pulmonary fibrosis. Clinical epidemiology 2013; Nov 25 (5): 483-92.
34. **Neurohr C, Behr J.** Diagnostik und Therapie interstitieller Lungenerkrankungen [Diagnosis and therapy of interstitial lung diseases]. Dtsch Med Wochenschr 2009; 134 (11): 524-9.
35. **Novelli F, Tavanti L, Cini S, Aquilini F, Melosini L, Romei C, Sbragia P, Falaschi F, Celi A, Paggiaro P.** Determinants of the prognosis of idiopathic pulmonary fibrosis. Eur Rev Med Pharmacol Sci 2014; 18 (6): 880-6.
36. **Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE, Jr., Kondoh Y, Myers J, Muller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schunemann HJ, ATS/ERS/JRS/ALAT**

- Committee on Idiopathic Pulmonary Fibrosis.** An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183 (6): 788-824.
37. **Richeldi L, Ryerson CJ, Lee JS, Wolters PJ, Koth LL, Ley B, Elicker BM, Jones KD, King TE, Jr., Ryu JH, Collard HR.** Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 2012; 67 (5): 407-11.
 38. **Song JW, Hong SB, Lim CM, Koh Y, Kim DS.** Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *The European respiratory journal* 2011; 37 (2): 356-63.
 39. **Swigris JJ, Brown KK, Behr J, du Bois RM, King TE, Raghu G, Wamboldt FS.** The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med* 2010; 104 (2): 296-304.
 40. **Swigris JJ, Han M, Vij R, Noth I, Eisenstein EL, Anstrom KJ, Brown KK, Fairclough D.** The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. *Respir Med* 2012; 106 (10): 1447-55.
 41. **Yorke J, Jones PW, Swigris JJ.** Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax* 2010; 65 (10): 921-6.
 42. **Zappala CJ, Latsi PI, Nicholson AG, Colby TV, Cramer D, Renzoni EA, Hansell DM, du Bois RM, Wells AU.** Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *The European respiratory journal* 2010; 35 (4): 830-6.