

Dokumentvorlage, Version vom 18.04.2013

Dossier zur Nutzenbewertung gemäß § 35a SGB V

Nintedanib (Ofev®)

Boehringer Ingelheim Pharma GmbH & Co. KG

Modul 4 A

Idiopathische Lungenfibrose

Medizinischer Nutzen und
medizinischer Zusatznutzen,
Patientengruppen mit therapeutisch
bedeutsamem Zusatznutzen

Stand: 12.03.2015

Inhaltsverzeichnis

	Seite
Inhaltsverzeichnis	1
Tabellenverzeichnis	4
Abbildungsverzeichnis	8
Abkürzungsverzeichnis	9
4 Modul 4 – allgemeine Informationen	12
4.1 Zusammenfassung der Inhalte von Modul 4.....	13
4.2 Methodik.....	28
4.2.1 Fragestellung.....	28
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung.....	29
4.2.3 Informationsbeschaffung.....	29
4.2.3.1 Studien des pharmazeutischen Unternehmers.....	30
4.2.3.2 Bibliografische Literaturrecherche.....	30
4.2.3.3 Suche in Studienregistern.....	31
4.2.3.4 Selektion relevanter Studien.....	32
4.2.4 Bewertung der Aussagekraft der Nachweise.....	32
4.2.5 Informationssynthese und -analyse.....	35
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien.....	35
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien.....	36
4.2.5.3 Meta-Analysen.....	50
4.2.5.4 Sensitivitätsanalysen.....	52
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren.....	52
4.2.5.6 Indirekte Vergleiche.....	55
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen.....	57
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	60
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel.....	60
4.3.1.1.1 Studien des pharmazeutischen Unternehmers.....	60
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche.....	62
4.3.1.1.3 Studien aus der Suche in Studienregistern.....	64
4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	65
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	67
4.3.1.2.1 Studiendesign und Studienpopulationen.....	67
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	76
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien.....	77
4.3.1.3.1 Mortalität – RCT.....	79
4.3.1.3.1.1 Endpunkte <i>Zeit bis zum Tod</i>	80
4.3.1.3.1.2 Endpunkte <i>Forcierte Vitalkapazität</i>	87
4.3.1.3.2 Morbidität – RCT.....	94
4.3.1.3.2.1 Endpunkte <i>akute Exazerbationen</i>	95
4.3.1.3.3 Gesundheitsbezogene Lebensqualität – RCT.....	102

4.3.1.3.3.1	Endpunkte zum <i>SGRQ</i> , <i>SGRQ-I</i> , <i>SOBQ</i> , <i>CASA-Q</i> , <i>PGI-C</i> und <i>EQ-5D</i>	103
4.3.1.3.4	Unerwünschte Ereignisse– RCT	118
4.3.1.3.4.1	Endpunkt <i>Gesamtzahl UE</i>	119
4.3.1.3.4.2	Endpunkt <i>schwerwiegende UE</i>	123
4.3.1.3.4.3	Endpunkt <i>UE, die zum Therapieabbruch führten</i>	127
4.3.1.3.4.4	Endpunkt <i>UE, die zum Tod führten</i>	131
4.3.1.3.4.5	Endpunkt <i>UE von besonderem Interesse</i>	135
4.3.1.3.5	Subgruppenanalysen – RCT.....	153
4.3.1.3.5.1	Mortalität	157
4.3.1.3.5.1.1	Endpunkt <i>Forcierte Vitalkapazität</i>	158
4.3.1.3.5.2	Morbidität	161
4.3.1.3.5.2.1	Endpunkt <i>Zeit bis zur ersten akuten Exazerbation</i>	162
4.3.1.3.5.3	Gesundheitsbezogene Lebensqualität.....	164
4.3.1.3.5.3.1	Endpunkte zum <i>SGRQ</i> , <i>SGRQ-I</i> , <i>CASA-Q</i> und <i>PGI-C</i>	165
4.3.1.3.5.4	Unerwünschte Ereignisse	176
4.3.1.3.5.4.1	Endpunkt <i>Gesamtzahl UE</i>	177
4.3.1.3.5.4.2	Endpunkt <i>UE, die zum Therapieabbruch führen</i>	178
4.3.1.3.5.4.3	Endpunkt <i>UE, die zum Tod führten</i>	181
4.3.1.3.5.4.4	Endpunkt <i>UE von besonderem Interesse</i>	183
4.3.1.3.5.5	Ländereffekte.....	189
4.3.1.3.6	Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien.....	189
4.3.2	Weitere Unterlagen.....	197
4.3.2.1	Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien	197
4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche	197
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	197
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen	197
4.3.2.1.3.1	<Endpunkt xxx> – indirekte Vergleiche aus RCT	198
4.3.2.1.3.2	Subgruppenanalysen – indirekte Vergleiche aus RCT	200
4.3.2.2	Nicht randomisierte vergleichende Studien.....	201
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien	201
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	201
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien	202
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien....	202
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien	203
4.3.2.3	Weitere Untersuchungen.....	204
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen	204
4.3.2.3.2	Charakteristika der weiteren Untersuchungen	204
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen	204
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen	204
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen	205
4.3.2.4	Zusammenfassung der Ergebnisse aus weiteren Unterlagen.....	205
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	206
4.4.1	Beurteilung der Aussagekraft der Nachweise	207

4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	210
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht	220
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte	220
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	220
4.5.2	Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen.....	220
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen	220
4.5.4	Verwendung von Surrogatendpunkten.....	221
4.6	Liste der eingeschlossenen Studien.....	227
4.7	Referenzliste.....	227
Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche		234
Anhang 4-B : Suchstrategien – Suche in Studienregistern.....		236
Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....		237
Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern).....		238
Anhang 4-E : Methodik der eingeschlossenen Studien – RCT		239
Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten		261

Tabellenverzeichnis

	Seite
Tabelle 4-1: Ergebnisse zum Ausmaß des medizinischen Zusatznutzen.....	25
Tabelle 4-2: Übersicht der a priori geplanten Subgruppen, zu denen Interaktionstests durchgeführt wurden.	54
Tabelle 4-3: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel	60
Tabelle 4-4: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	62
Tabelle 4-5: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel	64
Tabelle 4-6: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	66
Tabelle 4-7: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel	68
Tabelle 4-8: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	71
Tabelle 4-9: Charakterisierung der Studienpopulation (Baseline) – RCT mit dem zu bewertenden Arzneimittel	72
Tabelle 4-10: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	76
Tabelle 4-11: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel	77
Tabelle 4-12: Operationalisierung von <i>Zeit bis zum Tod</i>	80
Tabelle 4-13: Bewertung des Verzerrungspotenzials für den Endpunkt <i>Zeit bis zum Tod</i> in RCT mit dem zu bewertenden Arzneimittel	81
Tabelle 4-14: Ergebnisse für <i>Zeit bis zum Tod</i> aus RCT mit dem zu bewertenden Arzneimittel.....	83
Tabelle 4-15: Operationalisierung von <i>FVC</i>	87
Tabelle 4-16: Bewertung des Verzerrungspotenzials für den Endpunkt <i>FVC</i> in RCT mit dem zu bewertenden Arzneimittel.....	88
Tabelle 4-17: Ergebnisse für <i>jährliche FVC-Abnahme</i> (ml) aus RCT mit dem zu bewertenden Arzneimittel	90
Tabelle 4-18: Ergebnisse für <i>FVC-Responder</i> Schwellenwert 10 % (5 %) vom Sollwert aus RCT mit dem zu bewertenden Arzneimittel	92
Tabelle 4-19: Operationalisierung von <i>akute Exazerbationen</i>	95
Tabelle 4-20: Bewertung des Verzerrungspotenzials für den Endpunkt <i>akute Exazerbationen</i> in RCT mit dem zu bewertenden Arzneimittel	96
Tabelle 4-21: Ergebnisse für <i>Zeit bis zur ersten adjudizierten akuten Exazerbationen</i> und <i>Zeit bis zur ersten akuten Exazerbation</i> aus RCT mit dem zu bewertenden Arzneimittel	98

Tabelle 4-22: Operationalisierung zu <i>SGRQ</i> , <i>SGRQ-I</i> , <i>SOBQ</i> , <i>CASA-Q</i> , <i>PGI- C</i> und <i>EQ-5D</i>	103
Tabelle 4-23: Bewertung des Verzerrungspotenzials für die Endpunkte <i>SGRQ</i> , <i>SGRQ-I</i> , <i>SOBQ</i> , <i>CASA-Q</i> , <i>PGI- C</i> und <i>EQ-5D</i> in RCT mit dem zu bewertenden Arzneimittel	105
Tabelle 4-24: Ergebnisse für <i>Veränderung SGRQ-Gesamtscore</i> und <i>Veränderung SGRQ-I-Gesamtscore</i> aus RCT mit dem zu bewertenden Arzneimittel.....	107
Tabelle 4-25: Ergebnisse für <i>SGRQ-(≤ -4 Punkte)-Responder</i> und <i>SGRQ-I-(≤ -4 Punkte)-Responder</i> aus RCT mit dem zu bewertenden Arzneimittel	109
Tabelle 4-26: Ergebnisse für <i>Veränderung SOBQ-Gesamtscore</i> aus RCT mit dem zu bewertenden Arzneimittel	111
Tabelle 4-27: Ergebnisse für <i>Veränderung CASA-Q (Husten-Symptome bzw. Husten-Impact)</i> aus RCT mit dem zu bewertenden Arzneimittel.....	113
Tabelle 4-28: Ergebnisse für <i>PGI-C-Responder</i> aus RCT mit dem zu bewertenden Arzneimittel.....	115
Tabelle 4-29: Ergebnisse für <i>EQ-5D-VAS</i> aus RCT mit dem zu bewertenden Arzneimittel.	116
Tabelle 4-30: Operationalisierung von <i>Gesamtzahl UE</i>	119
Tabelle 4-31: Bewertung des Verzerrungspotenzials für den Endpunkt <i>Gesamtzahl UE</i> in RCT mit dem zu bewertenden Arzneimittel	119
Tabelle 4-32: Ergebnisse für die <i>Gesamtzahl UE</i> aus RCT mit dem zu bewertenden Arzneimittel.....	121
Tabelle 4-33: Operationalisierung von <i>SUE</i>	123
Tabelle 4-34: Bewertung des Verzerrungspotenzials für die Endpunkte zu den <i>SUE</i> in RCT mit dem zu bewertenden Arzneimittel	123
Tabelle 4-35: Ergebnisse für die <i>SUE</i> aus RCT mit dem zu bewertenden Arzneimittel	125
Tabelle 4-36: Operationalisierung von <i>UE, die zum Therapieabbruch führten</i>	127
Tabelle 4-37: Bewertung des Verzerrungspotenzials für den Endpunkt <i>UE, die zum Therapieabbruch führten</i> , in RCT mit dem zu bewertenden Arzneimittel	127
Tabelle 4-38: Ergebnisse für <i>UE, die zum Therapieabbruch führten</i> , aus RCT mit dem zu bewertenden Arzneimittel	129
Tabelle 4-39: Operationalisierung von <i>UE, die zum Tod führten</i>	131
Tabelle 4-40: Bewertung des Verzerrungspotenzials für den Endpunkt <i>UE, die zum Tod führten</i> , in RCT mit dem zu bewertenden Arzneimittel.....	131
Tabelle 4-41: Ergebnisse für <i>UE, die zum Tod führten</i> , aus RCT mit dem zu bewertenden Arzneimittel.....	133
Tabelle 4-42: Operationalisierung von <i>UE von besonderem Interesse</i>	135
Tabelle 4-43: Bewertung des Verzerrungspotenzials für den Endpunkt <i>UE von besonderem Interesse</i> in RCT mit dem zu bewertenden Arzneimittel.....	135
Tabelle 4-44: Ergebnisse für <i>UE – gastrointestinale Ereignisse</i> aus RCT mit dem zu bewertenden Arzneimittel	137

Tabelle 4-45: Ergebnisse für <i>UE – Leber- und Gallenerkrankungen (Leberenzym-, Bilirubin-, ALKP-Erhöhung)</i> aus RCT mit dem zu bewertenden Arzneimittel.....	142
Tabelle 4-46: Ergebnisse für <i>UE - kardiale Ereignisse</i> aus RCT mit dem zu bewertenden Arzneimittel.....	145
Tabelle 4-47: Ergebnisse für <i>UE - respiratorische Ereignisse</i> aus RCT mit dem zu bewertenden Arzneimittel	148
Tabelle 4-48: Ergebnisse für <i>UE - Blutungsereignisse</i> aus RCT mit dem zu bewertenden Arzneimittel.....	151
Tabelle 4-49: Subgruppen, zu denen Interaktionstests bezüglich Wirksamkeit durchgeführt wurden	154
Tabelle 4-50: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Mortalität.....	157
Tabelle 4-51: Subgruppenergebnisse für die <i>jährliche Abnahme der FVC</i> für die gepoolte Auswertung der INPULSIS Studien	158
Tabelle 4-52: Subgruppenergebnisse für <i>FVC-Responder</i> (Schwellenwert 10 % und 5 % vom Sollwert) für die gepoolte Auswertung der INPULSIS Studien	159
Tabelle 4-53: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Morbidität.....	161
Tabelle 4-54: Subgruppenergebnisse für die <i>Zeit bis zur ersten akuten Exazerbation</i> die gepoolte Auswertung der INPULSIS Studien.....	162
Tabelle 4-55: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Lebensqualität	164
Tabelle 4-56: Subgruppenergebnisse für die <i>Veränderung des SGRQ-Gesamtscores</i> und <i>SGRQ-I-Gesamtscores</i> für die gepoolte Auswertung der INPULSIS Studien	165
Tabelle 4-57: Subgruppenergebnisse für <i>SGRQ-(≤ -4 Punkte)-Responder</i> für die gepoolte Auswertung der INPULSIS Studien	169
Tabelle 4-58: Subgruppenergebnisse für <i>CASA-Q</i> für die gepoolte Auswertung der INPULSIS Studien	173
Tabelle 4-59: Subgruppenergebnisse für <i>PGI-C-Responder</i> für die gepoolte Auswertung der INPULSIS Studien	175
Tabelle 4-60: Subgruppen, zu denen Interaktionstests bezüglich <i>unerwünschter Ereignisse</i> durchgeführt wurden	176
Tabelle 4-61: Übersicht zu den p-Werten der für die Endpunkte zu unerwünschten Ereignissen	176
Tabelle 4-62: Subgruppenergebnisse für die <i>Gesamtzahl UE</i> für die gepoolte Auswertung der INPULSIS Studien	177
Tabelle 4-63: Subgruppenergebnisse für die <i>UE, die zum Therapieabbruch führen</i> für die gepoolte Auswertung der INPULSIS Studien.....	178
Tabelle 4-64: Subgruppenergebnisse für die <i>UE, die zum Tod führten</i> , für die gepoolte Auswertung der INPULSIS Studien	181

Tabelle 4-65: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zu <i>UE von besonderem Interesse</i>	183
Tabelle 4-66: Subgruppenergebnisse für die <i>UE von besonderem Interesse</i> für die gepoolte Auswertung der INPULSIS Studien.....	184
Tabelle 4-67: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche	198
Tabelle 4-68: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	198
Tabelle 4-69: Operationalisierung von <Endpunkt xxx>.....	199
Tabelle 4-70: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche	199
Tabelle 4-71: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	199
Tabelle 4-72: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien	202
Tabelle 4-73: Operationalisierung von <Endpunkt xxx>.....	202
Tabelle 4-74: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien	203
Tabelle 4-75: Operationalisierung von <Endpunkt xxx> - weitere Untersuchungen	205
Tabelle 4-76: Bewertung des Zusatznutzens mit Erläuterung - Mortalität	212
Tabelle 4-77: Bewertung des Zusatznutzens mit Erläuterung - Morbidität	214
Tabelle 4-78: Bewertung des Zusatznutzens mit Erläuterung - Lebensqualität.....	216
Tabelle 4-79: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	220
Tabelle 4-80: Liste der eingeschlossenen Studien	227
Tabelle 4-81 (Anhang): Studiendesign und -methodik für Studie INPULSIS-1	239
Tabelle 4-82 (Anhang): Studiendesign und -methodik für Studie INPULSIS-2	250
Tabelle 4-83 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie INPULSIS-1	262
Tabelle 4-84 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie INPULSIS-2	280

Abbildungsverzeichnis**Seite**

Abbildung 4-1: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zum Tod, unabhängig von der Ursache</i> für die gepoolte Analyse der INPULSIS Studien.....	15
Abbildung 4-2: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis respiratorisch bedingtem Tod</i> für die gepoolte Analyse der INPULSIS Studien	16
Abbildung 4-3: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zur ersten adjudizierten akuten Exazerbation</i> für die gepoolte Analyse der INPULSIS Studien.....	18
Abbildung 4-4: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zur ersten akuten Exazerbation</i> für die gepoolte Analyse der INPULSIS Studien	19
Abbildung 4-5: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel	63
Abbildung 4-6: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zum Tod, unabhängig von der Ursache</i> für die gepoolte Analyse der INPULSIS Studien.....	84
Abbildung 4-7: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis respiratorisch bedingtem Tod</i> für die gepoolte Analyse der INPULSIS Studien	85
Abbildung 4-8: Mittlere beobachtete FVC-Veränderung (standard error of the mean, SEM) gegenüber dem Ausgangswert (ml) im Zeitverlauf, Studien INPULSIS-1 und INPULSIS-2 gepoolt	91
Abbildung 4-9: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zur ersten adjudizierten akuten Exazerbation</i> für die gepoolte Analyse der INPULSIS Studien.....	100
Abbildung 4-10: Kaplan-Meier-Kurven für Endpunkt <i>Zeit bis zur ersten akuten Exazerbation</i> für die gepoolte Analyse der INPULSIS Studien	100
Abbildung 4-11: Vergleich der FVC-Abnahme in ml – Nintedanib vs. Placebo	223
Abbildung 4-12: Vier-Jahres-Überleben in Abhängigkeit vom Ausmaß der FVC-Abnahme in % vom Sollwert über sechs Monate bei Patienten mit IPF (n = 84)	226
Abbildung 4-13: Patientenfluss INPULSIS-1 Studie.....	249
Abbildung 4-14: Patientenfluss INPULSIS-2 Studie.....	260

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ALKP	Alkalische Phosphatase
ALT	Alanin-Aminotransferase
AST	Aspartat-Aminotransferase
ATS	American Thoracic Society
BMI	Body-Mass-Index
CASA-Q	Cough and Sputum Assessment Questionnaire
CONSORT	Consolidated Standards of Reporting Trials
COPD	Chronic Obstructive Pulmonary Disease (Chronisch obstruktive Lungenerkrankung)
CRF	Case Report Form
DGP	Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DL _{CO}	Diffusing Capacity of the Lung for Carbon Monoxide (Kohlenmonoxid-Diffusionskapazität)
eCRF	Electronic Case Report Form
EG	Europäische Gemeinschaft
EU	Europäische Union
EQ-5D	EuroQol 5-Dimensional Quality of Life Questionnaire
ERS	European Respiratory Society
FEV ₁	Forced Expiratory Volume in one Second (Forcierte expiratorische Einsekunden-Kapazität)
FVC	Forced Vital Capacity (Forcierte Vitalkapazität)
G-BA	Gemeinsamer Bundesausschuss
HR	Hazard Ratio
HRCT	High Resolution Computed Tomography (Hochauflösende Computer-Tomographie)
HRQoL	Health-Related Quality of Life (Gesundheitsbezogene Lebensqualität)
IAC	Independent Adjudication Committee (Unabhängiges Bewertungskomitee)

IPF	Idiopathic Pulmonary Fibrosis (Idiopathische Lungenfibrose)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
KI	Konfidenzintervall
MedDRA	Medical Dictionary for Regulatory Activities
MCID	Minimal Clinically Important Difference (Minimaler Schwellenwert für eine klinisch relevante Verbesserung)
MTC	Mixed Treatment Comparison
MMRM	Mixed Effect Model for Repeated Measures
MW	Mittelwert
MWD	Mittelwertdifferenz
NICE	National Institute for Health and Care Excellence
OC	Observed Case (Auswertung nur der tatsächlich beobachteten Werte)
OR	Odds Ratio
PGI-C	Patient´s Global Impression of Change
RCT	Randomized Controlled Trial
RD	Risk Difference (Absolute Risikodifferenz)
RR	Risk Ratio (Relative Risikodifferenz)
RS	Randomized Set
SD	Standard Deviation (Standardabweichung)
SE	Standard Error (Standardfehler)
SEM	Standard Error of the Mean
SGB	Sozialgesetzbuch
SGRQ	St. George´s Respiratory Questionnaire
SGRQ-I	St. George´s Respiratory Questionnaire IPF-specific version
SpO ₂	Sauerstoffsättigung
SOBQ	Shortness of Breath Questionnaire
STE	Surrogate Threshold Effects
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SUE	Schwerwiegendes unerwünschtes Ereignis

TREND	Transparent Reporting of Evaluations with Non-Randomized Design
TS	Treated Set
UE	Unerwünschtes Ereignis
UEA	Unerwünschtes Ereignis, das zum Therapieabbruch führt
UIP	Usual Interstitial Pneumonia
ULN	Upper Limits of Normal (Oberer Normgrenzwert)
VAS	Visual Analog Scale
VerfO	Verfahrensordnung
WHO	World Health Organization (Weltgesundheitsorganisation)

4 Modul 4 – allgemeine Informationen

Modul 4 enthält folgende Angaben:

- Zusammenfassung (Abschnitt 4.1)
- Angaben zur Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens (Abschnitt 4.2)
- Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen (Abschnitt 4.3)
- eine abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens, einschließlich der Angabe von Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Abschnitt 4.4)
- ergänzende Informationen zur Begründung der vorgelegten Unterlagen (Abschnitt 4.5)

Für jedes zu bewertende Anwendungsgebiet ist eine separate Version des vorliegenden Dokuments zu erstellen. Die Kodierung der Anwendungsgebiete ist in Modul 2 hinterlegt. Sie ist je Anwendungsgebiet einheitlich für die Module 3, 4 und 5 zu verwenden.

Für Arzneimittel, die zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 zugelassen sind, müssen gemäß §35b SGB V keine Nachweise zum medizinischen Nutzen und medizinischen Zusatznutzen (ZN) vorgelegt werden, solange der Umsatz des Arzneimittels zu Lasten der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten zwölf Monaten einen Betrag von 50 Millionen Euro nicht übersteigt. Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, müssen jedoch vorgelegt werden. In diesem Fall müssen in Modul 4 nur die Abschnitte 4.2 und 4.4 ausgefüllt werden. Weitere Abschnitte von Modul 4 können zur strukturierten Darstellung der Ergebnisse ausgefüllt werden. In dem vorliegenden Dossier wurden daher zusätzlich weitere Abschnitte von Modul 4 entsprechend den Vorgaben der Dokumentenvorlage bearbeitet.

Im Dokument verwendete Abkürzungen sind in das Abkürzungsverzeichnis aufzunehmen. Sofern Sie für Ihre Ausführungen Tabellen und Abbildungen verwenden, sind diese im Tabellen- bzw. Abbildungsverzeichnis aufzuführen.

4.1 Zusammenfassung der Inhalte von Modul 4

Stellen Sie eine strukturierte Zusammenfassung der Inhalte von Modul 4 zur Verfügung.

Fragestellung

Ziel des Dossiers ist die Bewertung des Ausmaßes des medizinischen Zusatznutzens im Sinne des § 35a des Sozialgesetzbuches (SGB) V von Ofev[®] (Nintedanib) bei der Behandlung von erwachsenen Patienten mit idiopathischer Lungenfibrose (IPF) jeglicher Schweregrade. Aufgrund des Status eines Arzneimittels zur Behandlung einer seltenen Erkrankung ist der medizinische Nutzen und medizinische Zusatznutzen mit der Zulassung von Nintedanib bereits festgestellt (§ 35a SGB V Abs. 1 Satz 10 Halbs. 1 SGB V). Nachweise zum medizinischen Nutzen und zum medizinischen Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie müssen nicht vorgelegt werden (§ 35a SGB V Abs. 1 Satz 10 Halbs. 2 SGB V). Die Bewertung des Ausmaßes für den medizinischen Nutzen und Zusatznutzen erfolgt auf der Basis der Zulassung und der zugrunde liegenden Placebo-kontrollierten Studien (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c, Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d) anhand der patientenrelevanten Nutzendimensionen *Mortalität, Morbidität, Lebensqualität* sowie *unerwünschte Ereignisse*.

Datenquellen

Das Ausmaß des Zusatznutzens für Nintedanib (Ofev[®]) wird auf der Grundlage der Zulassung und der die Zulassung begründenden Studien definiert (Gemeinsamer Bundesausschuss, 2013). Analog zum Vorgehen der Zulassungsbehörden wird im vorliegenden Dossier zur frühen Nutzenbewertung von Nintedanib in der Indikation IPF auf der Basis der beiden Phase III-Studien (INPULSIS-1 und INPULSIS-2) das Ausmaß des Zusatznutzens abgeleitet.

Ein-/Ausschlusskriterien für Studien

Entsprechend den Vorgaben für die Nutzenbewertung von Arzneimitteln zur Behandlung seltener Erkrankungen (§ 35a SGB V) erfolgt die Bewertung des Ausmaßes des Zusatznutzens auf Grundlage der zulassungsrelevanten Studien. Die Zulassung für Nintedanib (Ofev[®]) in der Indikation IPF beruht auf zwei RCTs (INPULSIS-1 und INPULSIS-2). Dementsprechend erfolgt die Darstellung des Ausmaßes für den Zusatznutzen auf Basis dieser beiden Studien.

Methoden zur Bewertung der Aussagekraft der Nachweise und zur Synthese von Ergebnissen

Das Design und die Methodik der eingeschlossenen RCTs (INPULSIS-1 und INPULSIS-2) wurden anhand der Consolidated Standards of Reporting Trials (CONSORT)-Kriterien dargestellt. Die Studienpopulation wurde anhand demografischer und krankheitsspezifischer Faktoren beschrieben. Verzerrungsaspekte wurden endpunktübergreifend für die gesamte eingeschlossene Studie und endpunktspezifisch bewertet. Die Bewertung folgte den Vorgaben der Dossievorlage. Die Patientenrelevanz und die Validität der Endpunkte bzw. Messinstrumente wurden überprüft und diskutiert.

Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Der medizinische Nutzen und Zusatznutzen von Nintedanib zur Behandlung des seltenen Leidens IPF wurde mit der Zulassung durch die Europäische Arzneimittelbehörde (European Medicines Agency, EMA) bestätigt. Die Grundlage für die Bewertung des Ausmaßes des medizinischen Zusatznutzens sind die Studienberichte und -protokolle der pivotalen Zulassungsstudien INPULSIS-1 und INPULSIS-2 sowie die Veröffentlichungen von Richeldi et al. 2014 (Richeldi et al., 2014a, Richeldi et al., 2014b).

Mit den INPULSIS-1 und INPULSIS-2 Studien liegen zwei doppelblinde RCTs vor, so dass die zur Bewertung des Zusatznutzens herangezogene Evidenz der höchsten Evidenzstufe entspricht (Evidenzstufe Ib).

Die Phase III-Studien INPULSIS-1 und INPULSIS-2 sind analog aufgebaut und grundsätzlich für eine gepoolte Auswertung auf Patientenebene geeignet. Auf Basis der gepoolten Analyse wird eine wesentlich höhere Ergebnissicherheit im Vergleich zu den einzelnen Studien erreicht. Deshalb wird in der Ergebnisdarstellung und bei der Bewertung des Ausmaßes des Zusatznutzens auf Ergebnisse der gepoolten Auswertung fokussiert. Sofern nicht anders dargestellt, beziehen sich alle folgenden Angaben auf die gepoolten Ergebnisse der INPULSIS Studien.

Nachstehend wird das Ausmaß des medizinischen Zusatznutzens von Nintedanib bei erwachsenen IPF-Patienten beschrieben.

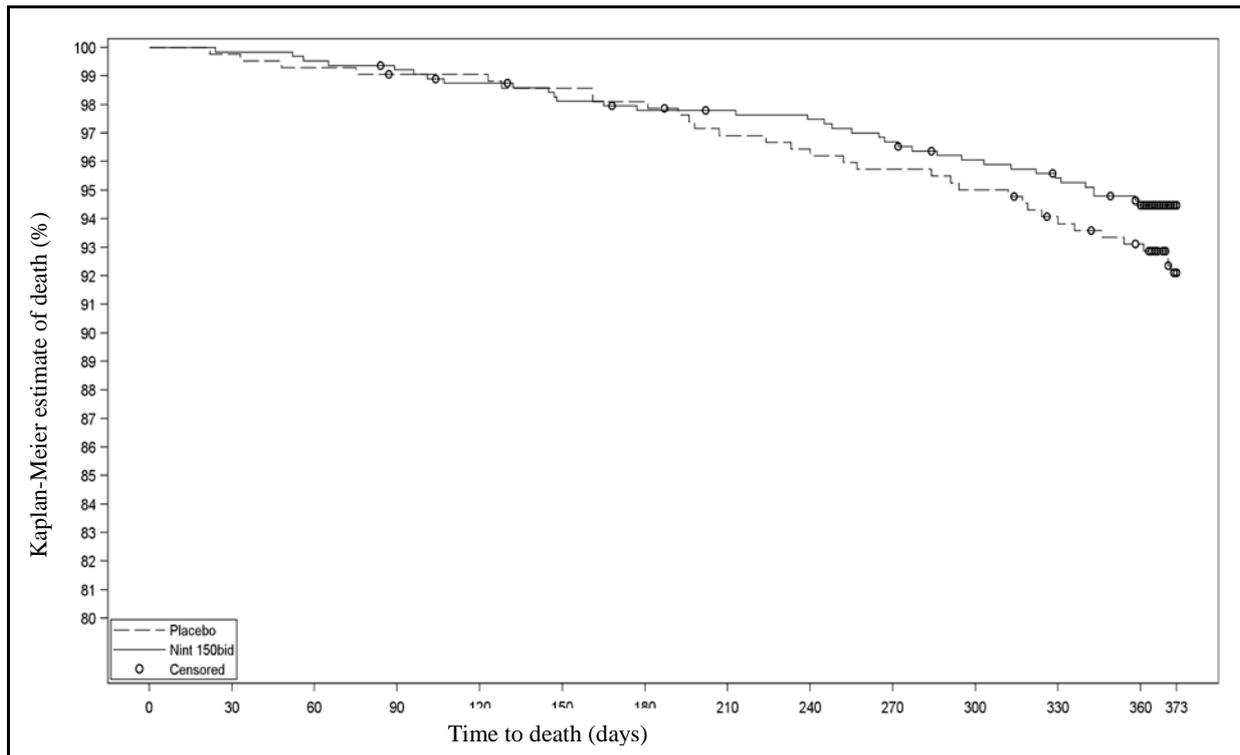
Mortalität

Die Mortalität wurde in diesem Dossier durch die Endpunkte *Zeit bis zum Tod (unabhängig von der Ursache bzw. respiratorisch bedingt)* sowie durch die *jährliche Abnahme der forcierten Vitalkapazität (FVC)* in ml und den Anteil *FVC-Responder* (Schwellenwerte 10 % und 5 % vom Sollwert) operationalisiert und hinsichtlich des medizinischen Zusatznutzens bewertet.

Die Überlebenswahrscheinlichkeit war unter Nintedanib höher als unter Placebo. Sowohl die *jährliche FVC Abnahme* (in ml) als auch der Anteil *FVC-Responder* war unter Nintedanib höher als unter Placebo. Der Unterschied ist jeweils statistisch signifikant. Als Surrogat für die Mortalität deutet dies ebenfalls auf eine Reduktion der Sterbewahrscheinlichkeit bei IPF-Patienten unter der Therapie mit Nintedanib im Vergleich zu Placebo-behandelten Patienten hin.

Zeit bis zum Tod

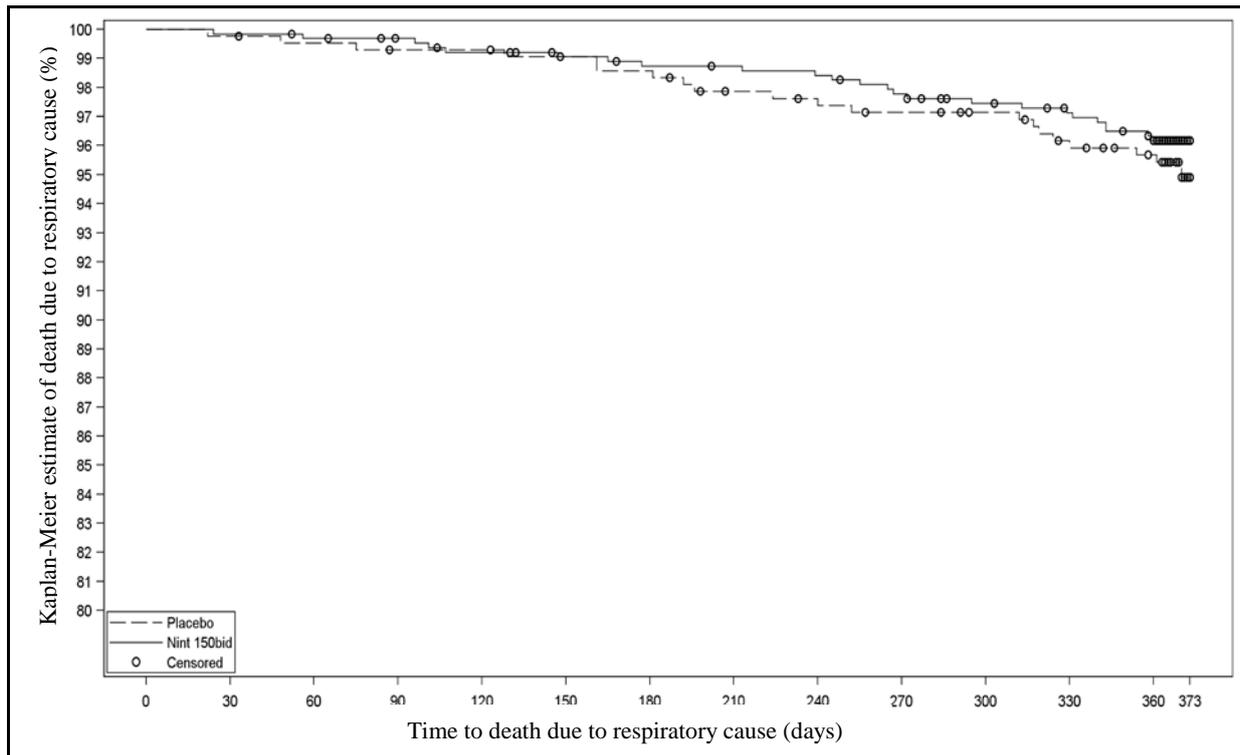
Die *Zeit bis Tod, unabhängig von der Ursache* wird als Überlebenswahrscheinlichkeit anhand des Kaplan-Meier-Schätzers berechnet. Basierend auf einer Cox-Regression, war die Überlebenswahrscheinlichkeit zum Studienende in den gepoolten INPULSIS Studien unter Nintedanib höher als unter Placebo (Hazard Ratio (HR) [95 % Konfidenzintervall (KI)]: 0,70 [0,43; 1,12]; $p = 0,1399$) (Abbildung 4-1).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-1: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zum Tod*, unabhängig von der *Ursache* für die gepoolte Analyse der INPULSIS Studien

Ein vergleichbarer Vorteil wurde auch für die *Zeit bis zum respiratorisch bedingten Tod* beobachtet (HR [95 % KI]: 0,74 [0,41; 1,34]; $p = 0,3435$). Die Kaplan-Meier-Kurven zeigen auch hier, dass mit zunehmender Beobachtungszeit der Unterschied der aus respiratorischen Gründen verstorbenen Studienteilnehmer zwischen den Behandlungsgruppen Nintedanib und Placebo größer wird (Abbildung 4-2).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-2: Kaplan-Meier-Kurven für Endpunkt *Zeit bis respiratorisch bedingtem Tod* für die gepoolte Analyse der INPULSIS Studien

Forcierte Vitalkapazität

Die *jährliche FVC-Abnahme* (ml) war in den gepoolten INPULSIS Studien unter Nintedanib mit 113,59 ml klinisch relevant und statistisch signifikant geringer als unter Placebo (223,53 ml). Der Unterschied zwischen den beiden Behandlungsgruppen hat eine Mittelwertdifferenz (MWD) von 109,94 ml (MWD [95 % KI]: 109,94 ml [75,85; 144,03]; $p < 0,0001$). Die *Abnahme der FVC* im Behandlungsarm mit Nintedanib fiel somit durchschnittlich um 50 % geringer aus als im Placeboarm.

Der Anteil der *FVC-Responder* war in den Phase III-Studien definiert als Anteil der Studienteilnehmer mit einem absoluten Abfall der FVC nicht größer als 10 % bzw. nicht größer als 5 % von Baseline bis Woche 52. Ein Responder ist demnach ein Studienteilnehmer, dessen FVC-Wert nicht weiter als 10 % bzw. 5 % des Sollwerts über den Studienverlauf abfällt. Der Cut Off-Wert von 10 % ist dabei in Fachkreisen der anerkannte Wert (Behr, 2013). Daher wurde in diesem Dossier primär das Ansprechen mit einem Schwellenwert von 10 % und supportiv mit einem Schwellenwert von 5 % ausgewertet (Abschnitt 4.5.4).

Die Analyse der *FVC-Responder* zeigt durch eine logistische Regression mit einem OR von 1,577 einen statistisch signifikanten Unterschied zwischen den Behandlungsgruppen (OR [95 % KI]: 1,577 [1,21; 2,05]; $p = 0,0007$).

Bei einem Schwellenwert von 5 % zeigt sich ebenfalls ein statistisch signifikanter Unterschied: auf Nintedanib sprachen gemäß dieser Analyse mehr Studienteilnehmer an als auf Placebo (OR [95 % KI]: 1,835 [1,43; 2,36]; $p < 0,0001$).

Morbidität

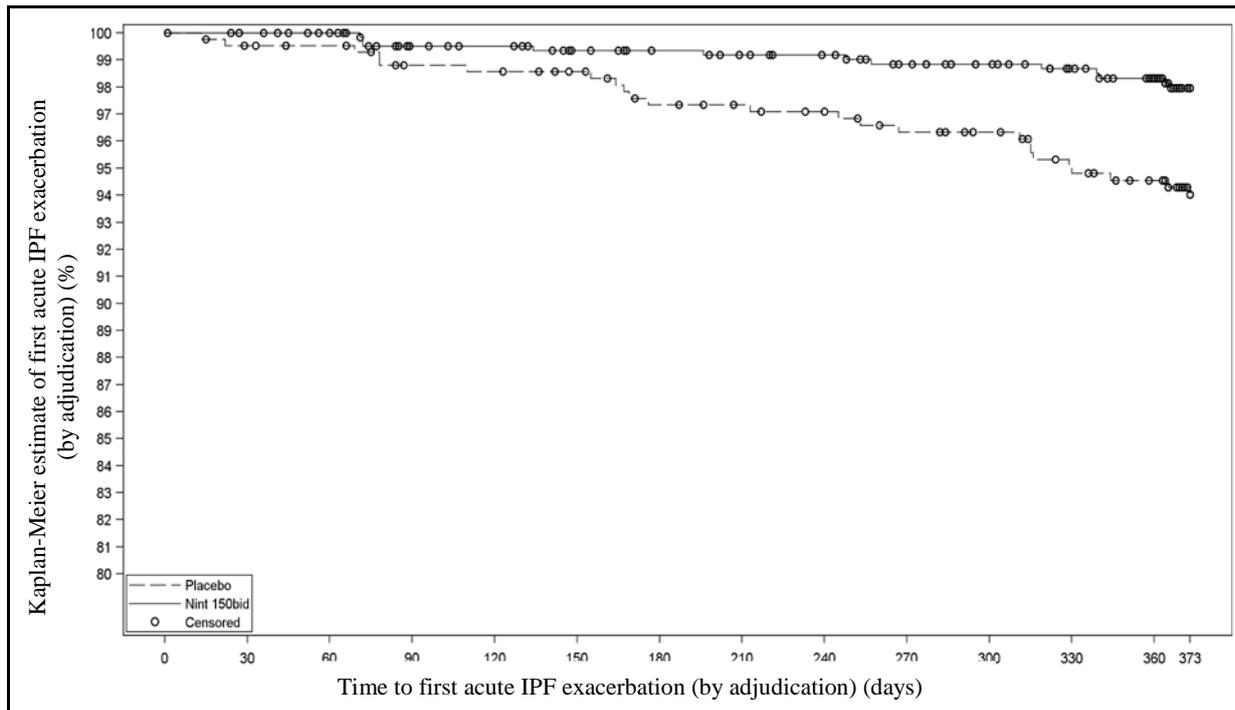
Die Morbidität wurde in diesem Dossier durch die Endpunkte *Zeit bis zur ersten adjudizierten akuten Exazerbation* (bewertet durch ein unabhängiges Bewertungskomitee, IAC), *Zeit bis zur ersten akuten Exazerbation* (bewertet durch den Prüfarzt) operationalisiert und hinsichtlich des Ausmaßes des medizinischen Zusatznutzens bewertet.

Die Ergebnisse aller Endpunkte zur Morbidität zeigen einen Vorteil der Nintedanib-Therapie im Vergleich zu Placebo. Für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* war der Unterschied statistisch signifikant.

Exazerbation

Adjudizierte akute Exazerbationen traten in den gepoolten INPULSIS Studien unter Nintedanib seltener auf als unter Placebo (1,9 % vs. 5,7 %). Die Wahrscheinlichkeit, keine Exazerbation zu erleiden (ausgedrückt durch den Kaplan-Meier-Schätzer), beträgt in der Nintedanib-Gruppe 0,980 und in der Placebo-Gruppe 0,940. Auch das HR [95 % KI] von 0,32 für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* zeigt einen Vorteil zugunsten von Nintedanib (HR [95 % KI]: 0,32 [0,16; 0,65]; $p = 0,0010$). Der Unterschied ist statistisch signifikant. Dies entspricht einer statistisch signifikanten Reduktion des Risikos einer ersten adjudizierten akuten Exazerbation um 68 %.

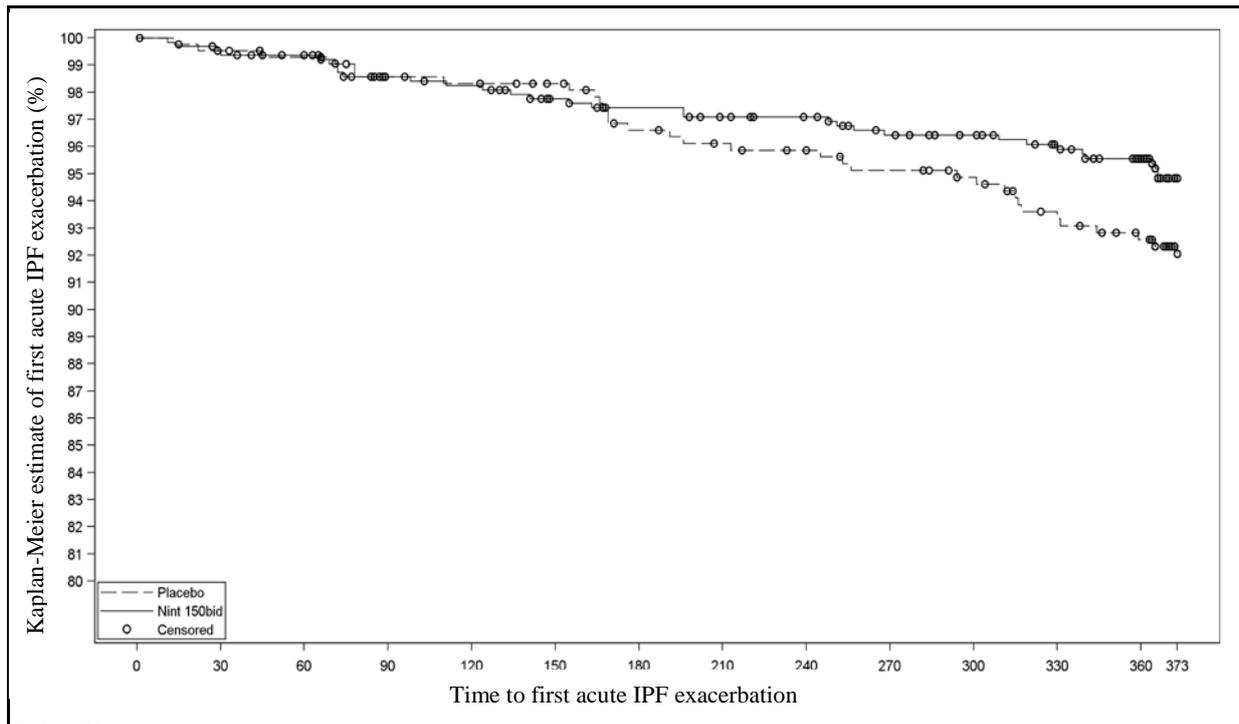
Die Kaplan-Meier-Kurven zeigen deutlich, dass mit zunehmender Beobachtungszeit die Kurven der *Zeit bis zur ersten adjudizierten akuten Exazerbation* zwischen den Behandlungsgruppen Nintedanib und Placebo kontinuierlich auseinandergehen (Abbildung 4-3).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-3: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zur ersten adjudizierten akuten Exazerbation* für die gepoolte Analyse der INPULSIS Studien

Ein Vorteil zugunsten von Nintedanib wurde auch für die *Zeit bis zur ersten akuten Exazerbation* in den gepoolten Daten der INPULSIS Studien (HR [95 % K]): 0,64 [0,39; 1,05]; $p = 0,0823$) beobachtet, jedoch ohne statistisch signifikanten Effekt. (Abbildung 4-4).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-4: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zur ersten akuten Exazerbation* für die gepoolte Analyse der INPULSIS Studien

Gesundheitsbezogene Lebensqualität

Die krankheitsspezifische Lebensqualität wurde in diesem Dossier durch die Endpunkte *Veränderung des St. George's Respiratory Questionnaire (SGRQ)*, *Veränderung des IPF-spezifischen SGRQ (SGRQ-I)* und *-Responder-Analysen* sowie *Veränderung des Shortness of Breath Questionnaire (SOBQ)* und der *Veränderung des Cough and Sputum Assessment Questionnaire (CASA-Q)* bewertet. Die allgemeine Lebensqualität wurde durch die Endpunkte *Responder-Analyse des Patient's Global Impression Change (PGI-C)* und *Veränderung der EuroQol 5-Dimensional Quality of Life Questionnaire Analogskala (EQ-5D-VAS)* dargestellt.

Die Ergebnisse aller Endpunkte zeigen einen Vorteil in Form einer geringeren Verschlechterung der Lebensqualität innerhalb des Studienzeitraums unter Nintedanib im Vergleich zu Placebo. Für die *Veränderung des SGRQ-I* und die *PGI-C-Responder* sind die Unterschiede statistisch signifikant.

Endpunkte zum SGRQ und SGRQ-I

Die Lebensqualität, ausgedrückt durch die Differenz des Wertes zu Woche 52 minus des Wertes zur Baseline, führt zu einer Zunahme des *SGRQ-Gesamtscores* in beiden Studienarmen. In der gepoolten Analyse der INPULSIS Studien fällt die Verschlechterung der Lebensqualität unter Nintedanib geringer als bei den Placebo-behandelten Studienteilnehmern aus (3,53 Punkte vs. 4,96 Punkte). Die *Veränderung des SGRQ-Gesamtscores* im Studienverlauf zeigt somit einen Vorteil von Nintedanib gegenüber Placebo (MWD [95 % KI]: -1,43 Punkte [-3,09; 0,23]; p = 0,0923).

Die dichotomisierte Analyse der *SGRQ-(≤ -4 Punkte)-Responder* (Änderung des SGRQ um ≤ -4 Punkte zu Woche 52) für die gepoolte Auswertung der INPULSIS Studien zeigt, dass unter Nintedanib in etwa gleich viele Studienteilnehmer den Schwellenwert von ≤ -4 Punkten im *SGRQ-Gesamtscore* erreichten wie unter Placebo (22,88 % vs. 20,33 % Responder; OR [95 % KI]: 1,186 [0,88; 1,61]; p = 0,2715).

Die Verschlechterung der Lebensqualität von Baseline bis Woche 52, ausgedrückt durch eine Zunahme des *SGRQ-I-Gesamtscore*, fällt in der gepoolten Auswertung der INPULSIS Studien unter Nintedanib geringer aus als bei den Placebo-behandelten Studienteilnehmern (3,48 Punkte vs. 5,47 Punkte). Dies entspricht einem Vorteil für Studienteilnehmer unter der Behandlung mit Nintedanib im Vergleich zu Studienteilnehmern aus der Placebo-Gruppe. Dieser Unterschied zwischen den Behandlungsgruppen ist statistisch signifikant (MWD [95 % KI]: -1,99 Punkte [-3,71; -0,27]; p = 0,0230).

Die Analyse der *SGRQ-I-(≤ -4 Punkte)-Responder* für die gepoolte Auswertung der INPULSIS Studien zeigt, dass unter Nintedanib mehr Studienteilnehmer den Schwellenwert von ≤ -4 Punkten im *SGRQ-I-Gesamtscore* erreichten und somit auf Nintedanib ansprachen als unter Placebo (23,51 % vs. 19,62 % Responder; OR [95 % KI]: 1,284 [0,95; 1,74]; p = 0,1091).

Weitere Endpunkte zur krankheitsspezifischen Lebensqualität

Die *Veränderung des SOBQ-Score*, der die subjektiv bewertete Dyspnoe der Studienteilnehmer misst (Wertebereich: 0-120; wobei ein hoher Wert eine hohe Beeinträchtigung widerspiegelt, Minimal Clinically Important Difference (MCID): 5 Punkte), zeigt eine geringere Verschlechterung unter Nintedanib im Vergleich zu Placebo in der gepoolten Auswertung der INPULSIS Studien (6,71 Punkte vs. 8,40 Punkte; MWD [95 % KI]: -1,69 Punkte [-4,09; 0,70]; p = 0,1648).

Auch die Ergebnisse zu *CASA-Q Husten-Symptome* und zum *Husten-Impact* (Wertebereich: 0-100; wobei ein hoher Wert eine bessere Lebensqualität bzw. weniger Beeinträchtigungen widerspiegelt) in der gepoolten Auswertung der INPULSIS Studien zeigen eine im Vergleich zu Placebo bessere subjektiv empfundene krankheitsspezifische Lebensqualität unter der Behandlung mit Nintedanib (*Symptome*: MWD [95 % KI]: 0,95 Punkte [-1,49; 3,38]; $p = 0,4450$, *Impact*: MWD [95 % KI]: 1,73 Punkte [-0,46; 3,92]; $p = 0,1212$). Mit Nintedanib behandelte Studienteilnehmer berichteten über weniger Husten und hustenbedingte Einschränkungen als Studienteilnehmer im Placeboarm. Dieser Unterschied bestätigt in der Tendenz die Ergebnisse der anderen Auswertungen zur krankheitsspezifischen Lebensqualität.

Endpunkte zur allgemeinen Lebensqualität

Der Anteil *PGI-C-Responder*, definiert als Studienteilnehmer die ihre gesundheitsbezogene Lebensqualität in Woche 52 im Vergleich zur Baseline als „sehr viel besser“, „viel besser“, „ein wenig besser“ oder „ohne Veränderung“ einschätzen, ist in der gepoolten Auswertung der INPULSIS Studien unter Nintedanib höher als unter Placebo (61,29 % vs. 54,37 % Responder). Der Unterschied ist statistisch signifikant (OR [95 % KI]: 1,328 [1,04; 1,70]; $p = 0,0254$).

Auch die *Veränderung der EQ-5D-VAS* (Wertebereich: 0-100) von Baseline bis Woche 52 zeigt in der gepoolten Auswertung der INPULSIS Studien eine geringere Verschlechterung der subjektiv empfundenen Lebensqualität unter Nintedanib im Vergleich zu Placebo (1,69 Punkte vs. 5,73 Punkte). Eine Analyse zur vergleichenden Statistik wurde gemäß Studienprotokoll nicht durchgeführt.

Unerwünschte Ereignisse

Im Rahmen dieser Nutzenbewertung wurden *unerwünschte Ereignisse (UE)*, *schwerwiegende unerwünschte Ereignisse (SUE)*, *unerwünschte Ereignisse, die zum Therapieabbruch führten (UEA)*, sowie *unerwünschte Ereignisse, die zum Tod führten*, berücksichtigt. Zusätzlich wurden *unerwünschte Ereignisse von besonderem Interesse* sowie unerwünschte Ereignisse betrachtet, die in > 10 % der Studienteilnehmer eines Behandlungsarms auftraten. Die Darstellung der UE erfolgt auf Grundlage der gepoolten INPULSIS Studien. In Modul 5 werden die Daten zu den UE auf Einzelstudienebene dargestellt (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f).

- Alle UE: OR [95 % KI]: 2,44 [1,50; 3,96]; $p < 0,001$
- Alle SUE: OR [95 % KI]: 1,02 [0,78; 1,33]; $p = 0,894$
- UE, die zum Therapieabbruch führten: OR [95 % KI]: 1,60 [1,13; 2,26]; 0,007
- UE, die zum Tod führten: OR [95 % KI]: 0,78 [0,48; 1,28]; $p = 0,319$
- UE von besonderem Interesse:
Zu den UE von besonderem Interesse zählen *gastrointestinalen Störungen*,
Leberwerterhöhungen, *kardiale UE*, *respiratorische UE* und *Blutungsereignisse*.

Unter den *gastrointestinalen Störungen* waren die Ereignisse *Diarrhoe, Übelkeit* und *Erbrechen* sowie *Appetitlosigkeit, Gewichtsverlust* und *gastrointestinale Perforation* von besonderem Interesse.

- Diarrhoe: OR [95 % KI]: 7,33 [5,47; 9,84]; p < 0,001
- Übelkeit: OR [95 % KI]: 4,57 [2,99; 6,97]; p < 0,001
- Erbrechen: OR [95 % KI]: 4,91 [2,58; 9,38]; p < 0,001
- Appetitlosigkeit: OR [95 % KI]: 1,98 [1,22; 3,21]; p = 0,005
- Gewichtsverlust: OR [95 % KI]: 2,93 [1,64; 5,22]; p < 0,001
- Gastrointestinale Perforation: 0% Placebo versus 0,3 % Nintedanib

Unerwünschte Leberenzym- und Bilirubinveränderungen

UE, welche das Leber- und Gallenwegssystem betrafen, wurden in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien häufiger unter Nintedanib als unter Placebo dokumentiert: Die Behandlung mit Nintedanib war im Vergleich zu Placebo häufiger verbunden mit einer Erhöhung der Leberenzyme (Alanin-Aminotransferase (ALT), Aspartat-Aminotransferase (AST), Alkalische Phosphatase (ALKP)) oder des Bilirubin-Wertes (upper limit of normal (ULN), oberer Normgrenzwert).

- AST- und ALT-Werte um mindestens 3 x ULN erhöht (Nintedanib vs. Placebo): 5 % vs. 0,7% (OR [95 % KI]: 7,39 [2,25; 24,3]; p < 0,001)
- AST- und ALT-Werte um mindestens 5 x ULN erhöht (Nintedanib vs. Placebo): 2,2% vs. 0,2% (OR [95 % KI]: 9,47 [1,24; 72,3]; p = 0,008)
- AST- und ALT-Werte um mindestens 8 x ULN erhöht (Nintedanib vs. Placebo): 0,8% vs. 0,2% (OR [95 % KI]: 3,33 [0,39; 28,6]; p = 0,244)
- Bilirubinwert um mindestens 1,5 x ULN erhöht (Nintedanib vs. Placebo): 2,4% vs. 0,7% (OR [95 % KI]: 3,37 [0,97; 11,7]; p = 0,043)
- Bilirubinwert um mindestens 2 x ULN erhöht (Nintedanib vs. Placebo): 0,5 % vs. 0,5 % (OR [95 % KI]: 0,99 [0,17; 5,98]; p = 0,995)
- ALKP-Wert um mindestens 1,5 x ULN erhöht (Nintedanib vs. Placebo): 5,8% vs. 0,9% (OR [95 % KI]: 6,45 [2,28; 18,2]; p < 0,001)
- ALKP-Wert um mindestens 2 x ULN erhöht (Nintedanib vs. Placebo): 2,7% vs. 0,2% (OR [95 % KI]: 11,6 [1,53; 87,1]; p = 0,003)

Unerwünschte kardiale Ereignisse

Unerwünschte kardiale Ereignisse wurden in den gepoolten INPULSIS Studien vergleichbar häufig unter Nintedanib und Placebo (10,0 % vs. 10,6 %) beobachtet (OR [95 % KI] 0,94 [0,63; 1,40]; $p = 0,750$). Bei *schwerwiegenden unerwünschten kardialen Ereignissen* wurden ebenfalls keine Unterschiede zwischen den Behandlungsgruppen beobachtet (OR [95 % KI] 0,92 [0,53; 1,59]; $p = 0,762$). *Tödliche kardiale Ereignisse* traten numerisch seltener in der Nintedanib-Gruppe (0,5 %) als in der Placebo-Gruppe (1,4 %) auf (OR [95 % KI]: 0,33 [0,08; 1,32]; $p = 0,099$). *Arterielle thromboembolische Ereignisse* wurden selten berichtet: bei 0,7 % der Patienten in der Placebogruppe und bei 2,5 % in der mit Nintedanib behandelten Gruppe (OR [95 % KI]: 3,60 [1,04; 12,44]; $p = 0,0305$). Unerwünschte Ereignisse, die auf eine *ischämische Herzerkrankung* deuten, waren zwar zwischen der Nintedanib- und der Placebogruppe ausgeglichen, aber in der Nintedanib-Gruppe erlitt ein numerisch höherer Prozentsatz der Patienten (1,6 %) einen Myokardinfarkt als dies in der Placebogruppe (0,5 %) der Fall war (OR [95 % KI]: 3,35 [0,73; 15,37]; $p = 0,099$).

Unerwünschte respiratorische Ereignisse

Als häufigste *respiratorische Ereignisse* wurden eine *Progression der IPF, Husten, Dyspnoe, Bronchitis und Nasopharyngitis* im Rahmen der Erfassung von UE angegeben.

- Eine *Progression der IPF* wurde in der gepoolten Analyse der INPULSIS Studien in der Nintedanib-Gruppe von 10,0 % der Studienteilnehmer als UE angegeben, verglichen mit 14,4 % in der Placebo-Gruppe. Der Unterschied ist statistisch signifikant zugunsten von Nintedanib (OR [95 % KI]: 0,66 [0,45; 0,96]; $p = 0,030$).
- *Husten* als UE trat in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig auf (13,3 % vs. 13,5 %) (OR [95 % KI]: 0,99 ([0,69; 1,42]; $p = 0,943$).
- *Dyspnoe* als UE wurde in den gepoolten INPULSIS Studien signifikant seltener bei Nintedanib-behandelten Studienteilnehmern (7,7 %) als bei Placebo-behandelten Studienteilnehmern (11,3 %) erfasst (OR [95 % KI]: 0,65; [0,43; 0,99]; $p = 0,042$).
- *Bronchitis* und *Nasopharyngitis* als UE fanden sich in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig: *Bronchitis* in der Nintedanib-Gruppe bei 10,5 % der Studienteilnehmer vs. 10,6 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,99 [0,66; 1,47]; $p = 0,943$). *Nasopharyngitis* in der Nintedanib-Gruppe bei 13,6 % der Studienteilnehmer vs. 16,1 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,82 [0,58; 1,16]; $p = 0,271$).

Unerwünschte Blutungsereignisse

In den INPULSIS Studien traten *Blutungsereignisse* im Nintedanib-Behandlungsarm numerisch häufiger (10,3 %) als im Placebo-Arm (7,8 %) auf (OR [95 % KI]: 1,36 [0,88; 2,11]; $p = 0,1631$). *Nicht schwerwiegende Epistaxis und Kontusionen* stellten die häufigsten *Blutungsereignisse* dar. Die Häufigkeit *schwerwiegender Blutungsereignisse* war in beiden Behandlungsgruppen gering und vergleichbar (Placebo: 1,4 %; Nintedanib: 1,3 %).

Unerwünschte Ereignisse mit einer Häufigkeit von > 10 %

Im Sinne der Vollständigkeit wurde geprüft, welche *UE* in den gepoolten INPULSIS Studien in mindestens einer der Gruppen mit einer Häufigkeit > 10 % auftraten. Hierbei wurde kein *UE* identifiziert, das nicht bereits in einer der oben genannten Kategorien nach Organsystemen von besonderem Interesse berücksichtigt wurden.

Subgruppenanalysen

Es kann von einem konstanten Therapieeffekt im Hinblick auf die dargestellten Wirksamkeitsendpunkte über die analysierten Subgruppen hinweg ausgegangen werden.

Die beobachteten Nachteile einzelner Endpunkte zu den unerwünschten Ereignissen sind für die gesamte Studienpopulation dargestellt.

Schlussfolgerungen zum Ausmaß des therapeutisch bedeutsamen Zusatznutzen

Bei Ofev[®] handelt es sich um ein Arzneimittel zur Behandlung eines seltenen Leidens. Der medizinische Zusatznutzen gilt gemäß § 35a Abs. 1 Satz 10 SGB V bereits durch die Zulassung als belegt. Lediglich das Ausmaß des Zusatznutzens muss für die Anzahl der Patienten, für die dieser Zusatznutzen besteht, auf Grundlage der Zulassungsstudien nachgewiesen werden.

Der therapeutische Bedarf in der Behandlung der IPF ist hoch. Ofev[®] (Nintedanib) wurde am 15.01.2015 im Rahmen eines „accelerated assessments“ als erstes Arzneimittel zur Behandlung aller Krankheitsschweregrade der IPF durch die Europäische Kommission zugelassen. Die US-amerikanische Gesundheitsbehörde FDA hatte Nintedanib bereits den Status eines Therapiedurchbruchs (Breakthrough Therapy Designation) erteilt und ebenfalls in einem beschleunigten Verfahren am 15. Oktober 2014 in den USA zugelassen. Gegenüber Placebo weist Nintedanib einen beträchtlichen Zusatznutzen in den Endpunktkategorien Mortalität (FVC) und Morbidität (Reduktion adjudizierter akuter Exazerbationen) auf. In der Endpunktkategorie Lebensqualität besteht ein nicht quantifizierbarer Zusatznutzen. Bezüglich der unerwünschten Ereignisse ist ein Nachteil anzunehmen.

Aus den dargestellten Daten geht hervor, dass bei der Behandlung mit Nintedanib für erwachsene Patienten mit IPF der Nutzen den potenziellen Nachteil überwiegt.

Tabelle 4-1: Ergebnisse zum Ausmaß des medizinischen Zusatznutzen

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 therapierelevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht quantifizierbaren Verlängerung der Lebensdauer	nicht quantifizierbar
Zeit bis 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 therapierelevanten 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 therapierelevanten 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 therapierelevanten Nutzens im Sinne einer auf Basis der wissenschaftlichen Datengrundlage nicht quantifizierbaren Reduktion akuter Exazerbationen	nicht quantifizierbar

Endpunkt	Effektschätzer [95 % KI] p-Wert	Beurteilung nach Arzneimittel- Nutzenbewertungsverordnung (AM-Nutzen V)	Ausmaß des Zusatznutzens
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		nicht quantifizierbar
SGRQ-I-(≤ -4 Punkte)-Responder	OR: 1,284 [0,95; 1,74] 0,1091		nicht quantifizierbar
SOBQ-Gesamtscore	MWD: -1,69 [-4,09; 0,70] 0,1648		nicht quantifizierbar
CASA-Q Husten-Symptome	MWD: 0,95 [-1,49; 3,38] 0,4450		nicht quantifizierbar
CASA-Q Husten-Impact	MWD: 1,73 [-0,46; 3,92] 0,1212		nicht quantifizierbar
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

Endpunkt	Effektschätzer [95 % KI] p-Wert	Beurteilung nach Arzneimittel- Nutzenbewertungsverordnung (AM-Nutzen V)	Ausmaß des Zusatznutzens
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

4.2 Methodik

Zusammenfassung des Abschnitts 4.2

Die Daten zum Nachweis zum Ausmaß des medizinischen Zusatznutzens von Nintedanib (Ofev[®]) basieren auf den pivotalen, randomisiert kontrollierten, doppelblinden Phase III-Studien (RCT) INPULSIS-1 und INPULSIS-2. Das Design und die Methodik der eingeschlossenen RCT wurden anhand der CONSORT-Kriterien dargestellt. Die Studienpopulation wurde anhand demografischer und krankheitsspezifischer Faktoren beschrieben. Verzerrungsaspekte wurden endpunktübergreifend für die gesamte eingeschlossene Studie und endpunktspezifisch bewertet. Die Bewertung folgte den Vorgaben der Dossievorlage. Die Patientenrelevanz und die Validität der Endpunkte bzw. Messinstrumente wurden überprüft und diskutiert.

Abschnitt 4.2 soll die Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens beschreiben. Der Abschnitt enthält Hilfestellungen für die Darstellung der Methodik sowie einige Vorgaben, die aus den internationalen Standards der evidenzbasierten Medizin abgeleitet sind. Eine Abweichung von diesen methodischen Vorgaben ist möglich, bedarf aber einer Begründung.

4.2.1 Fragestellung

Nach den internationalen Standards der evidenzbasierten Medizin soll eine Bewertung unter einer definierten Fragestellung vorgenommen werden, die mindestens folgende Komponenten enthält:

- Patientenpopulation
- Intervention
- Vergleichstherapie
- Endpunkte
- Studientypen

Unter Endpunkte sind dabei alle für die frühe Nutzenbewertung relevanten Endpunkte anzugeben (d.h. nicht nur solche, die ggf. in den relevanten Studien untersucht wurden).

Geben Sie die Fragestellung der vorliegenden Aufarbeitung von Unterlagen zur Untersuchung des medizinischen Nutzens und des medizinischen Zusatznutzens des zu bewertenden Arzneimittels an. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Am 15.01.2015 wurde von der EMA die Zulassung für Nintedanib zur Behandlung der IPF erteilt (European Commission (EC), 2015b). Das Anwendungsgebiet der Zulassung gemäß Fachinformation lautet: Ofev[®] wird angewendet bei Erwachsenen zur Behandlung der idiopathischen Lungenfibrose.

Ziel des Dossiers ist die Bewertung des Ausmaßes des medizinischen Zusatznutzens im Sinne des § 35a SGB V von Ofev® (Nintedanib) bei der Behandlung von erwachsenen Patienten mit IPF. Aufgrund des Status eines Arzneimittels zur Behandlung einer seltenen Erkrankung ist der medizinische Nutzen und medizinische Zusatznutzen mit der Zulassung von Nintedanib bereits festgestellt (§ 35a SGB V Abs. 1 Satz 10 Halbs. 1 SGB V). Nachweise zum medizinischen Nutzen und zum medizinischen Zusatznutzen im Verhältnis zur zweckmäßigen Vergleichstherapie müssen nicht vorgelegt werden (§ 35a SGB V Abs. 1 Satz 10 Halbs. 2 SGB V). Die Bewertung des Ausmaßes für den medizinischen Nutzen und Zusatznutzen erfolgt auf der Basis der Zulassung und der zugrunde liegenden multizentrischen, randomisierten, kontrollierten klinischen Studien der Phase III (INPULSIS-1 und INPULSIS-2) (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c, Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d). Die Bewertung erfolgt anhand der patientenrelevanten Nutzendimensionen *Mortalität, Morbidität, Lebensqualität* sowie *Unerwünschte Ereignisse*.

4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

Die Untersuchung der in Abschnitt 4.2.1 benannten Fragestellung soll auf Basis von klinischen Studien vorgenommen werden. Für die systematische Auswahl von Studien für diese Untersuchung sollen Ein- und Ausschlusskriterien für die Studien definiert werden. Dabei ist zu beachten, dass eine Studie nicht allein deshalb ausgeschlossen werden soll, weil keine in einer Fachzeitschrift veröffentlichte Vollpublikation vorliegt. Eine Bewertung der Studie kann beispielsweise auch auf Basis eines ausführlichen Ergebnisberichts aus einem Studienregister erfolgen, während ein Kongressabstract allein in der Regel nicht für eine Studienbewertung ausreicht.

Benennen Sie die Ein- und Ausschlusskriterien für Studien zum medizinischen Nutzen und Zusatznutzen. Machen Sie dabei mindestens Aussagen zur Patientenpopulation, zur Intervention, zur Vergleichstherapie, zu den Endpunkten, zum Studientyp und zur Studiendauer und begründen Sie diese. Stellen Sie die Ein- und Ausschlusskriterien zusammenfassend in einer tabellarischen Übersicht dar.

Entsprechend der Vorgaben für die Nutzenbewertung von Arzneimitteln zur Behandlung seltener Erkrankungen (§ 35a SGB V) erfolgt die Bewertung des Ausmaßes des Zusatznutzens auf Grundlage der zulassungsrelevanten Studien. Die Zulassung für Nintedanib (Ofev®) in der Indikation IPF beruht auf zwei RCTs (INPULSIS-1 und INPULSIS-2). Dementsprechend erfolgt die Darstellung des Ausmaßes für den Zusatznutzen auf Basis dieser beiden Studien.

4.2.3 Informationsbeschaffung

In den nachfolgenden Abschnitten ist zu beschreiben, nach welcher Methodik Studien identifiziert wurden, die für die Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens in dem in diesem Dokument bewerteten Anwendungsgebiet herangezogen werden. Dies bezieht sich sowohl auf publizierte als auch auf unpublizierte Studien. Die Methodik muss dazu geeignet sein, die relevanten Studien (gemäß den in Abschnitt 4.2.2 genannten Kriterien) systematisch zu identifizieren (systematische Literaturrecherche).

4.2.3.1 Studien des pharmazeutischen Unternehmers

Für die Identifikation der Studien des pharmazeutischen Unternehmers ist keine gesonderte Beschreibung der Methodik der Informationsbeschaffung erforderlich. Die vollständige Auflistung aller Studien, die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie aller Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, erfolgt in den Abschnitten 4.3.1 und 4.3.2, jeweils im Unterabschnitt „Studien des pharmazeutischen Unternehmers“. Die Darstellung soll auf Studien mit Patienten in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wird, beschränkt werden.

4.2.3.2 Bibliografische Literaturrecherche

Die Durchführung einer bibliografischen Literaturrecherche ist erforderlich, um sicherzustellen, dass ein vollständiger Studienpool in die Bewertung einfließt.

Eine bibliografische Literaturrecherche muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine bibliografische Literaturrecherche immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die bibliografische Literaturrecherche soll mindestens in den Datenbanken MEDLINE und EMBASE sowie in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Datenbanken (z. B. CINAHL, PsycINFO etc.) durchgeführt werden.

Die Suche soll in jeder Datenbank einzeln und mit einer für die jeweilige Datenbank adaptierten Suchstrategie durchgeführt werden. Die Suchstrategien sollen jeweils in Blöcken, insbesondere getrennt nach Indikation, Intervention und ggf. Studientypen, aufgebaut werden. Wird eine Einschränkung der Strategien auf bestimmte Studientypen vorgenommen (z. B. randomisierte kontrollierte Studien), sollen aktuelle validierte Filter hierfür verwendet werden. Alle Suchstrategien sind in Anhang 4-A zu dokumentieren.

Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Datenbanken eine bibliografische Literaturrecherche durchgeführt wurde. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Sprach- oder Jahreseinschränkungen), und begründen Sie diese.

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden (European Commission (EC)). Das Ausmaß des Zusatznutzens ist bei Orphan Drugs auf der Grundlage der Zulassung und der die Zulassung begründenden Studien zu bestimmen. Eine bibliographische Literaturrecherche nach weiteren Studien wurde daher nicht durchgeführt.

4.2.3.3 Suche in Studienregistern

Eine Suche in öffentlich zugänglichen Studienregistern ist grundsätzlich durchzuführen, um sicherzustellen, dass laufende Studien sowie abgeschlossene Studien auch von Dritten vollständig identifiziert werden und in Studienregistern vorliegende Informationen zu Studienmethodik und –ergebnissen in die Bewertung einfließen.

Eine Suche in Studienregistern muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche in Studienregistern immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Suche soll mindestens in den Studienregistern [clinicaltrials.gov](http://www.clinicaltrials.gov) (www.clinicaltrials.gov), EU Clinical Trials Register (EU-CTR, www.clinicaltrialsregister.eu), Klinische Prüfungen PharmNet.Bund (<http://www.pharmnet-bund.de/dynamic/de/klinische-pruefungen/index.htm>) sowie über das International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal, Suchportal der WHO: <http://apps.who.int/trialsearch/>) durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Studienregistern (z. B. krankheitsspezifische Studienregister oder Studienregister einzelner pharmazeutischer Unternehmen) durchgeführt werden. Die Suche in Studienregistern anderer pharmazeutischer Unternehmen ist insbesondere bei indirekten Vergleichen sinnvoll, wenn Studien zu anderen Arzneimitteln identifiziert werden müssen.

Die Suche soll in jedem Studienregister einzeln und mit einer für das jeweilige Studienregister adaptierten Suchstrategie durchgeführt werden. Die Suche soll abgeschlossene, abgebrochene und laufende Studien erfassen. Alle Suchstrategien sind in Anhang 4-B zu dokumentieren.

Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Studienregistern die Suche durchgeführt wurde. Begründen Sie dabei Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Jahreseinschränkungen), und begründen Sie diese.

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status´ in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden (European Commission (EC)). Das Ausmaß des Zusatznutzens ist bei Orphan Drugs auf der Grundlage der Zulassung und der die Zulassung begründenden Studien zu bestimmen. Eine Suche in Studienregistern nach weiteren Studien wurde daher nicht durchgeführt.

4.2.3.4 Selektion relevanter Studien

Beschreiben Sie das Vorgehen bei der Selektion relevanter Studien aus dem Ergebnis der in den Abschnitten 4.2.3.2 und 4.2.3.3 beschriebenen Rechenschritte. Begründen Sie das Vorgehen, falls die Selektion nicht von zwei Personen unabhängig voneinander durchgeführt wurde.

Aufgrund des Orphan Drug Status´ von Nintedanib (Ofev[®]) wurde weder eine bibliographische Literaturrecherche noch eine Suche in Studienregistern durchgeführt. Aus diesem Grund war keine Selektion relevanter Studien notwendig. Das Ausmaß des Zusatznutzens wird auf der Grundlage der Zulassung und der die Zulassung begründenden Studien definiert (Gemeinsamer Bundesausschuss, 2013). Analog zum Vorgehen der Zulassungsbehörden wird im vorliegenden Dossier zur frühen Nutzenbewertung von Nintedanib in der Indikation IPF auf der Basis der beiden Phase III-Studien (INPULSIS-1 und INPULSIS-2) das Ausmaß des Zusatznutzens abgeleitet.

4.2.4 Bewertung der Aussagekraft der Nachweise

Zur Bewertung der Aussagekraft der im Dossier vorgelegten Nachweise sollen Verzerrungsaspekte der Ergebnisse für jede eingeschlossene Studie beschrieben werden, und zwar separat für jeden patientenrelevanten Endpunkt. Dazu sollen insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte systematisch extrahiert werden (zur weiteren Erläuterung der einzelnen Aspekte siehe Bewertungsbogen in Anhang 4-F):

A: Verzerrungsaspekte der Ergebnisse auf Studienebene

- Erzeugung der Randomisierungssequenz (*bei randomisierten Studien*)
- Verdeckung der Gruppenzuteilung (*bei randomisierten Studien*)
- zeitliche Parallelität der Gruppen (*bei nicht randomisierten vergleichenden Studien*)
- Vergleichbarkeit der Gruppen bzw. Berücksichtigung prognostisch relevanter Faktoren (*bei nicht randomisierten vergleichenden Studien*)
- Verblindung des Patienten sowie der behandelnden Personen
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

Für randomisierte Studien soll darüber hinaus das Verzerrungspotenzial bewertet und als „niedrig“ oder „hoch“ eingestuft werden. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Eine zusammenfassende Bewertung der Verzerrungsaspekte soll nicht für nicht randomisierte Studien erfolgen.

Für die Bewertung eines Endpunkts soll für randomisierte Studien zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter A aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft werden. Falls diese Einstufung als „hoch“ erfolgt, soll das Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet werden, Abweichungen hiervon sind zu begründen. Ansonsten sollen die unter B genannten endpunktspezifischen Aspekte Berücksichtigung finden.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ soll nicht zum Ausschluss der Daten führen. Die Klassifizierung soll vielmehr der Diskussion heterogener Studienergebnisse und der Einschätzung der Aussagekraft der Nachweise dienen. Für nicht randomisierte Studien können für solche Diskussionen einzelne Verzerrungsaspekte herangezogen werden.

Beschreiben Sie die für die Bewertung der Verzerrungsaspekte und des Verzerrungspotenzials eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Die Bewertung der eingeschlossenen Studien hinsichtlich ihrer Aussagekraft wurde gemäß der Vorgaben der Dossievorlage anhand der zur Verfügung stehenden Informationen aus den Studienberichten vorgenommen. Sie erfolgte zunächst endpunktübergreifend auf Studienebene, anschließend für jeden patientenrelevanten Endpunkt und gegebenenfalls für jede Operationalisierung separat.

Es wurden die folgenden endpunktübergreifenden (A) und endpunktspezifischen (B) Aspekte systematisch extrahiert und bewertet (Anhang 4-F):

A: Verzerrungsaspekte der Ergebnisse auf Studienebene:

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung der Patienten sowie des behandelnden Arztes
- Ergebnisgesteuerte Berichterstattung
- Sonstige Aspekte

B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene:

- Verblindung der Endpunkterheber
- Umsetzung des Intention to treat (ITT)-Prinzips
- Ergebnisgesteuerte Berichterstattung
- Sonstige Aspekte

Gemäß der Verfahrensordnung (VerfO) des Gemeinsamen Bundesausschuss (G-BA) wurde das Verzerrungspotenzial auf Studienebene für RCT zusammenfassend als „niedrig“ oder „hoch“ eingestuft.

Zur Bewertung eines Endpunkts wurde für randomisierte Studien zunächst das Verzerrungspotenzial endpunktübergreifend als „niedrig“ oder „hoch“ eingestuft. Falls die Einstufung als „hoch“ erfolgte, wurde auch das Verzerrungspotenzial für den Endpunkt als „hoch“ bewertet.

Die Einstufung „hoch“ für das Verzerrungspotenzial eines Endpunktergebnisses führte nicht zum Ausschluss der Daten. Die Klassifizierung diente zur Einschätzung der Aussagekraft der Nachweise. Unter „sonstige Aspekte“ wurden z. B. relevante Dateninkonsistenzen innerhalb der Studienunterlagen oder zwischen den Studienunterlagen oder die Anwendung inadäquater statistischer Verfahren bewertet. Die Vorgehensweise bei der Bewertung des Verzerrungspotenzials für jede Studie und jeden Endpunkt (Operationalisierung) richtete sich nach den Hinweisen der Bewertungsbögen (Anhang 4-F).

4.2.5 Informationssynthese und -analyse

4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien

Das Design und die Methodik der eingeschlossenen Studien soll in den Abschnitten 4.3.1 und 4.3.2, jeweils in den Unterabschnitten „Charakteristika der in die Bewertung eingeschlossenen Studien“ und den dazugehörigen Anhängen, dargestellt werden. Die Darstellung der Studien soll für randomisierte kontrollierte Studien mindestens die Anforderungen des CONSORT-Statements erfüllen (Items 2b bis 14, Informationen aus dem CONSORT-Flow-Chart)¹. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-² bzw. STROBE-Statements³ folgen. Design und Methodik weiterer Untersuchungen sollen gemäß den verfügbaren Standards dargestellt werden.

Beschreiben Sie, nach welchen Standards und mit welchen Informationen (Items) Sie das Design und die Methodik der eingeschlossenen Studien in Modul 4 dargestellt haben. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Die Darstellung von Design und Methodik erfolgte bei den zugrundeliegenden RCT (INPULSIS Studien) entsprechend der Anforderungen des CONSORT-Statements (Items 2b bis 14 sowie CONSORT-Flussdiagramm). Die Anwendung des CONSORT-Statements stützte sich neben der Checkliste zusätzlich auf weitere Ausführungen zu den einzelnen Items. Danach sollten das Ziel der Studie und die dazugehörigen Hypothesen, das Studiendesign, die Studienpopulation, die Interventionen und die betrachteten Zielgrößen mittels hinreichender Kriterien klar definiert sein (Item 2b-6). Zur Bewertung der Aussagekraft der Studien waren die Details zur Bestimmung der Fallzahl, die Erzeugung der Randomisierungssequenz, die Zuteilung der Studienteilnehmer zu den Behandlungsarmen, die Verblindung und zuletzt die Verwendung adäquater statistischer Analysemethoden darzustellen (Item 7-12). Darüber hinaus wurden Angaben zur Zahl der eingeschlossenen, behandelten und analysierten Studienteilnehmer gemacht. Studienabbrüche oder Ausschlüsse von Studienteilnehmer nach der Randomisierung wurden klar und anschaulich beschrieben. Zudem wurde der Zeitraum, in dem die Rekrutierung und die Studie selbst stattfanden, angegeben, um diese später möglicherweise in einen historischen Zusammenhang stellen zu können (Item 13-14) (Moher et al., 2010, Schulz et al., 2010).

Als Informationsquellen für die Beschreibung des Designs und der Methodik der in die Nutzenbewertung eingeschlossenen Studien wurden Studienbericht, Protokoll und statistischer Analyseplan herangezogen.

¹ Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.

² Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Publ Health* 2004; 94(3): 361-366.

³ Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.

Die Methodik der eingeschlossenen Studien, die Interventionen sowie die Charakteristika der Studienpopulationen wurden in Tabelle 4-7, Tabelle 4-8 und Tabelle 4-9 dargestellt.

4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien

Die Ergebnisse der einzelnen Studien sollen in den Abschnitten 4.3.1 und 4.3.2 in den entsprechenden Unterabschnitten zunächst für jede eingeschlossene Studie separat dargestellt werden. Die Darstellung soll die Charakteristika der Studienpopulationen sowie die Ergebnisse zu allen in den eingeschlossenen Studien berichteten patientenrelevanten Endpunkten (Verbesserung des Gesundheitszustands, Verkürzung der Krankheitsdauer, Verlängerung des Überlebens, Verringerung von Nebenwirkungen, Verbesserung der Lebensqualität) umfassen. Anforderungen an die Darstellung werden in den Unterabschnitten beschrieben.

Benennen Sie die Patientencharakteristika und patientenrelevanten Endpunkte, die in den relevanten Studien erhoben wurden. Begründen Sie, wenn Sie von den oben benannten Vorgaben abgewichen sind. Beschreiben Sie für jeden Endpunkt, warum Sie ihn als patientenrelevant einstufen, und machen Sie Angaben zur Validität des Endpunkts (z. B. zur Validierung der eingesetzten Fragebögen). Geben Sie für den jeweiligen Endpunkt an, ob unterschiedliche Operationalisierungen innerhalb der Studien und zwischen den Studien verwendet wurden. Benennen Sie die für die Bewertung herangezogene(n) Operationalisierung(en) und begründen Sie die Auswahl. Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.

Relevant für die Bewertung waren die Phase III-Studien INPULSIS-1 und INPULSIS-2, die nachfolgend beschrieben werden.

Patientencharakteristika

Die Studienpopulation (erwachsene Patienten mit IPF) wurde durch die folgenden demografischen Daten sowie krankheitsspezifischen Charakteristika zu Studienbeginn beschrieben:

Demografie

- Alter (Jahre)
- Geschlecht (männlich, weiblich)
- Gewicht (kg)
- Body-Mass-Index (BMI) (kg/m²)
- Ethnie (kaukasisch, asiatisch)
- Raucherstatus (Nichtraucher, Ex-Raucher, Raucher)
- Zeit seit IPF-Diagnose (Jahre)

Krankheitsspezifische Charakteristika:

- Radiologische Beurteilung der IPF (usual interstitial pneumonia (UIP), vermutlich UIP, mögliche UIP, keine UIP)
- Lungenbiopsie (ja, nein)
- FVC (ml, % vom Sollwert)
- forcierte expiratorische Einsekunden-Kapazität (FEV₁)/FVC-Verhältnis (%)
- Kohlenmonoxid-Diffusionskapazität (DLco) (mmol/min/kPa)
- Sauerstoffsättigung (SpO₂) (%)
- SGRQ-Gesamtscore (Punkte)
- Systemische Kortikosteroid-Therapie (ja, nein)

Die Patientenflüsse der berücksichtigten Studien wurden in Flow-Charts dargestellt (Abbildung 4-13, Abbildung 4-14), die die Verteilung der Studienteilnehmer auf die Behandlungsgruppen sowie deren Gründe für einen ggf. vorzeitigen Studienabbruch enthalten.

Endpunkte

Im vorliegenden Dossier wurden die Ergebnisse der folgenden patientenrelevanten Endpunkte dargestellt:

- **Mortalität**

- *Zeit bis Tod, unabhängig von der Ursache*
- *Zeit bis respiratorisch bedingtem Tod*
- *Jährliche FVC-Abnahme (ml)*
- *Anteil FVC-Responder (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert)*

- **Morbidität**

- *Zeit bis zur ersten akuten Exazerbation*
- *Zeit bis zur ersten adjudizierten akuten Exazerbation*

- **Lebensqualität**

Krankheitsspezifische Messinstrumente

- *Veränderung SGRQ-Gesamtscore*
- *Veränderung SGRQ-I-Gesamtscore*
- *SGRQ-(≤ -4 Punkte)-Responder*
- *SGRQ-I-(≤ -4 Punkte)-Responder*
- *Veränderung SOBQ-Gesamtscore*
- *Veränderung CASA-Q Husten-Symptome*
- *Veränderung CASA-Q Husten-Impact*

Generische Messinstrumente

- *PGI-C Responder*
- *Veränderung EQ-5D-VAS*

- **Unerwünschte Ereignisse**

- *UE*
- *SUE*
- *UE, die zum Therapieabbruch führten (UEA)*
- *UE, die zum Tod führten*
- *UE von besonderem Interesse*
- *Weitere häufige UE (Inzidenz > 10 % in mindestens einer Behandlungsgruppe)*

Mortalität

Die IPF ist eine schwerwiegende und in der Regel zum Tod führende Erkrankung (Behr et al., 2013). Somit ist die Verlängerung der Gesamtüberlebenszeit ein wichtiges Therapieziel und daher patientenrelevant. Auch seitens des Institutes für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und des G-BA ist die Verlängerung des Überlebens als patientenrelevante Nutzendimension anerkannt. Aufgrund der eindeutigen und klaren Bestimmbarkeit ist der Überlebensstatus auch im Fall des vorzeitigen Studienabbruchs eines Patienten in der Regel zu späteren Zeitpunkten noch zweifelsfrei feststellbar.

Effekte der Behandlung mit Nintedanib auf die Mortalität wurden unmittelbar durch die Endpunkte *Zeit bis Tod, unabhängig von der Ursache* und *Zeit bis respiratorisch bedingtem Tod* sowie indirekt durch die validierten Surrogatendpunkte *jährliche FVC-Abnahme* (ml) und *Anteil FVC-Responder* (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert) erfasst und bewertet. Zur Validität dieser Surrogatendpunkte siehe Abschnitt 4.5.4.

Zeit bis Tod***Zeit bis Tod, unabhängig von der Ursache***

Die *Zeit bis Tod, unabhängig von der Ursache* wurde als die Zeit von Randomisierung bis zum Todeszeitpunkt eines Studienteilnehmers definiert. Überlebende Studienteilnehmer wurden zum Zeitpunkt des letzten Patientenkontaktes (längstenfalls 372 Tage) zensiert. Für diesen Endpunkt wurden tabellarisch die Anzahl der Studienteilnehmer unter Risiko, die Anzahl und der Anteil Studienteilnehmer mit dem betreffenden Ereignis angegeben. Aufgrund der niedrigen Ereignisrate konnte kein Median angegeben werden. Stattdessen wurde die Wahrscheinlichkeit, das betreffende Ereignis nicht zu erleiden, anhand des Kaplan-Meier-Schätzers pro Behandlungsgruppe zu Tag 372 angegeben. Zusätzlich zur tabellarischen Darstellung wurden Kaplan-Meier-Kurven erstellt und präsentiert. Der Behandlungseffekt wurde als HR mit 95 %-Konfidenzintervall und p-Wert, basierend auf Cox-Regression, angegeben. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitts 4.3.1 als Fußnoten.

Zeit bis respiratorisch bedingtem Tod

Die *Zeit bis respiratorisch bedingtem Tod* wurde als die Zeit von Randomisierung bis zum Zeitpunkt des Todes aufgrund einer respiratorisch bedingten Ursache eines Studienteilnehmers definiert. Alle Todesfälle wurden hinsichtlich ihrer Ursache durch ein verblindetes IAC bewertet, dem drei ausgewiesene 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 (längstenfalls 372 Tage) zensiert. Auswertung und Darstellung der Ergebnisse zu diesem Endpunkt erfolgten analog zu derjenigen für *Zeit bis Tod, unabhängig von der Ursache*.

Forcierte Vitalkapazität (FVC)

Die FVC ist das Lungenvolumen, das nach maximaler Einatmung forciert ausgeatmet werden kann. Bei der IPF kommt es aufgrund der zunehmenden Vernarbung der Lunge zu einer mit der Zeit fortschreitenden Abnahme der FVC, sodass die Belüftung der Lunge zunehmend eingeschränkt wird bzw. eine erhebliche Atemanstrengung des Patienten erfordert. Da die Vernarbungsprozesse bei der IPF 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 folglich das Mortalitätsrisiko zu verringern. Die FVC ist der am häufigsten verwendete physiologische Messparameter in klinischen Studien in der Pneumologie (Taniguchi et al., 2010). Eine detaillierte Beschreibung der FVC als Surrogatendpunkt für die Mortalität erfolgt in Abschnitt 4.5.4.

Jährliche FVC-Abnahme (ml)

Die Bestimmung der FVC erfolgte bei Studienbeginn sowie zu Woche 2, 4, 6, 12, 24, 36 und 52 mittels Spirometrie unter Einhaltung der Kriterien der American Thoracic Society/European Respiratory Society (ATS/ERS) (Miller et al., 2005). Die *jährliche FVC-Abnahme* (ml) wurde mittels Random-Koeffizienz-Modell unter Berücksichtigung aller vorhandenen Messwerte berechnet. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitts 4.3.1 als Fußnoten. Insbesondere berücksichtigt dieses Modell auch Daten von Studienteilnehmern mit fehlenden Messungen. Die Anzahl der analysierten Studienteilnehmer sowie die beobachteten Mittelwerte (MW) mit Standardabweichung (SD) zu Studienbeginn und -ende (Woche 52) wurden tabellarisch dargestellt. Weiterhin zeigt die Tabelle die auf dem Random-Koeffizienz-Modell basierende Veränderung der FVC von Studienbeginn bis Woche 52 pro Behandlungsgruppe als adjustierte MWD mit Standardfehler (SE) und den daraus resultierenden Behandlungseffekt ebenfalls als adjustierte MWD mit SE zwischen den Behandlungsgruppen mit 95 %-Konfidenzintervall und p-Wert.

FVC-Responder (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert)

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. Falls zu Woche 52 kein Messwert vorlag, wurde der entsprechende Studienteilnehmer konservativ als Non-Responder bewertet (siehe Abschnitt 4.5.4 bezüglich der Validität der gewählten Schwellenwerte von 10 % und 5 % für den Surrogatendpunkt FVC).

Der Endpunkt *FVC-Responder* (Schwellenwerte 10 % vom Sollwert und 5 % vom Sollwert) wurde mittels logistischer Regression ausgewertet. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitt 4.3.1 als Fußnoten. Die tabellarische Ergebnisdarstellung zeigt die Anzahl der Studienteilnehmer, die in die Analyse eingegangen sind, sowie die Anzahl und den Anteil der *FVC-Responder*. Der Behandlungseffekt wird durch das Odds Ratio (OR) mit 95 %-Konfidenzintervall und p-Wert dargestellt.

Morbidität

Die IPF geht mit einer erheblichen Beeinträchtigung des Gesundheitszustands einher. Diese Beeinträchtigung hat zwei wesentliche Komponenten: Einerseits die alltäglich vom Patienten verspürten Symptome, andererseits das Erleben krisenhafter Verschlechterungen des Gesundheitszustands durch akute Exazerbationen. Effekte der Behandlung mit Nintedanib auf den Gesundheitszustand wurden in den eingeschlossenen Studien mithilfe der Erfassung der patientenrelevanten Ereignisse *akute Exazerbationen* und *adjudizierte akute Exazerbationen* beurteilt. Darüber hinaus wurde die Beeinträchtigung durch Symptome mithilfe von Fragebogeninstrumenten erfasst, die die Auswirkungen der Symptomintensität auf die Lebensqualität abbilden. Daher werden letztere nicht in der Nutzenkategorie Morbidität sondern unter Verbesserung der Lebensqualität betrachtet.

Akute Exazerbationen

Akute Exazerbationen sind akut lebensbedrohliche Ereignisse, welche bei Patienten mit IPF mit einer Inzidenz von 5-15 % pro Jahr auftreten (King et al., 2008, Azuma et al., 2005, Song et al., 2011, Collard et al., 2007b). Sie können in jedem Erkrankungsstadium, jedoch vermehrt bei höherem Schweregrad der IPF, ohne Vorwarnung auftreten. Leitsymptom der *akuten Exazerbation* ist das Neuauftreten oder die Zunahme einer Dyspnoe. Weitere Symptome können Verschlimmerung von Husten, Auftreten von Fieber und ggf. vermehrtes Sputum sein (Kim et al., 2006, Parambil et al., 2005, Johannson and Collard, 2013). Ätiologisch werden sie als akute, intrinsische Akzeleration des IPF Krankheitsfortschreitens gewertet (Johannson and Collard, 2013).

Klinisch führen *akute Exazerbationen* unmittelbar zu einer rapiden Verschlechterung des Gesundheitszustands der Patienten und gehen in den meisten Fällen mit einer stationären Behandlung einher (Brown et al., 2015). Überdies kommt es aufgrund des Alveolarschadens und der Zunahme der Fibrosierung zu einer Verschlechterung der Lungenfunktionsparameter (Johansson and Collard, 2013). *Akute Exazerbationen* sind mit einer erhöhten Mortalitätsrate assoziiert: In einer retrospektiven Studie lag das mediane Überleben von Patienten mit einer akuten Exazerbation bei 2,2 Monaten und die Letalität bei stationärer Behandlung bei 50 % (Song et al., 2011). Innerhalb eines Jahres nach einer Exazerbation liegt die Sterblichkeit bei 80 %-90 % der betroffenen Patienten (Collard et al., 2007b). Schwere und Häufigkeit akuter Exazerbationen stellen somit ein wichtiges prognostisches Kriterium dar. Daher sind akute Exazerbationen patientenrelevante Ereignisse, und deren Vermeidung ist wichtiges Therapieziel.

Eine *akute Exazerbation* war in den berücksichtigten Studien definiert als eine erhebliche akute Verschlechterung der respiratorischen klinischen Situation des Patienten, die auf keine andere Ursache als die zugrunde liegende IPF zurück geführt werden konnte. Zu den Kriterien einer akuten Exazerbationen zählten gemäß Studienprotokollen:

- 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)
- Ausschluss von Infektionen
- Ausschluss anderer möglicher Erklärungen, wie beispielsweise Linksherzversagen, Lungenembolie oder akute Lungenschädigung bekannter Ursache

Die Details hinsichtlich aller klinischen Daten bei Verdacht einer *akuten Exazerbation* wurden in den berücksichtigten Studien durch einen spezifischen Fragebogen dokumentiert. Zu den klinisch relevanten Daten gehören:

- Zusammenfassung des klinischen Verlaufs, der Anzeichen und Symptome
- Ergebnisse der Laborparameter, Lungenfunktion und Bildgebung
- notwendige Behandlung

Adjudizierte akute Exazerbationen

Bei *akuten Exazerbationen* handelt es sich um in klinischen Studien relativ seltene Ereignisse. Zudem stellt ihre Abgrenzung zu akuten Verschlechterungen des Gesundheitszustands aufgrund anderer medizinischer Ursachen eine besondere Herausforderung dar. Aus diesem Grund ist die alleinige Bewertung durch den Prüfarzt, ob ein Verdachtsereignis tatsächlich eine akute Exazerbation ist oder nicht, mit einer gewissen Unsicherheit behaftet. Daher wurden die durch die Prüfarzte erhobenen Verdachtsereignisse adjudiziert: Die Daten für jede vom Prüfarzt vermutete akute Exazerbation wurden durch ein verblindetes IAC überprüft, dem drei ausgewiesene Experten aus dem Bereich der Pneumologie mit besonderen Fachkenntnissen in der Behandlung der IPF angehörten. Jedes aufgetretene Ereignis wurde einer der drei Kategorien *bestätigte Exazerbation*, *vermutete Exazerbation* oder *keine Exazerbation* zugeordnet. *Bestätigte* oder *vermutete akute Exazerbationen* wurden dem Endpunkt *adjudizierte akute Exazerbationen* als Ereignis zugeordnet.

Die Operationalisierung *akute Exazerbationen* und *adjudizierte akute Exazerbationen* erfolgte als *Zeit bis zur ersten akuten IPF-Exazerbation* und *Zeit bis zur ersten adjudizierten akuten IPF-Exazerbation*.

Zeit bis zur ersten akuten IPF-Exazerbation

Der Endpunkt *Zeit bis zur ersten akuten Exazerbation* wurde als die Zeit von Randomisierung bis zum Auftreten einer für den betroffenen Studienteilnehmer ersten vom Prüfarzt berichteten *akuten Exazerbation* definiert. Studienteilnehmer ohne Exazerbationsereignis wurden zum Zeitpunkt des letzten Patientenkontaktes (längstenfalls 372 Tage) zensiert. Die Anzahl der Studienteilnehmer unter Risiko, die Anzahl und der Anteil Studienteilnehmer mit dem betreffenden Ereignis wurden tabellarisch dargestellt. Aufgrund der niedrigen Ereignisrate konnte kein Median angegeben werden. Stattdessen wurde die Wahrscheinlichkeit, das betreffende Ereignis nicht zu erleiden, anhand des Kaplan-Meier-Schätzers pro Behandlungsgruppe zu Tag 372 angegeben. Der Behandlungseffekt wurde als HR mit 95 %-Konfidenzintervall und p-Wert, basierend auf Cox-Regression, angegeben. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitt 4.3.1 als Fußnoten.

Zeit bis zur ersten adjudizierten akuten IPF-Exazerbation

Der Endpunkt *Zeit bis zur ersten adjudizierten akuten Exazerbation* wurde analog zum Endpunkt *Zeit bis zur ersten akuten Exazerbation* operationalisiert. Darüber hinaus wurden Kaplan-Meier-Kurven erstellt und präsentiert.

Gesundheitsbezogene Lebensqualität

Die IPF ist eine schwere chronische Erkrankung, die mit zunehmender Dauer mit einer erheblichen Einschränkung der gesundheitsbezogenen Lebensqualität (HRQoL) einhergeht. Deren Verbesserung oder Erhalt stellt daher ein wichtiges Therapieziel dar. Für G-BA und IQWiG gehört die Verbesserung der Lebensqualität zu den zentralen betrachteten Nutzendimensionen. Die HRQoL der Studienteilnehmer wird in klinischen Studien anhand von validierten Fragebogeninstrumenten sowohl spezifisch für ein bestimmtes Krankheitsbild oder Organsystem als auch generisch erfasst.

In den berücksichtigten Studien wurde die HRQoL mittels der folgenden spezifischen Fragebogeninstrumente beurteilt.

Krankheitsspezifische Messinstrumente:

- *SGRQ*
- *SGRQ-I*
- *SOBQ*
- *CASA-Q*

Generische Messinstrumente:

- *PGI-C*
- *EQ-5D-VAS*

Bei der Beurteilung der krankheitsspezifischen und der generischen Lebensqualität mittels der oben aufgeführten Fragebögen handelt es sich um weltweit eingesetzte und validierte Instrumente zur Bewertung der Lebensqualität (Aaronson, 1993, Osoba et al., 1998, Velikova et al., 2010).

St. George's Respiratory Questionnaire (SGRQ)

Der *SGRQ* ist ein mehrdimensionales Erhebungsinstrument zur Dokumentation der krankheitsspezifischen Lebensqualität speziell für Patienten mit einer respiratorischen Erkrankung. Dieser validierte Fragebogen wurde von Jones et al. ursprünglich für die chronisch obstruktive Lungenerkrankung (COPD) entwickelt (Jones et al., 1992). Betrachtet wird die Lebensqualität bezogen auf 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). Der *SGRQ* ist auch für Patienten mit Bronchiektasien validiert und wird ebenfalls bei Erkrankungen wie Sarkoidose und zystischer Fibrose erfolgreich eingesetzt (Wilson et al., 1997). Im Rahmen der frühen Nutzenbewertung von Pirfenidon hat das IQWiG den *SGRQ* als validiertes Instrument zur Bewertung der HRQoL von Patienten mit IPF anerkannt (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2011).

Die Patienten beantworten den Fragebogen selbstständig. 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. Eine Verminderung des *SGRQ-Gesamtscores* um vier Punkte im Vergleich zum Ausgangswert ist allgemein als MCID der subjektiv empfundenen Lebensqualität akzeptiert (Swigris et al., 2010, Jones, 2014, Jones et al., 2011).

IPF-spezifischer St. George's Respiratory Questionnaire (SGRQ-I)

Mit dem *SGRQ-I* wurde eine IPF-spezifische Form des *SGRQ* entwickelt, die auf Basis des erhobenen *SGRQ* durch einen präspezifizierten Algorithmus ermittelt werden kann (Yorke et al., 2010). Das Ziel dieser Weiterentwicklung war laut Autoren die Generierung einer IPF-spezifischen *SGRQ*-Fragebogenversion, die spezifisch die Einschränkungen der HRQoL durch die IPF abbildet. Dieser für IPF-Patienten angepasste Fragebogen wurde 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. (Yorke et al., 2010, Swigris et al., 2010). Der *SGRQ-I* besteht ebenfalls 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/schlechte Lebensqualität) ermittelt. Für die Indikation IPF gibt es derzeit noch keine allgemein akzeptierte MCID. Um eine Responder-Analyse durchführen zu können, wurde in Anlehnung an die MCID für einen klinisch relevanten Effekt bzgl. des *SGRQ-Gesamtscores* auch für den *SGRQ-I* ein Schwellenwert von -4 Punkte als MCID festgelegt.

Die Bewertung der Effekte von Nintedanib auf die HRQoL von IPF-Studienteilnehmern mittels *SGRQ* und *SGRQ-I* erfolgte in den berücksichtigten Studien mittels der folgenden Operationalisierungen:

Veränderung des SGRQ-Gesamtscores

Die Bestimmung des *SGRQ* erfolgte bei Studienbeginn sowie zu Woche 4, 6, 12, 24 und 52. Die *Veränderung des SGRQ-Gesamtscores* wurde mittels eines Mixed Effect Model for Repeated Measures (MMRM) unter Berücksichtigung aller vorhandenen Messwerte berechnet. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitts 4.3.1 als Fußnoten. Insbesondere berücksichtigt dieses Modell auch Daten von Studienteilnehmern mit fehlenden Messungen. Die Anzahl der analysierten Studienteilnehmer sowie die beobachteten MW mit SD zu Studienbeginn und –ende (Woche 52) wurden tabellarisch dargestellt. Weiterhin zeigt die Tabelle die auf dem MMRM basierende *Veränderung des SGRQ-Gesamtscores* von Studienbeginn bis Woche 52 pro Behandlungsgruppe als adjustierte MWD mit SE und den daraus resultierenden Behandlungseffekt ebenfalls als adjustierte MWD mit SE zwischen den Behandlungsgruppen mit 95 %-Konfidenzintervall und p-Wert.

SGRQ-(≤ -4 Punkte)-Responder

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. Der Endpunkt Anteil *SGRQ-(≤ -4 Punkte)-Responder* wurde mittels logistischer Regression ausgewertet. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitts 4.3.1 als Fußnoten. Die tabellarische Ergebnisdarstellung zeigt die Anzahl der Studienteilnehmer, die in die Analyse eingegangen sind, sowie die Anzahl und den Anteil der *SGRQ-(≤ -4 Punkte)-Responder*. Der Behandlungseffekt wird durch das OR mit 95 %-Konfidenzintervall und p-Wert dargestellt.

Veränderung des SGRQ-I-Gesamtscores

Die Bestimmung des *SGRQ-I* erfolgte, wie im Studienprotokoll präspezifiziert, aus den vorliegenden *SGRQ*-Daten. Die Auswertung und Darstellung der *Veränderung des SGRQ-I-Gesamtscores* erfolgte analog zu derjenigen für die *Veränderung des SGRQ-Gesamtscores*.

SGRQ-I-(≤ -4 Punkte)-Responder

Die Bestimmung der *SGRQ-I-(≤ -4 Punkte)-Responder* erfolgte analog zur Bestimmung der *SGRQ-(≤ -4 Punkte)-Responder*. Dieser Endpunkt war Post hoc definiert.

Shortness of Breath Questionnaire (SOBQ)

Beim *SOBQ* handelt es sich um ein validiertes Fragebogeninstrument, das die Selbsteinschätzung der Beeinträchtigung der Aktivitäten des täglichen Lebens durch Dyspnoe widerspiegelt (Archibald and Guidotti, 1987, Eakin et al., 1998). Eine kürzlich veröffentlichte Arbeit unterstützt die Validität dieses ursprünglich für COPD-Patienten entwickelten Messinstruments für IPF-Patienten (Gries et al., 2013). Die Beurteilung der Schwere der Atemnot während täglicher Aktivitäten erfolgt auf einer 120-Punkte-Skala, wobei ein hoher Gesamtscore eine schwere Beeinträchtigung durch Atemnot reflektiert. Als MCID wurde eine Veränderung von fünf Punkten angegeben (Kupferberg et al., 2005).

Veränderung des SOBQ-Gesamtscores

Die Bestimmung des *SOBQ-Gesamtscores* erfolgte bei Studienbeginn sowie zu Woche 4, 6, 12, 24 und 52. Die *Veränderung des SOBQ-Gesamtscores* wurde analog zum Vorgehen für die *Veränderung des SGRQ-Gesamtscores* berechnet und dargestellt.

Cough and Sputum Assessment Questionnaire (CASA-Q)

Der *CASA-Q* ist in der Indikation COPD und chronische Bronchitis ein validiertes Fragebogeninstrument und dient der Evaluierung der Intensität der klinischen Symptome *Husten* und *Sputum* (*Husten-Symptome*, *Sputum-Symptome*) sowie deren Auswirkungen auf die HRQoL (*Husten-Impact*, *Sputum-Impact*) (Crawford et al., 2008). Die Validität dieses ursprünglich für COPD-Patienten entwickelten Messinstruments wurde kürzlich für IPF-Patienten bestätigt (Gries et al., 2013). Die Beurteilung erfolgt auf einer 100-Punkte-Skala, wobei ein geringerer Wert eine stärkere Intensität bzw. Beeinträchtigung durch die genannten Symptome anzeigt.

Veränderung des CASA-Q (Husten-Symptome bzw. Husten-Impact)

In den berücksichtigten Studien wurden ausschließlich die beiden Husten-Domänen (*Husten-Symptome* und *Husten-Impact*) des *CASA-Q* erhoben, da Sputum als Symptom bei der IPF – im Gegensatz zur COPD – eine untergeordnete Rolle spielt. Die Bestimmung des *CASA-Q* erfolgte bei Studienbeginn sowie zu Woche 4, 6, 12, 24 und 52. Die *Veränderung des CASA-Q (Husten-Symptome bzw. Husten-Impact)* wurde analog zum Vorgehen für die *Veränderung des SGRQ-Gesamtscores* berechnet und dargestellt.

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

Seit den Anfängen der Heilkunst ist die Frage an den Patienten nach dem aktuellen Befinden im Vergleich zum Zustand vor Beginn einer Behandlung ein zentrales Element bei der Wirksamkeitsbewertung durch den Arzt. Der *PGI-C* stellt den Versuch dar, die Antwort auf diese ureigenste ärztliche Frage („*Wie geht es Ihnen im Vergleich zu früher?*“) im Kontext einer klinischen Studie in standardisierter Weise zu dokumentieren. Mit Hilfe des *PGI-C* gibt der Patient seinen subjektiven Eindruck zur Verbesserung oder Verschlechterung seines Gesundheitszustands im Vergleich zum Referenzzeitpunkt bei Studienbeginn an. Die Änderungen werden auf einer 7-Punkte-Likert-Skala von eins (sehr viel besser) über vier (unverändert) bis sieben (sehr viel schlechter) dokumentiert. Bei einer chronisch progredienten Erkrankung wie der IPF ist ohne eine adäquate Therapie von einer stetigen Verschlechterung über die Zeit auszugehen. Das Ausbleiben einer solchen subjektiven Verschlechterung stellt einen Therapieerfolg und somit einen patientenrelevanten Nutzen dar.

PGI-C Responder

Die Bestimmung des *PGI-C* erfolgte zu Woche 12, 24 und 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 auf der 7-Punkte-Likert-Skala keine Kategorie der Verschlechterung seines Zustands angegeben hatte, also seinen Gesundheitszustand gegenüber Studienbeginn als „*sehr viel besser*“, „*viel besser*“, „*ein wenig besser*“ oder „*unverändert*“ bewertete. 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. Die entsprechenden Adjustierungsparameter finden sich in den Ergebnistabellen des Abschnitts 4.3.1 als Fußnoten. Die tabellarische Ergebnisdarstellung erfolgte in Analogie zur Darstellung für den Endpunkt Anteil *SGRQ-(≤ -4 Punkte)-Responder*.

EuroQol 5-Dimensional Quality of Life Questionnaire (EQ-5D)

Der *EQ-5D* ist ein unter anderem für respiratorische Erkrankungen validiertes Fragebogeninstrument zur generischen Erfassung der HRQoL zu den fünf Dimensionen *Mobilität, Fähigkeit sich selbst zu versorgen, alltägliche Aktivitäten, Schmerz/körperliche Beschwerden* und *Ängste/Niedergeschlagenheit* (Brazier et al., 1993, de Boer et al., 2004, Rabin and de Charro, 2001, Pickard et al., 2008). Durch eine weitere Domäne, die *EQ-5D-VAS*, bewertet der Patient zudem gesamthaft seinen Gesundheitsstatus auf einer 100-Punkte-Skala, wobei niedrigere Werte eine schlechtere HRQoL anzeigen. Im Kontext medizinischer Nutzenbewertungen wird der *EQ-5D* (beispielsweise durch das britische National Institute for Health and Care Excellence (NICE)) breit eingesetzt.

Veränderung der EQ-5D-VAS

In den berücksichtigten Studien wurde der *EQ-5D* bei Studienbeginn sowie zu Woche 12, 24 und 52 bestimmt. Die *Veränderung der EQ-5D VAS* wurde analog zum Vorgehen für die *Veränderung des SGRQ-Gesamtscores* berechnet und dargestellt.

Unerwünschte Ereignisse

Neben den erwünschten therapielevanten Wirkungen üben Arzneimittel auch unerwünschte Wirkungen auf den menschlichen Körper aus. Die Bedeutung unerwünschter Arzneimittelwirkungen ist abhängig von ihrer Schwere, der Häufigkeit ihres Auftretens sowie der Reversibilität ihrer Symptome. Insbesondere schwerwiegende und irreversible unerwünschte Arzneimittelwirkungen sind Gegenstand der Entscheidung über die Verkehrsfähigkeit von Arzneimitteln im Rahmen ihrer arzneimittelrechtlichen Zulassung (Bundesministerium der Justiz und für Verbraucherschutz, 2014).

Mit Blick auf die Bewertung des medizinischen Nutzens sind unerwünschte Arzneimittelwirkungen aus zwei Gründen patientenrelevant. Zum einen kann ihr Auftreten zu einer Veränderung der initial gewählten Therapie oder ihrem Abbruch führen, womit eine Verminderung der Effektivität einer Behandlung einhergehen kann. Aus gesellschaftlicher Sicht können unerwünschte Arzneimittelwirkungen bzw. deren Behandlung darüber hinaus mit einem erheblichen Ressourcenmehrverbrauch verbunden sein. Zum anderen können unerwünschte Arzneimittelwirkungen die Compliance vermindern und auf diese Weise mittelbar die Wahrscheinlichkeit eines Behandlungserfolgs reduzieren.

Die Erfassung von Nebenwirkungen als *unerwünschte Ereignisse* sollte bei einer Studie generell durchgeführt werden (European Medicines Agency (EMA), 1996). Die Verringerung von Nebenwirkungen zählt zu den Aspekten der therapiebedingten Morbidität und ist ein vom IQWiG anerkannter Endpunkt (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013).

In den berücksichtigten Studien wurden unerwünschte Ereignisse durch die folgenden Endpunkte erfasst:

- *UE*
- *SUE*
- *UE, die zum Therapieabbruch führten (UEA)*
- *UE, die zum Tod führten*
- *UE von besonderem Interesse*

Als *UE von besonderem Interesse* wurden *gastrointestinale Störungen (Diarrhoe, Übelkeit, Erbrechen, Appetitlosigkeit, Gewichtsverlust, gastrointestinales Perforation), Leberenzym- und Bilirubinerhöhungen, kardiale Ereignisse, respiratorische Ereignisse (Progression der IPF, Husten, Dyspnoe, Bronchitis, Nasopharyngitis), Blutungsereignisse* sowie alle weiteren Störungen, die in mindestens einem Studienarm mit einer Häufigkeit von $\geq 10\%$ der Studienteilnehmer auftraten, betrachtet (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f). Alle *UE* wurden in den berücksichtigten Studien routinemäßig bei jeder Studienvisite ab der Randomisierung bis 28 Tage nach der letzten Einnahme der Studienmedikation erfasst. Die Anzahl der Studienteilnehmer, die Anzahl und der Anteil der Studienteilnehmer mit einem *UE* der betreffenden Kategorie wurden tabellarisch dargestellt. Der Gruppenunterschied wurde als Risk Ratio (RR), OR und absolute Risikodifferenz (RD), jeweils mit 95 %-Konfidenzintervall und p-Wert, basierend auf einem Chi-Quadrat-Test, angegeben.

Auswertung der Endpunkte

In den berücksichtigten Studien wurden Analysen bezüglich der Mortalität, Morbidität, Lebensqualität und *UE* gemäß statistischem Analyseplan (SAP) 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. Dem gegenüber schloss das Randomized Set (RS) alle randomisierten Studienteilnehmer ein, unabhängig davon, ob mindestens eine dokumentierte Dosis der Studienmedikation verabreicht wurde. Die Anzahl der Studienteilnehmer im TS unterschied sich lediglich in 2 von insgesamt 515 randomisierten Studienteilnehmer der INPULSIS-1 Studie bzw. 3 von insgesamt 551 randomisierten Studienteilnehmer der INPULSIS-2 Studie, die keine Dosis der Studienmedikation erhalten hatten, von dem RS (entsprechend 0,39 % bzw. 0,54 % des RS). Dies ist als nicht ergebnisrelevante Verzerrung einzuschätzen. Es erfolgte somit eine adäquate Umsetzung des ITT-Prinzips.

Die Ergebnisse der Wirksamkeits-Endpunkte wurden für die Studien INPULSIS-1 und INPULSIS-2 sowie für die gepoolten INPULSIS Studien ausgewertet und dargestellt. Für die Bewertung der unerwünschten Ereignisse, die auf dem Zählen von Einzelereignissen in der Gesamtpopulation beruht, wurde auf eine Darstellung auf Einzelstudien-Ebene verzichtet und nur die gepoolte Patientenpopulation betrachtet. Vor dem Hintergrund des identischen Studiendesigns und ähnlicher Inzidenzen der erhobenen UE in den Einzelstudien hat die gepoolte Darstellung der UE eine höhere Trennschärfe und ist damit aussagekräftiger für die Grundgesamtheit. Eine Darstellung auf Einzelstudien-Ebene findet sich in Modul 5 (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c, Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d, Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f).

4.2.5.3 Meta-Analysen

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse quantitativ zusammengefasst werden, wenn die Studien aus medizinischen (z.B. Patientengruppen) und methodischen (z.B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Für Meta-Analysen soll die im Folgenden beschriebene Methodik eingesetzt werden.

Für die statistische Auswertung sollen primär die Ergebnisse aus Intention-to-treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Die Meta-Analysen sollen in der Regel auf Basis von Modellen mit zufälligen Effekten⁴ erfolgen. In begründeten Ausnahmefällen sollen zusätzlich Modelle mit festen Effekten eingesetzt werden. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, sollen diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt werden.

Für kontinuierliche Variablen soll die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' g, als Effektmaß eingesetzt werden. Bei binären Variablen sollen Meta-Analysen primär sowohl anhand des Odds Ratios als auch des Relativen Risikos durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet⁵ werden.

⁴ DerSimonian R, Laird N. *Meta-analysis in clinical trials*. *Control Clin Trials* 1986;7(3):177-188.

⁵ Deeks JJ, Higgins JPT, Altman DG. *Analysing data and undertaking meta-analyses*. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.

Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität⁶ erfolgen. Ist die Heterogenität der Studienergebnisse nicht bedeutsam, soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.2.5.5).

Beschreiben Sie die für Meta-Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Von den im Dossier zur Bewertung des Ausmaßes des Zusatznutzens zugrunde liegenden Studien, erfüllen alle Studien die Kriterien der medizinischen und methodischen Vergleichbarkeit, so dass diese gepoolt auf individuellen Patientendaten (IPD) in einer Meta-Analyse quantitativ zusammengefasst werden können. (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013).

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Ein Fixed-Effect Modell wurde gemäß IQWiG-Methoden in Kapitel 4.1 (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013) als ausreichend angesehen, da sowohl die klinische als auch statistische Homogenität Voraussetzung für die Durchführung der Meta-Analyse war:

„[...]Liegen Informationen darüber vor, dass die Effekte der einzelnen Studien homogen sind, ist eine Meta-Analyse unter der Annahme eines festen Effekts ausreichend. [...]“

Der pU verweist an dieser Stelle auf Abschnitt 4.3.1, wo sowohl die Ergebnisse der Einzelstudien als auch die Ergebnisse der IPD-Meta-Analysen, also die gepoolten Studienergebnisse, dargestellt und diskutiert werden.

⁶ Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560.

4.2.5.4 Sensitivitätsanalysen

Zur Einschätzung der Robustheit der Ergebnisse sollen Sensitivitätsanalysen hinsichtlich methodischer Faktoren durchgeführt werden. Die methodischen Faktoren bilden sich aus den im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ soll für Sensitivitätsanalysen verwendet werden.

Das Ergebnis der Sensitivitätsanalysen kann die Einschätzung der Aussagekraft der Nachweise beeinflussen.

Begründen Sie die durchgeführten Sensitivitätsanalysen oder den Verzicht auf Sensitivitätsanalysen. Beschreiben Sie die für Sensitivitätsanalysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

In den Studien INPULSIS-1 und INPULSIS-2 wurden zahlreiche Sensitivitätsanalysen durchgeführt. Eine Übersicht der Sensitivitätsanalysen, die sich auf die in diesem Dossier dargestellten Endpunkte beziehen, befindet sich in Modul 5 (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014e).

Die Sensitivitätsanalysen bestätigen die Robustheit der berichteten Ergebnisse.

4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse sollen hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht werden. Dies können beispielsweise direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen (z. B. die Dosis) sein. Im Gegensatz zu den in Abschnitt 4.2.5.4 beschriebenen methodischen Faktoren für Sensitivitätsanalysen besteht hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Eine potenzielle Effektmodifikation soll anhand von Homogenitäts- bzw. Interaktionstests oder von Interaktionstermen aus Regressionsanalysen (mit Angabe von entsprechenden Standardfehlern) untersucht werden. Subgruppenanalysen auf der Basis individueller Patientendaten haben in der Regel eine größere Ergebnissicherheit als solche auf Basis von Meta-Regressionen oder Meta-Analysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren, sie sind deshalb zu bevorzugen. Es sollen, soweit sinnvoll, folgende Faktoren bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte

Sollten sich aus den verfügbaren Informationen Anzeichen für weitere mögliche Effektmodifikatoren ergeben, können diese ebenfalls begründet einbezogen werden. Die Ergebnisse von in Studien a priori geplanten und im Studienprotokoll festgelegten Subgruppenanalysen für patientenrelevante Endpunkte sind immer darzustellen.

Bei Identifizierung möglicher Effektmodifikatoren kann gegebenenfalls eine Präzisierung der aus den für die Gesamtgruppe beobachteten Effekten abgeleiteten Aussagen erfolgen. Ergebnisse von Subgruppenanalysen können die Identifizierung von Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen unterstützen.

Benennen Sie die durchgeführten Subgruppenanalysen. Begründen Sie die durchgeführten Subgruppenanalysen bzw. die Untersuchung von Effektmodifikatoren oder den Verzicht auf solche Analysen. Beschreiben Sie die für diese Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Es wurden Interaktionstests für alle a priori geplanten Subgruppen (Tabelle 4-2) durchgeführt, um zu überprüfen, ob der Therapieeffekt über die Subgruppen hinweg konstant ist oder ob es Belege für bzw. Hinweise auf eine Effektmodifikation gibt. Die Details zur Berechnung der entsprechenden p-Werte zur Feststellung einer Interaktion sind in Abschnitt 4.3.1.3.5 als Fußnote in den jeweiligen Tabellen zu finden.

Ländereffekte wurden im Rahmen des Zulassungsverfahrens analysiert und mit der EMA diskutiert. Die entsprechenden Analyseergebnisse wurden in Abschnitt 4.3.1.3.5.5 dargestellt. Aufgrund der hieraus gewonnenen Erkenntnisse wurde auf die Durchführung weiterer dossierspezifischer Subgruppen-Analysen verzichtet.

Für die folgenden Subgruppen wurden Interaktionstests für alle im Dossier berichteten patientenrelevanten Endpunkte durchgeführt und in Abschnitt 4.3.1.3.5 dargestellt bzw. in Modul 5 hinterlegt.

Tabelle 4-2: Übersicht der a priori geplanten Subgruppen, zu denen Interaktionstests durchgeführt wurden.

Subgruppe	Merkmale	Endpunkte zur Wirksamkeit	Endpunkte zu UE
Geschlecht	männlich, weiblich	x	x
Alter	< 65 Jahre, ≥ 65 Jahre	x	--
	< 65 Jahre, 65-74 Jahre, ≥ 75 Jahre	--	x
Ethnie	kaukasisch, asiatisch	x	x
FVC (% pred.) ^a	≤ 70 %, > 70 %	x	--
SGRQ	≤ 40 Punkte, > 40 Punkte	x	--
Raucherstatus	Nichtraucher, Ex-Raucher/Raucher	x	x
systemische Kortikosteroide	ja, nein	x	--
Bronchodilator	ja, nein	x	--
Niereninsuffizienz	ja, nein	--	x
Gewicht	< 65 kg, ≥ 65 kg	--	x
Studie	1199.32, 1199.34	x	--
a: FVC % als Maß für die Krankheitsschwere			

Wurde in einer Subgruppe durch einen Interaktionstest ein Hinweis auf ($p < 0,2$) oder ein Beleg ($p < 0,05$) für eine Modifikation des Behandlungseffekts durch das Subgruppenmerkmal ermittelt, so wurden dieser Interaktions-p-Wert sowie eine detaillierte Ergebnisdarstellung der Subgruppen-Ergebnisse auf Endpunktebene in Abschnitt 4.3.1.3.5 aufgeführt und diese Ergebnisse zwischen den Subgruppen verglichen. Die Beschreibung der Subgruppenergebnisse erfolgte dabei analog der Ergebnisbeschreibung für die Gesamtpopulation (Abschnitt 4.3.1.3.5).

4.2.5.6 Indirekte Vergleiche

Zurzeit sind international Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Nicht adjustierte indirekte Vergleiche (d. h. Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator) stellen dabei keine valide Analysemethode dar, der Einsatz einfacher adjustierter indirekter Vergleiche ist möglich⁷. Komplexe Verfahren für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche werden in der Literatur unterschiedlich bezeichnet, z. B. als „Mixed-Treatment-Comparison (MTC)-Meta-Analysen“⁸, „Multiple-Treatment-Meta-Analysen“⁹ oder auch „Netzwerk-Meta-Analysen“¹⁰, sie gehen aber im Prinzip von denselben wesentlichen Annahmen aus.

Grundannahme für solche komplexen Analysen ist die Annahme der Konsistenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nicht mehr nur durch Zufallsfehler oder Heterogenität erklärbar ist¹¹.

Da das Ergebnis eines indirekten Vergleichs maßgeblich von der Auswahl des Brückenkomparators bzw. der Brückenkomparatoren abhängen kann, ist die Wahl des Brückenkomparators bzw. der Brückenkomparatoren zu begründen. Dies gilt insbesondere dann, wenn eine Beschränkung auf ein oder mehrere Brückenkomparatoren vorgenommen wird, obwohl Daten zu anderen Therapieoptionen, die ebenfalls als Brückenkomparatoren in Frage kommen, vorliegen. Insgesamt ist es notwendig, die zugrunde liegende Methodik genau und reproduzierbar zu beschreiben und die Annahme der Konsistenz zu untersuchen¹².

⁷ Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005; 9(26): 1-148.

⁸ Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.

⁹ Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331(7521): 897-900.

¹⁰ Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17(3): 279-301.

¹¹ B. Schöttker, D. Lühmann, D. Boukhemair, and H. Raspe. Indirekte Vergleiche von Therapieverfahren. *Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.*

¹² Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DJ. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.

Beschreiben Sie detailliert und vollständig die zugrunde liegende Methodik des indirekten Vergleichs. Dabei sind mindestens folgende Angaben notwendig:

- *Benennung des Brückenkomparators bzw. der Brückenkomparatoren und Begründung für die Auswahl.*
- *Genaue Spezifikation des statistischen Modells inklusive aller Modellannahmen. Bei Verwendung eines Bayesianischen Modells sind dabei auch die angenommenen A-priori-Verteilungen (falls informative Verteilungen verwendet werden, mit Begründung), die Anzahl der Markov-Ketten und deren Startwerte und Länge zu spezifizieren.*
- *Art der Prüfung der Homogenität der Ergebnisse direkter paarweiser Vergleiche.*
- *Art der Prüfung der Konsistenz zwischen den Ergebnissen direkter und indirekter Vergleiche.*
- *Bilden Sie den Code des Computerprogramms in lesbarer Form ab und geben Sie an, welche Software Sie zur Berechnung eingesetzt haben (ggf. inklusive Spezifizierung von Modulen, Prozeduren, Packages etc.; siehe auch Modul 5 zur Ablage des Programmcodes).*
- *Art und Umfang von Sensitivitätsanalysen.*

Aufgrund der vorliegenden Placebo-kontrollierten, vergleichenden Studien (INPULSIS-1, INPULSIS-2) wurde die Bewertung des Ausmaßes des medizinischen Zusatznutzens auf Grundlage der Evidenzgüte Stufe Ib durchgeführt und daher auf indirekte Vergleiche verzichtet.

4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

Zusammenfassung des Abschnitts 4.3

Ergebnisse zum Ausmaß des medizinischen Zusatznutzens

Der medizinische Nutzen und medizinische Zusatznutzen von Nintedanib (Ofev®) in der Indikation IPF wurde mit der Zulassung als Orphan Drug durch die Europäische Kommission bestätigt. Nachfolgend wird das Ausmaß des medizinischen Zusatznutzens von Nintedanib bei erwachsenen Patienten mit IPF beschrieben.

Mortalität

Die *Zeit bis zum Tod, unabhängig von der Ursache* wird als Überlebenswahrscheinlichkeit anhand des Kaplan-Meier-Schätzers berechnet. Basierend auf einer Cox-Regression, war die Überlebenswahrscheinlichkeit zum Studienende in den gepoolten INPULSIS Studien unter Nintedanib um 30 % höher als unter Placebo (HR [95 % KI]: 0,70 [0,43; 1,12]; p = 0,1399). Ein vergleichbarer Vorteil wurde auch für die *Zeit bis respiratorisch bedingtem Tod* beobachtet (HR [95 % KI]: 0,74 [0,41; 1,34]; p = 0,3435). Die Kaplan-Meier-Kurven zeigen, dass mit zunehmender Beobachtungszeit der Unterschied der verstorbenen Studienteilnehmer zwischen den Behandlungsgruppen Nintedanib und Placebo größer wird.

Die *jährliche Abnahme der forcierten Vitalkapazität (FVC)* (ml) als Surrogat für die Gesamtmortalität war in den gepoolten INPULSIS Studien unter Nintedanib mit 113,59 ml klinisch relevant und statistisch signifikant geringer als unter Placebo (223,53 ml). Der Unterschied zwischen den beiden Behandlungsgruppen hat eine Mittelwertdifferenz von 109,94 ml (MWD [95 % KI]: 109,94 ml [75,85; 144,03]; p < 0,0001). Die Abnahme der FVC im Behandlungsarm mit Nintedanib fiel somit durchschnittlich um 50 % geringer aus als im Placeboarm. Der Anteil der *FVC-Responder* war in den Phase III-Studien definiert als Anteil der Studienteilnehmer mit einem absoluten Abfall der FVC nicht größer als 10 % bzw. nicht größer als 5 % von Baseline bis Woche 52. Die Analyse der *FVC-Responder* zum Schwellenwert von 10 % zeigt durch eine logistische Regression mit einem OR von 1,577 einen statistisch signifikanten Unterschied zwischen den Behandlungsgruppen (OR [95 % KI]: 1,577 [1,21; 2,05]; p = 0,0007). Bei einem Schwellenwert von 5 % zeigt sich ebenfalls ein statistisch signifikanter Unterschied: Auf Nintedanib sprachen gemäß dieser Analyse mehr Studienteilnehmer an als auf Placebo (OR [95 % KI]: 1,835 [1,43; 2,36]; p < 0,0001).

Morbidität

Im Dossier wurde die Morbidität durch den Endpunkt *Zeit bis zur ersten adjudizierten akuten Exazerbation* operationalisiert.

Adjudizierte akute Exazerbationen traten in den gepoolten INPULSIS Studien unter Nintedanib seltener auf als unter Placebo (1,9 % vs. 5,7 %). Die Wahrscheinlichkeit, keine Exazerbation zu erleiden (ausgedrückt durch den Kaplan-Meier-Schätzer), beträgt in der Nintedanib-Gruppe 0,980 und in der Placebo-Gruppe 0,940.

Auch das HR [95 % KI] von 0,32 für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* zeigt einen Vorteil zugunsten von Nintedanib (HR [95 % KI: 0,32 [0,16; 0,65]; $p = 0,0010$). Dies entspricht einer statistisch signifikanten Reduktion des Risikos einer ersten adjudizierten akuten Exazerbation um 68 %.

Lebensqualität

Die krankheitsspezifische Lebensqualität wurde in diesem Dossier durch die Endpunkte *St. George's Respiratory Questionnaire (SGRQ)*-, *St. George's Respiratory Questionnaire IPF-specific version (SGRQ-I)*-Veränderung und *-Responder* sowie *Veränderung des Shortness of Breath Questionnaire (SOBQ)* und der *Veränderung des Cough and Sputum Assessment Questionnaire (CASA-Q)* bewertet. Die allgemeine Lebensqualität wurde über die Endpunkte *Patient's Global Impression of Change (PGI-C)*-Responder und *Veränderung der EuroQol 5-Dimensional Quality of Life Questionnaire (EQ-5D)*-Visuellen Analogskala (VAS) dargestellt. Die Ergebnisse aller Endpunkte zeigen einen Vorteil in Form einer geringeren Verschlechterung der Lebensqualität innerhalb des Studienzeitraums unter Nintedanib im Vergleich zu Placebo. Für die *Veränderung des SGRQ-I* und die *PGI-C-Responder* war der Unterschied statistisch signifikant (*SGRQ-I-Gesamtscore*: MWD [95 % KI]: -1,99 Punkte [-3,71; -0,27]; $p = 0,0230$; *PGI-C-Responder*: OR [95 % KI] 1,328 [1,04; 1,70]; $p = 0,0254$).

Unerwünschte Ereignisse

Im Rahmen dieser Nutzenbewertung wurden *unerwünschte Ereignisse (UE)*, *schwerwiegende unerwünschte Ereignisse (SUE)*, *unerwünschte Ereignisse, die zum Therapieabbruch führten (UEA)*, und *unerwünschte Ereignisse, die zum Tod führten*, berücksichtigt. Zusätzlich wurden unerwünschte Ereignisse von besonderem Interesse sowie unerwünschte Ereignisse betrachtet, die in > 10 % der Patienten eines Behandlungsarms auftraten. Für *schwerwiegende unerwünschte Ereignisse* sowie *unerwünschte Ereignisse, die zum Tod führten*, liegt weder ein Zusatznutzen noch ein Nachteil von Nintedanib vor. Dahingegen ergibt sich für *unerwünschte Ereignisse* und *unerwünschte Ereignisse, die zum Therapieabbruch führten*, ein Hinweis auf einen Nachteil von Nintedanib gegenüber Placebo.

Ein Hinweis auf einen Zusatznutzen gegenüber Placebo ergibt sich für die *unerwünschten Ereignisse von besonderem Interesse Progression der IPF* und *Dyspnoe*. Für die unerwünschten Ereignisse von besonderem Interesse *kardiale Ereignisse, schwerwiegende kardiale Ereignisse, tödliche kardiale Ereignisse, Husten, Bronchitis* und *Nasopharyngitis* liegen weder ein Zusatznutzen noch ein Nachteil von Nintedanib gegenüber Placebo vor. Ein Nachteil von Nintedanib gegenüber Placebo ergibt sich für *gastrointestinale Ereignisse, arterielle thromboembolische Ereignisse* sowie *Leberenzym- und Bilirubinveränderungen*. Die Fachinformation für das Produkt Ofev[®] gibt Empfehlungen zu Maßnahmen hinsichtlich des Nebenwirkungsmanagements (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Fazit

Bei Ofev[®] handelt es sich um ein Arzneimittel zur Behandlung eines seltenen Leidens, der medizinische Zusatznutzen gilt gemäß § 35a Abs. 1 Satz 10 SGB V bereits durch die Zulassung als belegt. Lediglich das Ausmaß des Zusatznutzens muss für die Anzahl der Patienten, für die dieser Zusatznutzen besteht, auf Grundlage der Zulassungsstudien nachgewiesen werden.

Der therapeutische Bedarf in der Behandlung der IPF ist hoch. Ofev[®] (Nintedanib) wurde am 15.01.2015 im Rahmen eines „accelerated assessments“ als erstes Arzneimittel zur Behandlung aller Krankheitsschweregrade der IPF durch die Europäische Kommission zugelassen. Die US-amerikanische Gesundheitsbehörde FDA hatte Nintedanib bereits den Status eines Therapiedurchbruchs (Breakthrough Therapy Designation) erteilt und ebenfalls in einem beschleunigten Verfahren am 15. Oktober 2014 in den USA zugelassen. Gegenüber Placebo weist Nintedanib einen beträchtlichen Zusatznutzen in den Endpunktkategorien Mortalität (FVC) und Morbidität (Reduktion adjudizierter akuter Exazerbationen) auf. In der Endpunktkategorie Lebensqualität besteht ein nicht quantifizierbarer Zusatznutzen. Bezüglich der unerwünschten Ereignisse ist ein Nachteil anzunehmen.

Aus den dargestellten Daten geht hervor, dass bei der Behandlung mit Nintedanib für erwachsene Patienten mit IPF der Nutzen den potenziellen Nachteil überwiegt.

In den nachfolgenden Abschnitten sind die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen zu beschreiben. Abschnitt 4.3.1 enthält dabei die Ergebnisse aus randomisierten kontrollierten Studien, die mit dem zu bewertenden Arzneimittel durchgeführt wurden (Evidenzstufen Ia/Ib).

Abschnitt 4.3.2 enthält weitere Unterlagen anderer Evidenzstufen, sofern diese aus Sicht des pharmazeutischen Unternehmers zum Nachweis des Zusatznutzens erforderlich sind. Diese Unterlagen teilen sich wie folgt auf:

- Randomisierte, kontrollierte Studien für einen indirekten Vergleich mit der zweckmäßigen Vergleichstherapie, sofern keine direkten Vergleichsstudien mit der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen (Abschnitt 4.3.2.1)
- Nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2)
- Weitere Untersuchungen (Abschnitt 4.3.2.3)

4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel

4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel

4.3.1.1.1 Studien des pharmazeutischen Unternehmers

Nachfolgend sollen alle Studien (RCT), die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie alle Studien (RCT), für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, benannt werden. Beachten Sie dabei folgende Konkretisierungen:

- Es sollen alle RCT, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden und deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle RCT, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.
- Benennen Sie in der nachfolgenden Tabelle nur solche RCT, die ganz oder teilweise innerhalb des in diesem Dokument beschriebenen Anwendungsgebiets durchgeführt wurden. Fügen Sie dabei für jede Studie eine neue Zeile ein.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Angabe „Zulassungsstudie ja/nein“, Angabe über die Beteiligung (Sponsor ja/nein), Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-3: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer	Therapiearme
TOMORROW (1199.30)	nein	ja	abgeschlossen	52 Wochen	Nintedanib 50 mg (1x täglich) 50 mg 100 mg 150 mg vs. Placebo je 2 x täglich, per os
1199.31	nein	ja	abgeschlossen	14 Tage bzw. 28 Tage	Nintedanib 50 mg 100 mg 150 mg vs. Placebo je 2 x täglich, per os

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer	Therapiearme
INPULSIS-1 (1199.32)	ja	ja	abgeschlossen	52 Wochen	Nintedanib 150 mg vs. Placebo je 2 x täglich, per os
1199.33	nein	ja	laufend	-	Nintedanib 100 mg 150 mg je 2 x täglich, per os
INPULSIS-2 (1199.34)	Ja	ja	abgeschlossen	52 Wochen	Nintedanib 150 mg vs. Placebo je 2 x täglich, per os
1199.35	nein	ja	laufend	-	Nintedanib 50 mg (1 x täglich) 50 mg 100 mg 150 mg je 2 x täglich, per os; Studienteilnehmern wurde nach 3 Monaten angeboten auf die Dosierung 150 mg 2 x täglich zu eskalieren
1199.40	nein	ja	laufend	-	Nintedanib 150 mg 2 x täglich, per os (begleitend zu Pirfenidon)
1199.187	nein	ja	laufend	52–78 Wochen	Nintedanib 150 mg vs. Placebo je 2 x täglich, per os

Geben Sie an, welchen Stand die Information in Tabelle 4-3 hat, d. h. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Der Informationsstand bezieht sich auf den 15.01.2015.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-3 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-4: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
1199.30 (TOMORROW)	keine Zulassungsstudie; Intervention abweichend von Vorgaben der Fachinformation aufgrund fehlender Option zur Reeskalation auf Startdosis nach zwischenzeitlicher Behandlungsunterbrechung oder Dosisreduktion
1199.31	keine Zulassungsstudie
1199.33	open-label Extension der 1199.32/1199.34; keine Zulassungsstudie
1199.35	open-label Extension der 1199.30; keine Zulassungsstudie
1199.40	open-label Extension der 1199.31; keine Zulassungsstudie
1199.187	keine Zulassungsstudie

4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

Beschreiben Sie nachfolgend das Ergebnis der bibliografischen Literaturrecherche. Illustrieren Sie den Selektionsprozess und das Ergebnis der Selektion mit einem Flussdiagramm. Geben Sie dabei an, wie viele Treffer sich insgesamt (d. h. über alle durchsuchten Datenbanken) aus der bibliografischen Literaturrecherche ergeben haben, wie viele Treffer sich nach Entfernung von Dubletten ergeben haben, wie viele Treffer nach Sichtung von Titel und, sofern vorhanden, Abstract als nicht relevant angesehen wurden, wie viele Treffer im Volltext gesichtet wurden, wie viele der im Volltext gesichteten Treffer nicht relevant waren (mit Angabe der Ausschlussgründe) und wie viele relevante Treffer verblieben. Geben Sie zu den relevanten Treffern an, wie vielen Einzelstudien diese zuzuordnen sind. Listen Sie die im Volltext gesichteten und ausgeschlossenen Dokumente unter Nennung des Ausschlussgrunds in Anhang 4-C.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Geben Sie im Flussdiagramm auch das Datum der Recherche an. Die Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Orientieren Sie sich bei der Erstellung des Flussdiagramms an dem nachfolgenden Beispiel.

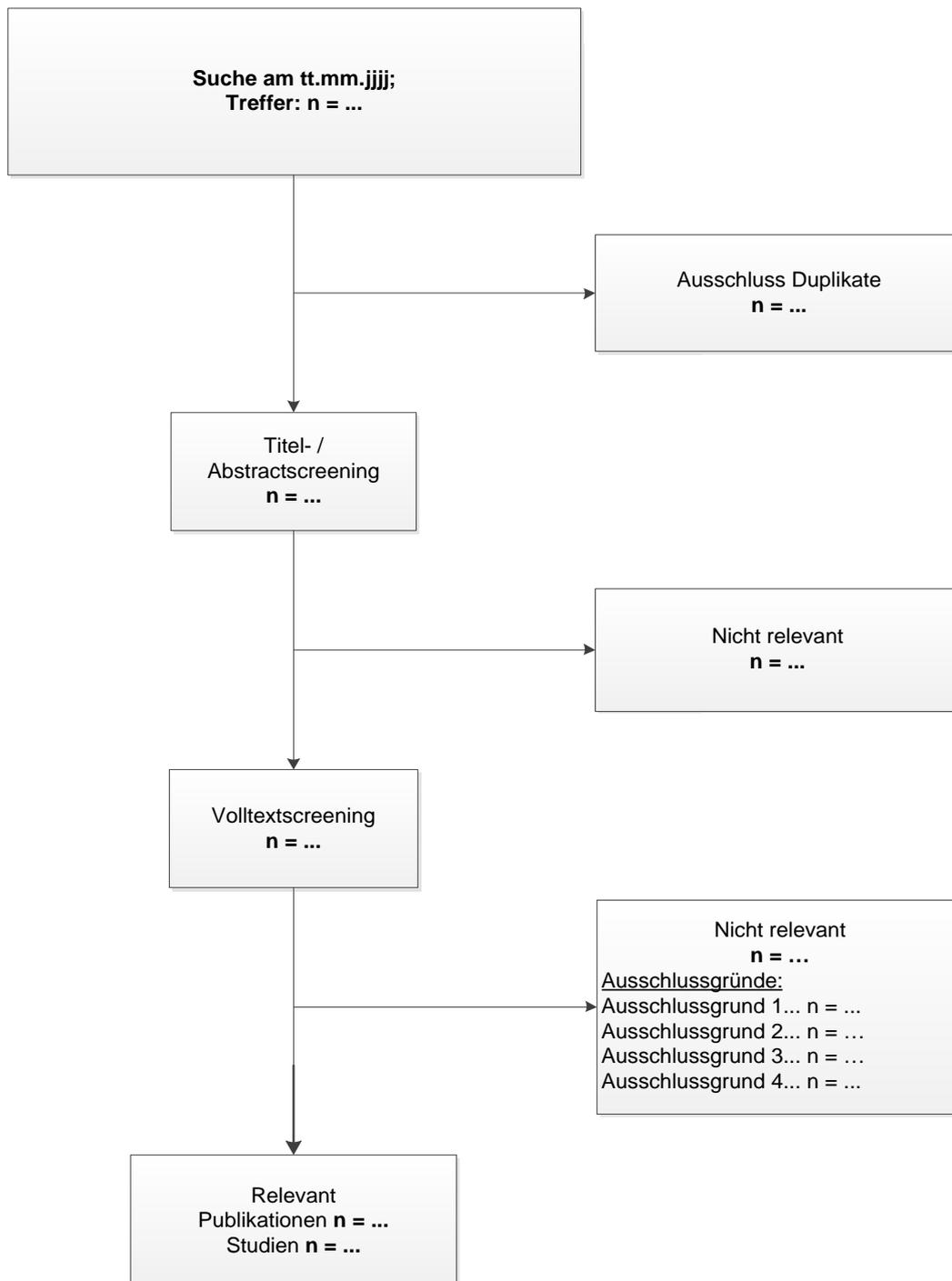


Abbildung 4-5: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status¹ in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden (European Commission (EC)). Somit gilt der Zusatznutzen von Nintedanib als belegt. Die Bewertung des Ausmaßes des Zusatznutzens basiert auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. Eine bibliographische Literaturrecherche ist daher nicht erforderlich.

4.3.1.1.3 Studien aus der Suche in Studienregistern

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern identifiziert wurden. Geben Sie dabei an, in welchem Studienregister die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (z. B. Studienregistereintrag, Bericht über Studienergebnisse etc.). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-3) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein. Listen Sie die ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-D.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei Ihren Angaben an der beispielhaften ersten Tabellenzeile.

Tabelle 4-5: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel

Studie	Identifikationsorte (Name des Studienregisters und Angabe der Zitate ^a)	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja / nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja / nein)	Status (abgeschlossen / abgebrochen / laufend)
<Studie 1>	clinicaltrials.gov [6, 7]	ja	nein	abgeschlossen
n. z.				
a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder –ergebnisse.				

Geben Sie an, welchen Stand die Information in Tabelle 4-5 hat, d. h. zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status´ in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden) (European Commission (EC)). Somit gilt der Zusatznutzen von Nintedanib als belegt, die Bewertung des Ausmaßes des Zusatznutzens basiert auf der Grundlage der Zulassung und der die Zulassung begründenden Studien. Eine Suche in Studienregistern ist daher nicht erforderlich.

4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel

Benennen Sie in der nachfolgenden Tabelle den aus den verschiedenen Suchschritten (Abschnitte 4.3.1.1.1, 4.3.1.1.2 und 4.3.1.1.3) resultierenden Pool relevanter Studien (exklusive laufender Studien) für das zu bewertende Arzneimittel, auch im direkten Vergleich zur zweckmäßigen Vergleichstherapie. Führen Sie außerdem alle relevanten Studien einschließlich der verfügbaren Quellen in Abschnitt 4.6 auf. Alle durch die vorhergehenden Schritte identifizierten und in der Tabelle genannten Quellen der relevanten Studien sollen für die Bewertung dieser Studien herangezogen werden.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Quellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften ggf. sinnvoll angeordnet werden, beispielsweise nach Therapieschema (Akut-/Langzeitstudien) und jeweils separat nach Art der Kontrolle (Placebo, zweckmäßige Vergleichstherapie, beides). Sollten Sie eine Strukturierung des Studienpools vornehmen, berücksichtigen Sie diese auch in den weiteren Tabellen in Modul 4.

Tabelle 4-6: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Quellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
placebokontrolliert						
INPULSIS-1 (1199.32)	ja	ja	nein	ja [(Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c)]	ja [(Boehringer Ingelheim Pharma GmbH & Co. KG, 2015c, Richeldi et al., 2014a, Richeldi et al., 2014b)]	ja (Richeldi et al., 2014a, Richeldi et al., 2014b)
INPULSIS-2 (1199.34)	ja	ja	nein	ja [(Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d)]	ja [(Boehringer Ingelheim Pharma GmbH & Co. KG, 2015d, Richeldi et al., 2014a, Richeldi et al., 2014b)]	ja (Richeldi et al., 2014a, Richeldi et al., 2014b)
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						
n. z.						
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregister-einträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.7 genannte Referenzliste. Darüber hinaus ist darauf zu achten, dass alle Quellen, auf die in dieser Tabelle verwiesen wird, auch in Abschnitt 4.6 (Liste der eingeschlossenen Studien) aufgeführt werden.</p> <p>b: Studie, für die der Unternehmer Sponsor war.</p> <p>c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder –ergebnisse.</p>						

4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

4.3.1.2.1 Studiendesign und Studienpopulationen

Beschreiben Sie das Studiendesign und die Studienpopulation der in die Bewertung eingeschlossenen Studien mindestens mit den Informationen in den folgenden Tabellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Fügen Sie für jede Studie eine neue Zeile ein.

Weitere Informationen zu Studiendesign, Studienmethodik und Studienverlauf sind in Anhang 4-E zu hinterlegen

Tabelle 4-7: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	Studiendesign	Population	Interventionen (Zahl der randomisierten Patienten)	Studiendauer <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
INPULSIS-1 (1199.32)	zweiarmig, randomisiert (3:2), doppelblind, Placebo-kontrolliert, multizentrisch, Phase III	Erwachsene (≥ 40 Jahre) mit IPF-Diagnose nach ATS/ERS-Richtlinien DL _{CO} 30-79 % FVC ≥ 50 %	<u>Nintedanib</u> 150 mg 2x täglich per os n = 309 (bei Auftreten bestimmter UE ggf. Dosisreduktion auf 100 mg möglich) <u>Placebo</u> 2x täglich, per os n = 206	52 Wochen <i>Follow up</i> : 28 Tage nach Beendigung der Therapie	98 Studienzentren in 13 Ländern (Australien, Belgien, China, Tschechische Republik, Frankreich, Deutschland, Indien, Irland, Israel, Italien, Japan, UK, USA) 5/2011 – 10/2013	<u>Primärer Endpunkt</u> : - Jährliche FVC-Abnahme (ml) <u>Sekundäre Endpunkte</u> : - Zeit bis zum Tod - Zeit bis respiratorisch bedingtem Tod - Zeit bis Tod unter Behandlung ^a - Zeit bis zum Tod oder Lungentransplantation ^b - Zeit bis zur ersten akuten Exazerbation (adjudiziert) - Inzidenzrate einer akuten Exazerbation - absolute und relative Veränderung der FVC (ml) und FVC (%) - kategoriale Veränderung der FVC (%) (Abfall > 5 %; Anstieg > 5 %; Änderung ≤ 5 %) - kategoriale Änderung der FVC (%) (Abfall > 10 %; Anstieg > 10 %; Änderung ≤ 10 %) - Anteil FVC-Responder (Schwellenwert 5 % bzw. 10 % vom Sollwert) - Veränderung SGRQ

Studie	Studiendesign	Population	Interventionen (Zahl der randomisierten Patienten)	Studiendauer <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
	<RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	<relevante Charakteristika, z. B. Schweregrad>				Gesamtscore - Veränderung SGRQ-Domänen von Baseline über 52 Wochen - Anteil SGRQ-Responder (≤ -4) - Veränderung SGRQ-I- Gesamtscore - Anteil SGRQ-I-Responder (≤ -4) - Veränderung SOBQ- Gesamtscore - Veränderung CASA-Q - Anteil PGI-C-Responder - Veränderung EQ-5D - Veränderung SpO ₂ - Veränderung DL _{CO}
INPULSIS-2 (1199.34)	zweiarmig, randomisiert (3:2), doppelblind, Placebo- kontrolliert, multizentrisch; Phase III	Erwachsene (≥ 40 Jahre) mit einer IPF- Diagnose nach ATS/ERS- Richtlinien DL _{CO} 30-79 % FVC ≥ 50 %	<u>Nintedanib</u> 150 mg 2x täglich per os n = 331 (bei Auftreten bestimmter UE ggf. Dosisreduktion auf 100 mg möglich) <u>Placebo</u> 2x täglich; per os n = 220	52 Wochen <i>Follow-up</i> : 28 Tage nach Beendigung der Therapie	107 Studienzentren in 17 Ländern (Kanada, Chile, China, Finnland, Frankreich, Deutschland, Griechenland, Indien, Japan, Korea, Mexiko, Niederlande, Portugal, Russland, Spanien, Türkei, USA) 05/2011 – 10/2013	homogenes Studiendesign wie INPULSIS-1

Studie	Studiendesign	Population	Interventionen (Zahl der randomisierten Patienten)	Studiendauer	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
	<RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	<relevante Charakteristika, z. B. Schweregrad>		<ggf. Run-in, Behandlung, Nachbeobachtung>		
a: Als UE, die zum Tod führten, berücksichtigt						
b: Zusammengesetzter Endpunkt, bei dem zwei verschiedene Ergebnissituationen für den Patienten vermischt werden						

Tabelle 4-8: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib	Placebo	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i>
INPULSIS-1 (1199.32)	Nintedanib 150 mg 2x täglich per os	Placebo 2x täglich per os	<p>Studienteilnehmer erhielten Nintedanib als Soft-Gelatine-Kapsel kontinuierlich über 52 Wochen in einer Dosis von 150 mg zweimal täglich. Placebo wurde ebenfalls als Soft-Gelatine Kapsel zweimal täglich verordnet. Die Einnahme sollte mit einem Glas Wasser (ca. 250 ml) in einem Intervall von 12 Stunden, wenn möglich täglich zur gleichen Zeit, erfolgen (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> <p>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>
INPULSIS-2 (1199.34)	Nintedanib 150 mg 2x täglich per os	Placebo 2x täglich per os	homogenes Studiendesign wie INPULSIS-1

Tabelle 4-9: Charakterisierung der Studienpopulation (Baseline) – RCT mit dem zu bewertenden Arzneimittel

Studie	INPULSIS-1		INPULSIS-2		INPULSIS (gepoolt)	
	Nintedanib	Placebo	Nintedanib	Placebo	Nintedanib	Placebo
N ^a	309	204	329	219	638	423
Demographie						
Alter (Jahre) MW (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)
Geschlecht (männlich) N (%)	251 (81,2)	163 (79,9)	256 (77,8)	171 (78,1)	507 (79,5)	334 (79,0)
Gewicht (kg) 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)
BMI (kg/m ²) 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)
Ethnie N(%) ^b - kaukasisch - asiatisch	198 (64,1) 66 (21,4)	135 (66,2) 41 (20,1)	162 (49,2) 128 (38,9)	113 (51,6) 86 (39,3)	360 (56,4) 194 (30,4)	248 (58,6) 128 (30,3)
Raucherstatus N(%) - Nicht-Raucher - Ex-Raucher - Raucher	71 (23,0) 217 (70,2) 21 (6,8)	51 (25,0) 144 (70,6) 9 (4,4)	103 (31,1) 218 (66,3) 8 (2,4)	71 (32,4) 139 (63,5) 9 (4,1)	174 (27,3) 435 (68,2) 29 (4,5)	122 (28,8) 283 (66,9) 18 (4,3)
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) ^a	1,57 (1,31)
krankheitsspezifische Charakteristika						
Radiologische Beurteilung der IPF N(%) - UIP - vermutlich UIP - mögliche UIP - keine UIP	301 (97,4) nicht erhoben 8 (2,6) 0	198 (97,1) nicht erhoben 6 (2,9) 0	321 (97,6) nicht erhoben 8 (2,4) 0	215 (98,2) nicht erhoben 4 (1,8) 0	622 (97,5) nicht erhoben 16 (2,5) 0	413 (97,6) nicht erhoben 10 (2,4) 0
Lungenbiopsie N (%)	60 (19,4)	33 (16,2)	84 (25,5)	52 (23,7)	144 (22,6) ^c	85 (20,1)
FVC MW (SD) - in ml - in % vom Sollwert	2.756,8 (735,1) 79,47 (17,03)	2.844,5 (820,1) 80,53 (17,34)	2.672,8 (776,0) 79,99 (18,08)	2.619,0 (787,3) 78,09 (18,97)	2.713,5 (757,0) 79,74 (17,57)	2.727,7 (810,2) 79,27 (18,22)

Studie	INPULSIS-1		INPULSIS-2		INPULSIS (gepoolt)	
	Nintedanib	Placebo	Nintedanib	Placebo	Nintedanib	Placebo
N^a	309	204	329	219	638	423
FEV₁: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)
DLco (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)
SpO₂ (%) 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)
SGRQ-Gesamtscore (Punkte)	39,55 (17,63)	39,79 (18,48)	39,46 (20,47)	39,39 (18,65)	39,51 (19,15)	39,58 (18,54)
Systemische Kortikosteroid- Therapie N (%)	68 (22,0)	43 (21,1)	68 (20,7)	46 (21,0)	136 (21,3)	89 (21,0)
<p>a: Anzahl Studienteilnehmer im Treated Set b: Das Merkmal Ethnie wurde in Frankreich aufgrund entsprechender gesetzlicher Vorschriften nicht erfasst. c: Zahlen wurden aus Daten der u. g. Quellen berechnet. Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.1.4.1, Tab. 15.1.4.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.1.4.1, Tab. 15.1.4.2, (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 1.1.6, Tab. 1.1.7, Tab. 1.1.8, Tab. 1.1.10, Tab. 1.1.11, Tab. 1.1.12</p>						

Beschreiben Sie die Studien zusammenfassend. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Sollte es Unterschiede zwischen den Studien geben, weisen Sie in einem erläuternden Text darauf hin.

Die Studien INPULSIS-1 und INPULSIS-2 waren zweiarmige, doppelblinde, Placebo-kontrollierte, multizentrische, multinationale, randomisierte (3:2) Phase III-Studien mit homogenem Studiendesign (Tabelle 4-7). Studienteilnehmer waren Erwachsene mit einer nach ATS/ERS-Richtlinien gesicherten IPF-Diagnose. Das Mindestalter der Studienteilnehmer betrug 40 Jahre. Weitere Einschlusskriterien waren ein DL_{CO} von 30-79 % vom Sollwert, eine FVC von ≥ 50 % vom Sollwert bei Studienbeginn sowie das Vorliegen einer HRCT innerhalb von zwölf Monaten vor Zeitpunkt der Randomisierung.

In den Studien INPULSIS-1 und INPULSIS-2 betrug die Behandlungsphase 52 Wochen, in der die Studienteilnehmer im Verum-Arm Nintedanib in einer kontinuierlichen Dosierung von 150 mg zweimal täglich per os erhielten. Die Einnahme sollte dabei im Intervall von 12 Stunden und nach einer Mahlzeit erfolgen. Beim Auftreten von UE konnte im Einzelfall in Rücksprache mit dem Studienteilnehmer eine Dosisreduktion auf 100 mg oder eine Therapie-Unterbrechung erwogen werden. Eine Aufdosierung auf 150 mg war bei Besserung der Beschwerden aufgrund von UE innerhalb von vier darauf folgenden Wochen zulässig. Bei schwerwiegender Toxizität oder bei Intoleranz trotz Dosisreduktion wurde die Therapie hingegen abgebrochen und nicht fortgesetzt. Im Kontrollarm erhielten die Studienteilnehmer ein äußerlich nicht unterscheidbares Placebopräparat im gleichen Gabe- und Zeitintervall. An die Behandlungsphase schloss sich eine 28-tägige Nachbeobachtungsphase an. Studienteilnehmer hatten des Weiteren die Möglichkeit an einer unverblindeten Verlängerungsstudie zur Untersuchung der Langzeit-Sicherheit (1199.33) teilzunehmen.

Bei den in die Bewertung eingehenden Studien INPULSIS-1 und INPULSIS-2 war als primärer Endpunkt die *jährliche Abnahme der FVC* in ml nach 52 Wochen definiert. Als sekundäre Endpunkte wurden die *Veränderung des SGRQ-Gesamtscore* sowie das Auftreten von akuten Exazerbationen zu Woche 52 bestimmt. Weitere sekundäre Endpunkte waren Analysen bezüglich der FVC (absolute und relative Veränderungen sowie Responder-Analysen mit Cut-Off-Werten bei 5 %- bzw. 10 %-iger Veränderung (bezogen auf den Sollwert) von Baseline bis Woche 52 sowie Analysen zur gesundheitsbezogenen Lebensqualität mittels verschiedener Fragebögen (*Veränderungen im SGRQ, SGRQ-I, SOBQ, CASA-Q* und der *EQ-5D-VAS* sowie *SGRQ- und SGRQ-I-* (Cut-Off-Wert: -4 Punkte) und *PGI-C-Responder*. Zusätzlich wurden Überlebenszeitanalysen (*Zeit bis zum Tod, unabhängig von der Ursache, Zeit bis zum respiratorisch bedingtem Tod, Zeit bis Tod unter Behandlung, Zeit bis Tod oder Lungentransplantation*) sowie Bestimmungen der Veränderung von Baseline bis Woche 52 für die Parameter SpO₂ und DL_{CO} durchgeführt.

Beide Studien fanden in einem Zeitraum von Mai 2011 bis Oktober 2013 statt. Die Studie INPULSIS-1 wurde in 98 Studienzentren in 13 verschiedenen Ländern und die Studie INPULSIS-2 in 107 Studienzentren in 17 verschiedenen Ländern weltweit durchgeführt.

Innerhalb der und zwischen den Studien waren die Behandlungsgruppen Nintedanib im Vergleich zu Placebo hinsichtlich der Patientencharakteristika grundsätzlich vergleichbar (Tabelle 4-9). Das mittlere Alter in den Studien betrug 66,4 Jahre bis 67,1 Jahre und der Anteil an männlichen Studienteilnehmern lag zwischen 76 % und 82 %. Auch die Zeit seit der IPF-Diagnose war in den Behandlungsgruppen der Studien INPULSIS-1 und INPULSIS-2 mit 1,55 Jahre bis 1,66 Jahren nahezu identisch.

Übertragbarkeit auf den deutschen Versorgungskontext

Die Studien wurden unter anderem in deutschen Studienzentren durchgeführt.

Das Durchschnittsalter der Studienteilnehmer von 64 Jahre bis 67 Jahren spiegelt die Prävalenzverteilung in Deutschland gut wider (vgl. Modul 3, Abschnitt 3.2.3). Wie in den zwei Studien, so tritt die IPF auch in der deutschen Versorgungsrealität bei Männern häufiger auf als bei Frauen. Allerdings ist der Anteil Männer in den Studien noch höher als unter IPF-Patienten in der Gesamtbevölkerung. Des Weiteren wurden in den betrachteten Studien die empfohlenen Diagnose- und Behandlungsrichtlinien der ATS/ERS-Kriterien angewandt, die auch im deutschen Versorgungsalltag zur Anwendung kommen (Behr et al., 2013).

Für Nintedanib liegen zudem keine Hinweise auf biodynamische oder –kinetische Unterschiede zwischen einzelnen Bevölkerungsgruppen vor, die sich auf die Studienergebnisse auswirken könnten.

Daher kann davon ausgegangen werden, dass die Studienergebnisse grundsätzlich auf den deutschen Versorgungskontext übertragbar sind.

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

Bewerten Sie das Verzerrungspotenzial der RCT auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-10: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene
			Patient	Behandelnde Personen			
INPULSIS-1 ^a	ja	ja	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	ja	ja	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

Beide Studien wurden als doppelblinde, randomisierte Placebo-kontrollierte, multizentrische, multinationale Studien durchgeführt.

Die Randomisierung erfolgte adäquat mittels *Interactive Voice Response System* (IVRS) und die Gruppenzuteilung blieb während der gesamten Studie verdeckt, so dass weder der Patient noch die behandelnden Personen Rückschlüsse auf die Behandlung ziehen konnten.

Die Auswertungskriterien für die Endpunkte waren in den Prüfplänen a priori definiert. Die Ergebnisse wurden gemäß der Prüfpläne ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

Der Patientenfluss der jeweiligen Studien sowie die Methodik und Ergebnisse der verwendeten Analysen sind transparent dargestellt.

Es finden sich keine weiteren das Verzerrungspotenzial beeinflussenden Punkte, so dass das Verzerrungspotenzial dieser Studien als *niedrig* eingestuft wird.

4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens beruht. Geben Sie dabei an, welche dieser Endpunkte in den relevanten Studien jeweils untersucht wurden. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-11: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Endpunkt	INPULSIS-1	INPULSIS-2	INPULSIS (gepoolt)
Mortalität			
Zeit bis Tod, unabhängig von der Ursache	ja	ja	ja
Zeit bis respiratorisch bedingtem Tod	ja	ja	ja
Jährliche FVC-Abnahme (ml)	ja	ja	ja
FVC-Responder (10 % vom Sollwert)	ja	ja	ja
FVC-Responder (5 % vom Sollwert)	ja	ja	ja
Morbidität			
Zeit bis zur ersten akuten Exazerbation	ja	ja	ja
Zeit bis zur ersten adjudizierten akuten Exazerbation	ja	ja	ja
Lebensqualität			
Veränderung SGRQ-Gesamtscore	ja	ja	ja
Veränderung SGRQ-I-Gesamtscore	ja	ja	ja
SGRQ-(≤ -4 Punkte)-Responder	ja	ja	ja
SGRQ-I-(≤ -4 Punkte)-Responder	ja ^a	ja ^a	ja ^a
Veränderung SOBQ-Gesamtscore	ja	ja	ja
PGI-C Responder	ja	ja	ja
Veränderung CASA-Q Husten-Symptome	ja	ja	ja

Endpunkt	INPULSIS-1	INPULSIS-2	INPULSIS (gepoolt)
Veränderung CASA-Q Husten-Impact	ja	ja	ja
Veränderung EQ-5D- VAS	ja	ja	ja
Verringerung von unerwünschten Ereignissen			
UE	ja	ja	ja
SUE	ja	ja	ja
UE, die zum Therapieabbruch führten	ja	ja	ja
UE, die zum Tod führten	ja	ja	ja
UE von besonderem Interesse	ja	ja	ja
a: Post hoc-Analyse			

Die Phase III-Studien INPULSIS-1 und INPULSIS-2 haben einen homogenen Studienaufbau. Sie sind für die gepoolte Analyse im Rahmen einer IPD Meta-Analyse ausgelegt und somit für eine solche Auswertung geeignet. Auf Basis der gepoolten Analyse wird eine wesentlich höhere Ergebnissicherheit im Vergleich zu den einzelnen Studien erreicht. Deshalb wird in der folgenden Ergebnisdarstellung und -bewertung auf die Ergebnisse der gepoolten Auswertung fokussiert.

4.3.1.3.1 Mortalität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Meta-Analyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des *Odds Ratios* und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (z. B. Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.2). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

4.3.1.3.1.1 Endpunkte Zeit bis zum Tod

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-12: Operationalisierung von *Zeit bis zum Tod*

Studie	Operationalisierung
INPULSIS-1 ^a	<p><u>Zeit bis Tod, unabhängig von der Ursache</u></p> <p>Definition: Zeit von Randomisierung bis Todeszeitpunkt bzw. zu letztem Patientenkontakt, definiert als Datum des letzten im Case Report Form (CRF) dokumentierten Ereignisses</p> <p>Erhebung: durch Prüfarzt, Ermittlung des letzten Patientenkontakts aus CRF-Dokumentation, von Baseline bis Tag 372 (52 Wochen + 7 Tage Toleranz)</p> <p><u>Zeit bis respiratorisch bedingtem Tod</u></p> <p>Definition: Zeit von Randomisierung bis zum Zeitpunkt des respiratorisch bedingten Todes bzw. letztem Patientenkontakt, definiert als Datum des letzten im CRF dokumentierten Ereignisses für Todesfälle, die durch das IAC als respiratorisch bedingt eingestuft wurden</p> <p>Erhebung: durch Prüfarzt, adjudiziert durch IAC, Ermittlung des letzten Patientenkontakts aus CRF-Dokumentation, von Baseline bis Tag 372 (Woche 52 + 7 Tage Toleranz)</p>
INPULSIS-2 ^a	homogenes Studiendesign wie INPULSIS-1
a: Die Operationalisierungen der Endpunkte zur gepoolten Auswertung der INPULSIS Studien waren identisch mit denen der Einzelstudien.	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-13: Bewertung des Verzerrungspotenzials für den Endpunkt *Zeit bis zum Tod* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	<u>Zeit bis Tod, unabhängig von der Ursache</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>Zeit bis respiratorisch bedingtem Tod</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-1 ^a	<u>Zeit bis Tod, unabhängig von der Ursache</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>Zeit bis respiratorisch bedingtem Tod</u>					
	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die Endpunkte *Zeit bis zum Tod* erfolgte verblindet. Die Feststellung eines Todesfalles ist zudem objektiv möglich. Die Beurteilung der Todesursache (z. B. respiratorische Ursache) erfolgte primär durch den Prüfarzt. Zusätzlich wurde die Todesursache durch ein unabhängiges Bewertungskomitee anhand der Patientenakten adjudiziert. Dadurch kann von einer objektiven Einschätzung der Todesursache ausgegangen werden.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen haben. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien für die Endpunkte *Zeit bis zum Tod* waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll ist transparent. Hinweise auf ein erhöhtes Verzerrungspotenzial liegen diesbezüglich nicht vor.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für die Endpunkte *Zeit bis zum Tod* ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt Zeit bis zum Tod für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

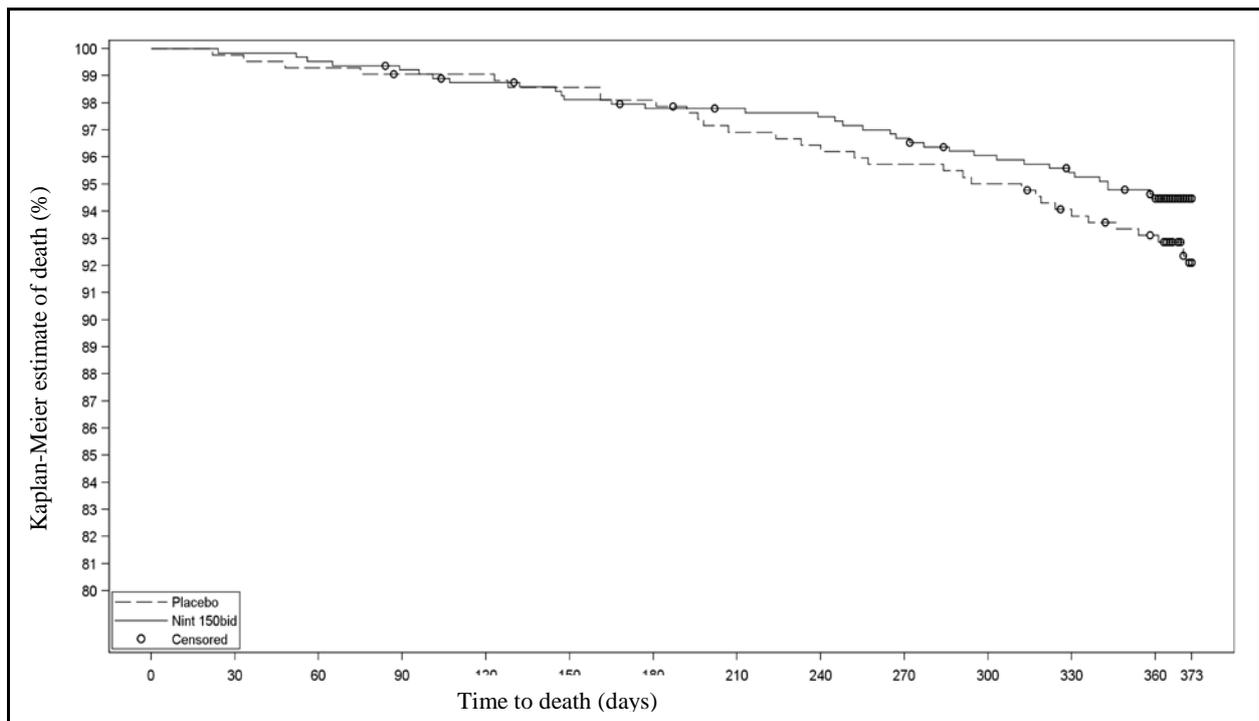
Tabelle 4-14: Ergebnisse für *Zeit bis zum Tod* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib			Placebo			Behandlungseffekt Nintedanib vs. Placebo ^d		
	N ^a	n ^b (%)	Überlebens- wahrschein- lichkeit ^c	N ^a	n ^b (%)	Überlebens- wahrschein- lichkeit ^c	HR ^d	95 % KI	P- Wert
Zeit bis Tod, unabhängig von der Ursache									
INPULSIS (gepoolt) ^e	638	35 (5,5)	0,945	423	33 (7,8)	0,921	0,70	0,43; 1,12	0,1399
INPULSIS-1 ^e	309	13 (4,2)	0,958	204	13 (6,4)	0,935	0,63	0,29; 1,36	0,2880
INPULSIS-2 ^e	329	22 (6,7)	0,933	219	20 (9,1)	0,908	0,74	0,40; 1,35	0,2995
Zeit bis respiratorisch bedingtem Tod									
INPULSIS (gepoolt) ^e	638	24 (3,8)	0,962	423	21 (5,0)	0,949	0,74	0,41; 1,34	0,3435
INPULSIS-1 ^e	309	10 (3,2)	0,967	204	10 (4,9)	0,950	0,61	0,25; 1,47	0,3515
INPULSIS-2 ^e	329	14 (4,3)	0,956	219	11 (5,0)	0,949	0,86	0,39; 1,90	0,6654
<p>a: Anzahl analysierter Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: Kaplan-Meier-Schätzer an Tag 372 nach Randomisierung für INPULSIS-1, INPULSIS-2 und INPULSIS (gepoolt) d: HR [95 % KI] INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels Cox-Regression mit Termen für Behandlung, Geschlecht, Alter und Größe. Bei INPULSIS (gepoolt) wurde noch der Term Studie hinzugefügt p-Wert INPULSIS-1, INPULSIS-2, INPULSIS (gepoolt): berechnet mittels Log-Rank Test e: TS; OC</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.4.1, Tab. 15.2.2.2.4.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.4.1, Tab. 15.2.2.2.4.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 2.3.1.1.1, Tab. 2.3.1.2.1</p>									

Für die Endpunkte *Zeit bis zum Tod, unabhängig von der Ursache* und *Zeit bis respiratorisch bedingtem Tod* war die Überlebenswahrscheinlichkeit unter der Behandlung mit Nintedanib höher als in der Placebo-Gruppe (Tabelle 4-14).

Der Anteil aller verstorbenen Studienteilnehmer war unter der Behandlung mit Nintedanib geringer als in der Placebo-Gruppe sowohl in den einzelnen Studien INPULSIS-1 (4,2 % vs. 6,4 %) und INPULSIS-2 (6,7 % vs. 9,1 %) als auch in den gepoolten Analysen der INPULSIS Studien (5,5 % vs. 7,8 %). Das HR [95 % KI] für die *Zeit bis zum Tod, unabhängig von der Ursache* beträgt in der INPULSIS-1 Studie 0,63 ([0,29; 1,36]; p = 0,2880), in der INPULSIS-2 Studie 0,74 ([0,40; 1,35]; p = 0,2995) und für die gepoolten Daten der INPULSIS Studien 0,70 ([0,43; 1,12]; p = 0,1399).

Der Anteil der aus respiratorischen Gründen verstorbenen Studienteilnehmer war ebenfalls in der Nintedanib-Gruppe niedriger als in der Placebo-Gruppe, sowohl in den einzelnen Studien INPULSIS-1 (3,2 % vs. 4,9 %) und INPULSIS-2 (4,3 % vs. 5,0 %) als auch in den gepoolten Daten der INPULSIS Studien (3,8 % vs. 5,0 %). Das HR [95 % KI] für die *Zeit bis respiratorisch bedingtem Tod* beträgt in der INPULSIS-1 Studie 0,61 ([0,25; 1,47]; $p = 0,3515$), in der INPULSIS-2 Studie 0,86 ([0,39; 1,90]; $p = 0,6654$) und für die gepoolte Auswertung der INPULSIS Studien 0,74 ([0,41; 1,34]; $p = 0,3435$).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-6: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zum Tod*, unabhängig von der Ursache für die gepoolte Analyse der INPULSIS Studien

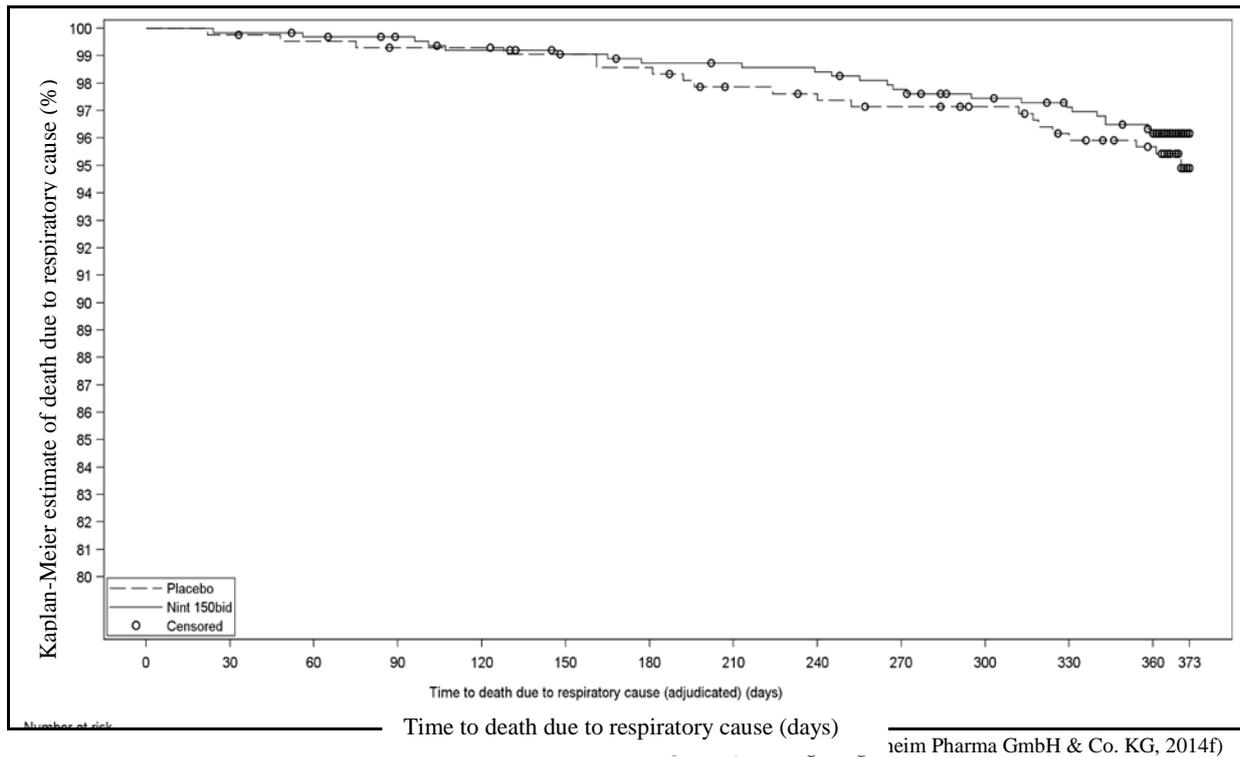


Abbildung 4-7: Kaplan-Meier-Kurven für Endpunkt *Zeit bis respiratorisch bedingtem Tod* für die gepoolte Analyse der INPULSIS Studien

Die Ergebnisse für die Endpunkte *Zeit bis zum Tod, unabhängig von der Ursache* und *Zeit bis respiratorisch bedingtem Tod* zeigen in allen der Bewertung zugrunde liegenden Studien einen Trend zugunsten von Nintedanib (Abbildung 4-6, Abbildung 4-7). Dies deutet auf ein geringeres Mortalitätsrisiko für Nintedanib-behandelte Studienteilnehmer hin.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.1.1 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.1.2 Endpunkte *Forcierte Vitalkapazität*

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-15: Operationalisierung von FVC

Studie	Operationalisierung
INPULSIS-1 ^a	<p><u>Jährliche FVC-Abnahme (ml)</u></p> <p>Definition: Jährliche FVC-Abnahme (ml) errechnet mittels Random-Koeffizienz-Modell aus der Veränderung der FVC im Zeitverlauf (Daten zu mehreren Erhebungszeitpunkten)</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Prüfarzt, mittels Spirometer gemäß ATS/ERS-Kriterien, von Baseline bis 52 Wochen - Test wurde dreifach in sitzender Position jeweils zur gleichen Tageszeit (\pm 60 min.) und ohne größere körperliche Belastung innerhalb von 12 Std. vor der Erhebung durchgeführt. Studienteilnehmer wurden gebeten am Tag der Erhebung nicht zu rauchen, 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. <p>Auswaschphase für Bronchodilatatortherapie von 24 Std. (für langwirksame Wirkstoffe) bzw. 8 Std. (für Wirkstoffe mit kurzer Halbwertszeit) vor Testdurchführung.</p> <p><u>FVC-Responder (Schwellenwert 10 % bzw. 5 % vom Sollwert^b)</u></p> <p>Definition: Anteil Studienteilnehmer mit einem Abfall der FVC nicht größer als 10 % bzw. 5 % vom Sollwert^b, jeweils bezogen auf den Sollwert</p> <p>Erhebung: durch Prüfarzt, mittels Spirometer gemäß ATS/ERS-Kriterien, von Baseline zu Woche 52</p>
INPULSIS-2 ^a	homogenes Studiendesign wie INPULSIS-1
<p>a: Die Operationalisierungen der Endpunkte zur gepoolten Auswertung der INPULSIS Studien waren identisch mit denen der Einzelstudien.</p> <p>b: Cut Off-Wert 10 % vom Sollwert: Hauptanalyse, Cut Off-Wert 5 % vom Sollwert: supportive Analyse (Abschnitt 4.5.4)</p>	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-16: Bewertung des Verzerrungspotenzials für den Endpunkt *FVC* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	<u>Jährliche FVC-Abnahme (ml)</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>FVC-Responder (Schwellenwert von 10 % und 5 % vom Sollwert^b)</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>Jährliche FVC-Abnahme (ml)</u>					
	niedrig	ja	ja	ja	ja	niedrig
	<u>FVC-Responder (Schwellenwert von 10 % und 5 % vom Sollwert^b)</u>					
	niedrig	ja	ja	ja	ja	niedrig
a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.						
b: Cut Off-Wert 10 % vom Sollwert: Hauptanalyse, Cut Off-Wert 5 % vom Sollwert: supportive Analyse (Abschnitt 4.5.4)						

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die Endpunkte zur *forcierten Vitalkapazität* erfolgte doppelblind. Die FVC ist zudem objektiv bestimmbar.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen haben. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien für die Endpunkte zur Lungenfunktion waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll ist transparent. Hinweise auf ein erhöhtes Verzerrungspotenzial liegen diesbezüglich nicht vor.

- Verzerrungspotenzial auf Endpunktebene

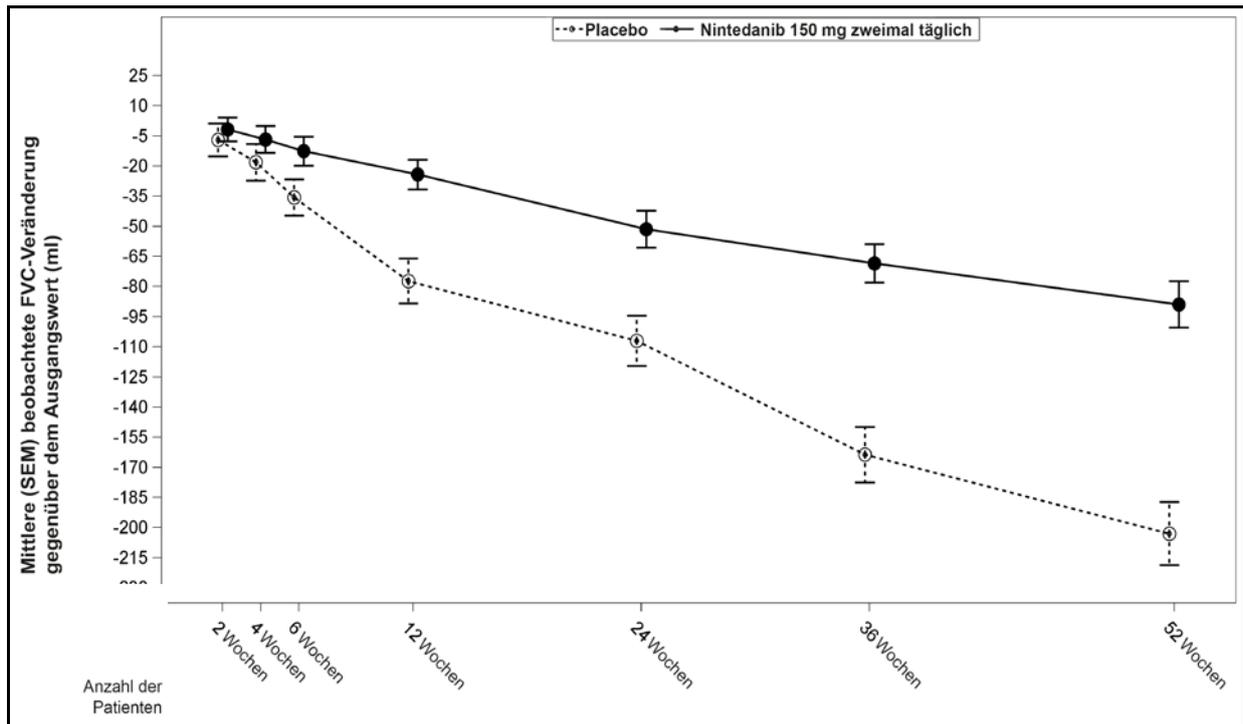
Abschließend ergibt sich für die Endpunkte zur *forcierten Vitalkapazität* ein niedriges Verzerrungspotenzial.

*Stellen Sie die Ergebnisse für den Endpunkt *forcierte Vitalkapazität* für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.*

Tabelle 4-17: Ergebnisse für jährliche FVC-Abnahme (ml) aus RCT mit dem zu bewertenden Arzneimittel

Studie	Behandlung	N ^a	Baseline in ml	Woche 52 in ml	jährliche Abnahme in ml	Behandlungseffekt Nintedanib vs. Placebo ^c in ml		
			MW ^b (SD)	MW ^b (SD)	MWD ^c (SE)	MWD (SE)	95 % KI	p-Wert
jährliche Abnahme der FVC								
INPULSIS (gepoolt) ^d	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)			
INPULSIS-1 ^d	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 ^d	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)			
<p>a: Anzahl analysierter Studienteilnehmer b: nicht adjustierte Mittelwerte c: MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels Random-Coeffizient-Regression mit festen Effekten für Behandlung, Geschlecht, Alter und Größe und zufälligem Effekt für Studienteilnehmer-spezifischen Achsenabschnitt und Zeit; zusätzlich zu Baseline und Woche 52 wurden alle weiteren Zeitpunkte zur Berechnung der jährlichen Abnahme einbezogen MWD [95 % KI] und p-Wert INPULSIS (gepoolt): berechnet mittels Random-Coeffizient-Regression mit festen Effekten für Studie, Behandlung, Geschlecht, Alter und Größe und zufälligem Effekt für Studienteilnehmer-spezifischen Achsenabschnitt und Zeit; zusätzlich zu Baseline und Woche 52 wurden alle weiteren Zeitpunkte zur Berechnung der jährlichen Abnahme einbezogen d: TS; OC Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.1.1, Tab. 15.2.2.2.1.1.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.1.1, Tab. 15.2.2.2.1.1.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 2.1.1.1.; Tab. 15.2.1.1.1., Tab. 15.2.1.3.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 24.10.1.1</p>								

Die jährliche Abnahme der FVC war sowohl in den betrachteten Einzelstudien als auch in den gepoolten Daten der INPULSIS Studien in der Nintedanib-Gruppe geringer als in der Placebo-Gruppe (FVC-Abnahme: 114,65 ml vs. 239,91 ml, 113,59 ml vs. 207,32 ml, 113,59 ml vs. 223,53 ml) (Tabelle 4-17). Die adjustierte Mittelwertdifferenz [95 % KI] beträgt in der INPULSIS-1 Studie 125,26 ml ([77,68; 172,84]; $p < 0,0001$) und in der INPULSIS-2 Studie 93,72 ml ([44,78; 142,68]; $p = 0,0002$) zugunsten einer geringeren jährlichen Abnahme der FVC im Nintedanib-Untersuchungsarm. Für die gepoolte Auswertung der INPULSIS Studien beträgt die adjustierte Mittelwertdifferenz [95 KI %] 109,94 ml ([75,85; 144,03]; $p < 0,0001$).



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-8: Mittlere beobachtete FVC-Veränderung (standard error of the mean, SEM) gegenüber dem Ausgangswert (ml) im Zeitverlauf, Studien INPULSIS-1 und INPULSIS-2 gepoolt

In allen Studien führte Nintedanib zu einer geringeren *jährlichen FVC-Abnahme* im Vergleich zu Placebo. Der Unterschied ist in allen Auswertungen statistisch signifikant.

Tabelle 4-18: Ergebnisse für *FVC-Responder* Schwellenwert 10 % (5 %) vom Sollwert aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c		
	N ^a	n ^b (%)	N ^a	n ^b (%)	OR	95 % KI	p-Wert
FVC-Responder (Schwellenwert 10 % vom Sollwert)							
INPULSIS (gepoolt) ^d	638	447 (70,06)	423	256 (60,52)	1,577	1,21; 2,05	0,0007
INPULSIS-1 ^d	309	218 (70,55)	204	116 (56,86)	1,914	1,32; 2,79	0,0007
INPULSIS-2 ^d	329	229 (69,60)	219	140 (63,93)	1,286	0,89; 1,86	0,1833
FVC-Responder (Schwellenwert 5 % vom Sollwert)							
INPULSIS (gepoolt) ^d	638	338 (53,0)	423	164 (38,8)	1,835	1,43; 2,36	< 0,0001
INPULSIS-1 ^d	309	163 (52,8)	204	78 (38,2)	1,857	1,28; 2,66	0,0010
INPULSIS-2 ^d	329	175 (53,2)	219	86 (39,3)	1,794	1,26; 2,55	0,0011
<p>a: Anzahl analysierter Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: OR [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Termen für Behandlung, Alter, Geschlecht, Größe und Baseline-FVC (%); bei INPULSIS (gepoolt) wurde noch der Term Studie dem Modell hinzugefügt d: TS; OC</p> <p>Studienteilnehmer waren FVC-Responder, wenn sie keinen absoluten Abfall der FVC von mehr als 10 % (5 %) der FVC (% vom Sollwert) hatten und ein Wert zu Woche 52 vorlag. Studienteilnehmer mit fehlenden Daten wurden als Non-Responder deklariert</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.1.1.3:3, Tab. 15.2.2.2.1.1.3:1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.1.1.3:3, Tab. 15.2.2.2.1.1.3:1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 2.3.2.1.3.3, Tab. 2.3.2.1.3.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.13.1.1</p>							

Für den Endpunkt *FVC-Responder* ist der Anteil an Studienteilnehmern mit einer Veränderung der FVC um < 10 % vom Sollwert von Baseline bis zum Erhebungszeitpunkt nach Woche 52 im Nintedanib-Untersuchungsarm größer als in der Placebo-Gruppe (Tabelle 4-18).

Sowohl in der INPULSIS-1 Studie (70,55 % vs. 56,86 %) als auch in der INPULSIS-2 Studie (69,60 % vs. 63,93 %) und folglich auch in den gepoolten Daten der INPULSIS Studien (70,06 % vs. 60,52 %), war der Anteil *FVC-Responder* (Schwellenwert < 10 % vom Sollwert) in der Nintedanib-Gruppe größer als in der Placebo-Gruppe. Das OR [95 % KI] beträgt für die INPULSIS-1 Studie 1,914 ([1,32; 2,79]; p = 0,0007) und für die INPULSIS-2 Studie 1,286 ([0,89; 1,86]; p = 0,1833). In der gepoolten Auswertung der INPULSIS Studien beträgt das OR [95 % KI] entsprechend 1,577 ([1,21; 2,05]; p = 0,0007). Die gepoolten Ergebnisse der INPULSIS Studien zeigen einen statistisch signifikanten Unterschied zwischen den Studiengruppen und belegen somit den Vorteil zugunsten von Nintedanib gegenüber Placebo.

Für die Responder-Analyse mit einer Veränderung der *FVC* bezogen auf den Schwellenwert von < 5 % vom Sollwert von Baseline bis zum Erhebungszeitpunkt nach Woche 52 zeigten sich vergleichbare Ergebnisse (Tabelle 4-18). In allen der Bewertung zugrunde liegenden Studien war der Anteil *FVC*-Responder in der Nintedanib-Gruppe größer als in der Placebo-Gruppe. Alle Ergebnisse zu diesem Endpunkt waren statistisch signifikant zugunsten von Nintedanib.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.1.2 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2 Morbidität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Meta-Analyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (z. B. Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

4.3.1.3.2.1 Endpunkte akute Exazerbationen

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-19: Operationalisierung von *akute Exazerbationen*

Studie	Operationalisierung
INPULSIS-1 ^a	<p><u>Zeit bis zur ersten adjudizierten akuten Exazerbation</u></p> <p>Definition: Zeit von Randomisierung bis zur ersten akuten Exazerbation (definiert gemäß Studienprotokoll), die nachträglich durch ein IAC adjudiziert wurde (präspezifiziert im Studienprotokoll).</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Prüfarzt (Adjudizierung durch IAC), gemäß im Studienprotokoll präspezifizierter Definition für Exazerbationen und deren Adjudizierung, von Baseline bis Tag 372 (52 Wochen + 7 Tage Toleranz) <p>Relevante klinische Daten wurden in spezifischem Fragebogen zusammengefasst (Zusammenfassung Krankheitsverlauf, Anzeichen und Symptome, Laborergebnisse, Lungenfunktion, Ergebnisse Bildgebungsverfahren, erforderliche Behandlung)</p> <p><u>Zeit bis zur ersten akuten Exazerbation</u></p> <p>Definition: Zeit von Randomisierung bis zur ersten akuten Exazerbation</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Prüfarzt, gemäß im Studienprotokoll präspezifizierter Definition für Exazerbationen, von Baseline bis Tag 372 (52 Wochen + 7 Tage Toleranz) - Studienteilnehmer wurden sensibilisiert, frühzeitig eine Verschlechterung in der Dyspnoe dem Untersucher zu berichten - Relevante klinische Daten wurden in spezifischem Fragebogen zusammengefasst (Zusammenfassung Krankheitsverlauf, Anzeichen und Symptome, Laborergebnisse, Lungenfunktion, Ergebnisse Bildgebungsverfahren, erforderliche Behandlung)
INPULSIS-2 ^a	homogenes Studiendesign wie INPULSIS-1
a: Die Operationalisierungen der Endpunkte zur gepoolten Auswertung der INPULSIS Studien waren identisch mit denen der Einzelstudien.	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-20: Bewertung des Verzerrungspotenzials für den Endpunkt *akute Exazerbationen* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	<u>Zeit bis zur ersten adjudizierten akuten Exazerbation</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>Zeit bis zur ersten akuten Exazerbation</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-1 ^a	<u>Zeit bis zur ersten adjudizierten akuten Exazerbation</u>					
	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	<u>Zeit bis zur ersten akuten Exazerbation</u>					
	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die Endpunkte *Zeit bis zur ersten (adjudizierten) akuten Exazerbation* erfolgte doppelblind. Das Vorliegen einer akuten Exazerbation wurde durch den Prüfarzt festgestellt. Zusätzlich wurden andere Ursachen als die zugrundeliegende IPF für die als Exazerbation imponierende akute Verschlechterung des Gesundheitszustands durch ein verblindetes IAC anhand der Patientenakten ausgeschlossen. Dadurch kann von einer objektiven Einschätzung des Ereignisses Exazerbation ausgegangen werden.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen haben. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien für die Endpunkte *Zeit bis zur ersten (adjudizierten) akuten Exazerbation* waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war transparent, Hinweise auf ein erhöhtes Verzerrungspotenzial liegen diesbezüglich nicht vor.

Exazerbation sind seltene aber potenziell lebensbedrohliche Ereignisse bei IPF-Patienten, die hinsichtlich ihrer Klinik manchmal nur schwer von ähnlichen Ereignissen, die mit einer plötzlichen Verschlechterung des Gesundheitszustandes einhergehen, abgegrenzt werden können. Um diesem Umstand Rechnung zu tragen, wurde in dieser Studie eine zusätzliche Überprüfung der durch den Prüfarzt gestellten Diagnose durchgeführt. Die Erhebung der *Zeit bis zur ersten adjudizierten akuten Exazerbation* unterscheidet sich dadurch maßgeblich von der Erhebung der *Zeit bis zur ersten akuten Exazerbation*. Fachärzte für Pneumologie mit spezieller Expertise auf dem Gebiet der IPF haben auf Basis einer strengen Definition für IPF-getriggerte Exazerbationen diesen patientenrelevanten Endpunkt - ebenfalls wie die Prüfarzte verblindet - bewertet. Damit ist der Stellenwert der adjudizierten Ereignisse als präziser einzustufen, als derjenige des Endpunkts *Zeit bis zur ersten akuten Exazerbation* (ohne anschließende Adjudizierung).

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* und *Zeit bis zur ersten akuten Exazerbation* ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt akute Exazerbationen für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-21: Ergebnisse für *Zeit bis zur ersten adjudizierten akuten Exazerbationen* und *Zeit bis zur ersten akuten Exazerbation* aus RCT mit dem zu bewertenden Arzneimittel

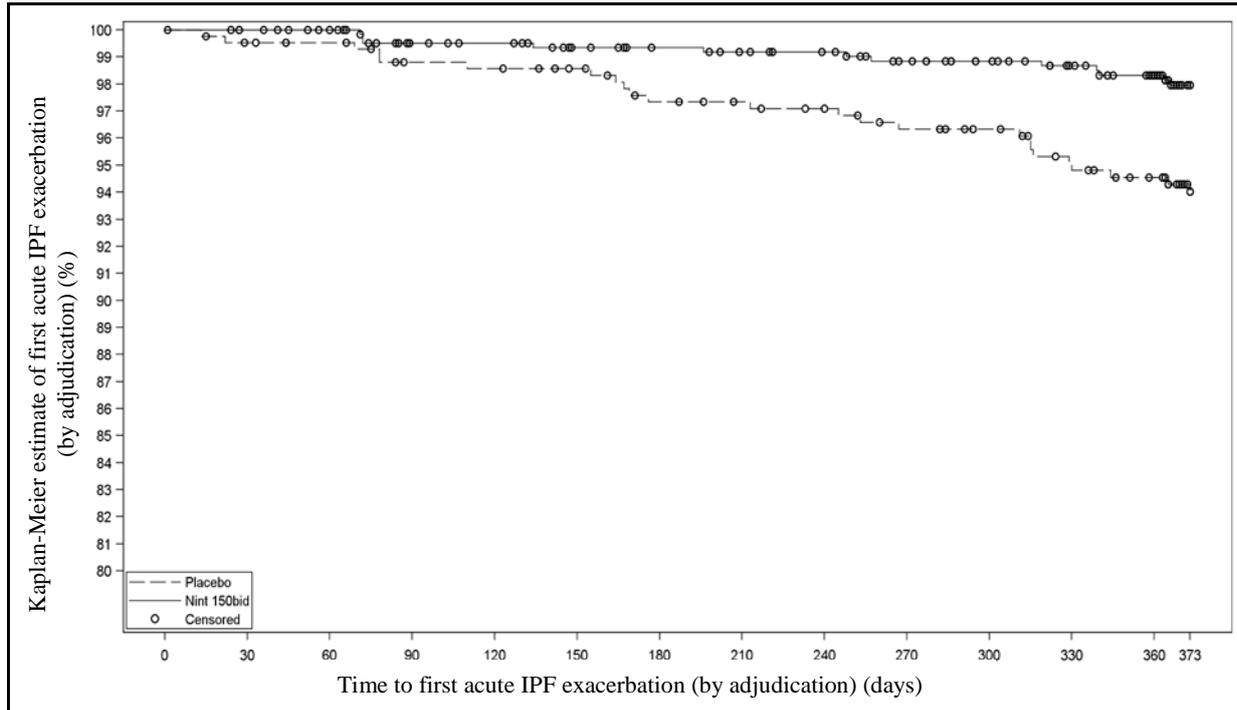
Studie	Nintedanib			Placebo			Behandlungseffekt Nintedanib vs. Placebo ^d		
	N ^a	n ^b (%)	Wahr- scheinlich- keit, keine Exazerbation zu erleiden ^c	N ^a	n ^b (%)	Wahr- scheinlich- keit, keine Exazerbation zu erleiden ^c	HR	95 % KI	p-Wert
Zeit bis zur ersten adjudizierten akuten Exazerbation									
INPULSIS ^e (gepoolt)	638	12 (1,9)	0,980	423	24 (5,7)	0,940	0,32	0,16; 0,65	0,0010
INPULSIS-1 ^e	309	7 (2,3)	0,975	204	8 (3,9)	0,959	0,55	0,20; 1,54	0,2551
INPULSIS-2 ^e	329	5 (1,5)	0,984	219	16 (7,3)	0,922	0,20	0,07; 0,56	0,0020
Zeit bis zur ersten akuten Exazerbation									
INPULSIS ^e (gepoolt)	638	31 (4,9)	0,948	423	32 (7,6)	0,920	0,64	0,39; 1,05	0,0823
INPULSIS-1 ^e	309	19 (6,1)	0,934	204	11 (5,4)	0,944	1,15	0,54; 2,42	0,6728
INPULSIS-2 ^e	329	12 (3,6)	0,961	219	21 (9,6)	0,898	0,38	0,19; 0,77	0,0050
<p>a: Anzahl analysierter Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: Kaplan-Meier-Schätzer an Tag 372 nach Randomisierung für INPULSIS-1, INPULSIS-2 und INPULSIS (gepoolt) d: HR [95 % KI] INPULSIS-1, INPULSIS-2 und INPULSIS (gepoolt) Studien: berechnet mittels Cox-Regression mit Termen für Behandlung, Geschlecht, Alter und Größe p-Wert INPULSIS-1, INPULSIS-2, INPULSIS (gepoolt): berechnet mittels Log-Rank-Test e: TS; OC f: RS; OC</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.1.3; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.1.3; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 2.2.2.1.1, Tab. 2.2.2.2.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.17.1.1, Tab. 86.17.1.3, Tab. 86.16.1.1, Tab. 86.16.1.3</p>									

Für die Endpunkte *adjudizierte akute Exazerbation* und *Zeit bis zur ersten akuten Exazerbation* wurde die Wahrscheinlichkeit, keine Exazerbation zu erleiden, mittels Kaplan-Meier-Schätzer pro Behandlungsgruppe nach Woche 52 dargestellt. Das HR [95 % KI] beschreibt den Effekt der Zeit bis zum Erreichen des Ereignisses *akute Exazerbation* (Tabelle 4-21).

Eine *adjudizierte akute Exazerbation* trat in den einzelnen Phase III-Studien und der gepoolten Analyse der INPULSIS Studien bei einem geringeren Anteil der mit Nintedanib behandelten Studienteilnehmer auf als dies in der Placebo-Gruppe der Fall war (INPULSIS-1: 2,3 % vs. 3,9 %, INPULSIS-2: 1,5 % vs. 7,3 %, INPULSIS Studien (gepoolt): 1,9 % vs. 5,7 %). Das HR [95 % KI] beträgt für die INPULSIS-1 Studie 0,55 ([0,20; 1,54]; $p = 0,2551$), für die INPULSIS-2 Studie 0,20 ([0,07; 0,56]; $p = 0,0020$) und für die gepoolte Auswertung dieser Studien 0,32 ([0,16; 0,65]; $p = 0,0010$). Der Unterschied zugunsten von Nintedanib im Auftreten von *adjudizierten akuten Exazerbationen* ist in der INPULSIS-2 Studie und der gepoolten Auswertung der Phase III-Studien statistisch signifikant.

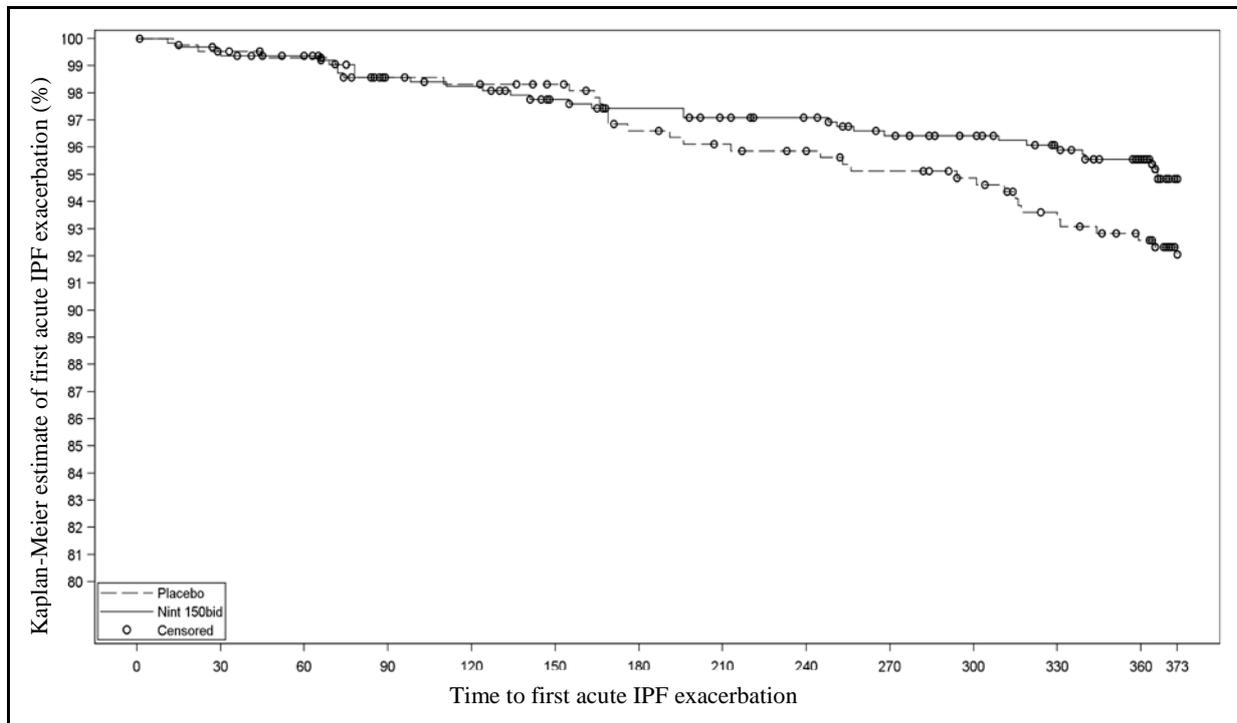
Vom Prüfarzt diagnostizierte *akute Exazerbationen (nicht adjudiziert)* traten in der INPULSIS-2 Studie sowie in den gepoolten Daten der INPULSIS Studien seltener im Nintedanib-Arm auf als unter Placebo (INPULSIS-2: 3,6 % vs. 9,6 %; INPULSIS Studien (gepoolt): 4,9 % vs. 7,6 %). In der INPULSIS-1 Studie traten solche Ereignisse in beiden Studienarmen ähnlich häufig auf (INPULSIS-1: 6,1 % vs. 5,4 %). Das HR [95 % KI] beträgt in der INPULSIS-1 Studie 1,15 ([0,54; 2,42]; $p = 0,6728$), in der INPULSIS-2 Studie 0,38 ([0,19; 0,77]; $p = 0,0050$) und in den gepoolten Daten der INPULSIS Studien 0,64 ([0,39; 1,05]; $p = 0,0823$). Insgesamt zeichnet sich für den Endpunkt *akute Exazerbationen* eine klare Tendenz hinsichtlich eines reduzierten Risikos unter der Behandlung mit Nintedanib ab.

Die Kaplan-Meier-Kurven der Studienteilnehmer mit akuten bzw. adjudizierten akuten Exazerbationen in den Behandlungsgruppen Nintedanib und Placebo sind in Abbildung 4-9 und in Abbildung 4-10 dargestellt.



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-9: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zur ersten adjudizierten akuten Exazerbation* für die gepoolte Analyse der INPULSIS Studien



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

Abbildung 4-10: Kaplan-Meier-Kurven für Endpunkt *Zeit bis zur ersten akuten Exazerbation* für die gepoolte Analyse der INPULSIS Studien

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits im Abschnitt 4.3.1.3.2.1 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.3 Gesundheitsbezogene Lebensqualität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Meta-Analyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse– Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des *Odds Ratios* und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (z. B. Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

4.3.1.3.3.1 Endpunkte zum SGRQ, SGRQ-I, SOBQ, CASA-Q, PGI- C und EQ-5D

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-22: Operationalisierung zu SGRQ, SGRQ-I, SOBQ, CASA-Q, PGI- C und EQ-5D

Studie	Operationalisierung
INPULSIS-1 ^a	<p><u>Veränderung SGRQ-Gesamtscore</u></p> <p>Definition: Veränderung des SGRQ-Gesamtscores über die Zeit. Der SGRQ-Gesamtscore setzt sich zusammen aus den drei Domänen Symptome, Aktivität und Impact. Ein niedrigerer Score steht für eine bessere subjektiv empfundene krankheitsspezifische Lebensqualität.</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer mittels SGRQ-Fragebogen, von Baseline bis Woche 52 <p><u>SGRQ-(≤ -4 Punkte)-Responder</u></p> <p>Definition: Anteil Studienteilnehmer mit einer Veränderung des SGRQ-Gesamtscores von ≤ -4 Punkten von Baseline bis Woche 52.</p> <p>Erhebung: durch Studienteilnehmer mittels SGRQ-Fragebogen, von Baseline bis Woche 52.</p> <p><u>Veränderung SGRQ-I-Gesamtscore</u></p> <p>Definition: Veränderung des IPF-spezifischen SGRQ-I-Gesamtscores über die Zeit. Beim SGRQ-I Fragebogen handelt sich um eine IPF-spezifische SGRQ-Fragebogenversion. Ein niedrigerer Score steht für eine bessere subjektiv empfundene krankheitsspezifische Lebensqualität.</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer mittels SGRQ-Fragebogen, von Baseline bis Woche 52. Der SGRQ-I-Gesamtscore wurde aus den Antworten im SGRQ-Fragebogen nach einem von den Entwicklern des SGRQ-I definierten Algorithmus ermittelt. <p><u>SGRQ-I-(≤ -4 Punkte)-Responder</u></p> <p>Definition: Anteil Studienteilnehmer mit einer Veränderung des SGRQ-I-Gesamtscores von ≤ -4 Punkten von Baseline bis Woche 52.</p> <p>Erhebung: durch Studienteilnehmer mittels SGRQ-Fragebogen, von Baseline bis Woche 52. Der SGRQ-I-Gesamtscore wurde aus den Antworten im SGRQ-Fragebogen nach einem von den Entwicklern des SGRQ-I definierten Algorithmus ermittelt.</p> <p><u>Veränderung SOBQ-Gesamtscore</u></p> <p>Definition: Veränderung des SOBQ-Gesamtscores über die Zeit. 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). Ein niedrigerer Score steht für eine geringere Beeinträchtigung der Lebensqualität durch Kurzatmigkeit.</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer mittels SOBQ-Fragebogen, von Baseline bis Woche 52 <p><u>Veränderung CASA-Q Husten-Symptome</u></p> <p>Definition: Veränderung der CASA-Q Husten-Symptome über die Zeit. Ein höherer Score ist mit weniger Husten-Symptomen assoziiert</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer, mittels CASA-Q-Fragebogen (Likert-Skala 1-5), von Baseline bis Woche 52

Studie	Operationalisierung
	<p><u>Veränderung CASA-Q Husten-Impact</u></p> <p>Definition: Veränderung der CASA-Q Husten-Impact über die Zeit. Ein höherer Score ist mit weniger Husten-Impact assoziiert</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer, mittels CASA-Q-Fragebogen (Likert-Skala 1-5), von Baseline bis Woche 52 <p><u>PGI-C-Responder</u></p> <p>Definition: Anteil Studienteilnehmer, die ihre Lebensqualität gegenüber dem Befinden bei Studienbeginn als „sehr viel besser“, „viel besser“, „ein wenig besser“ oder „ohne Veränderung“ einschätzten. Die Einschätzung erfolgte auf einer 7-Punkte-Likert-Skala von eins (sehr viel besser) bis sieben (sehr viel schlechter).</p> <p>Erhebung:</p> <ul style="list-style-type: none"> - durch Studienteilnehmer, mittels PGI-C-Fragebogen (7-Punkte-Likert-Skala), von Baseline bis Woche 52 <p><u>Veränderung EQ-5D-VAS</u></p> <p>Definition: Veränderung der EQ-5D-VAS (Skala 0-100). Ein höherer Score steht für eine bessere subjektiv empfundene Lebensqualität.</p> <p>Erhebung:</p> <p>durch Studienteilnehmer, mittels EQ-5D-Fragebogen, von Baseline bis Woche 52</p>
INPULSIS-2 ^a	homogenes Studiendesign wie INPULSIS-1
a: Die Operationalisierungen der Endpunkte zur gepoolten Auswertung der INPULSIS Studien waren identisch mit denen der Einzelstudien.	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-23: Bewertung des Verzerrungspotenzials für die Endpunkte *SGRQ*, *SGRQ-I*, *SOBQ*, *CASA-Q*, *PGI-C* und *EQ-5D* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	<u>Veränderung SGRQ-Gesamtscore</u>					
	niedrig	ja	ja	ja	ja	niedrig
	<u>SGRQ(< -4 Punkte)-Responder</u>					
	niedrig	ja	ja	ja	ja	niedrig
	<u>Veränderung SGRQ-I-Gesamtscore</u>					
	niedrig	ja	ja	ja	ja	niedrig
	<u>SGRQ-I(< -4 Punkte)-Responder</u>					
	niedrig	ja	ja	ja	ja	niedrig
	<u>Veränderung SOBQ-Gesamtscore</u>					
	niedrig	ja	ja	ja	ja	niedrig
<u>Veränderung CASA-Q Husten-Symptome</u>						
niedrig	ja	ja	ja	ja	niedrig	
<u>Veränderung CASA-Q Husten-Impact</u>						
niedrig	ja	ja	ja	ja	niedrig	
<u>PGI-C-Responder</u>						
niedrig	ja	ja	ja	ja	niedrig	
<u>Veränderung EQ-5D-VAS</u>						
niedrig	ja	ja	ja	ja	niedrig	
INPULSIS-2 ^a	homogenes Studiendesign wie INPULSIS-1					
a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.						

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die Endpunkte *SGRQ*, *SGRQ-I*, *SOBQ*, *CASA-Q*, *PGI-C* und *EQ-5D* erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen haben. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

Die Anzahl Studienteilnehmer, die die Fragebögen zu Woche 52 ausreichend ausfüllten, wick in den Endpunkten zur Lebensqualität um mehr als 10 % von der Anzahl randomisierter Studienteilnehmer ab. Diese Beobachtung ist für die Erhebungen zur Lebensqualität nicht ungewöhnlich. Zudem unterschieden sich die Anteile zwischen den Behandlungsgruppen innerhalb der Studien hinsichtlich ausgewerteter randomisierter Studienteilnehmer um nicht mehr als 5 %. Somit liegt ein niedriges Verzerrungspotenzial vor.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien für die Endpunkte zur Lebensqualität waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für alle Endpunkte zur Lebensqualität transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für alle Endpunkte zur Lebensqualität ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für die Endpunkte SGRQ, SGRQ-I, SOBQ, CASA-Q, PGI- C und EQ-5D für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Ergebnisse zum SGRO und SGRO-I

Tabelle 4-24: Ergebnisse für Veränderung SGRQ-Gesamtscore und Veränderung SGRQ-I-Gesamtscore aus RCT mit dem zu bewertenden Arzneimittel

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ^c		
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SE)	MWD (SE)	95 % KI	P-Wert
Veränderung SGRQ-Gesamtscore (Punkte)										
IMPULSIS ^d (gepoolt)	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)			
IMPULSIS-1 ^d	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)			
IMPULSIS-2 ^d	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)			
Veränderung SGRQ-I-Gesamtscore (Punkte)										
IMPULSIS ^d (gepoolt)	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
	Placebo	419	41,94 (19,412)	342	45,08 (21,534)	413	5,47 (0,676)			
IMPULSIS-1 ^d	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)			
IMPULSIS-2 ^d	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)			

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ^c		
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SE)	MWD (SE)	95 % KI	p-Wert
<p>a: Anzahl analysierter Studienteilnehmer b: nicht adjustierte Mittelwerte c: MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels MMRM mit festen Effekten für Behandlung, Behandlung Visite, Baseline SGRQ-Gesamtscore bzw. Baseline SGRQ-I-Gesamtscore, Baseline SGRQ-Gesamtscore*Visite bzw. Baseline SGRQ-I-Gesamtscore*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, Baseline SGRQ-Gesamtscore bzw. Baseline SGRQ-I-Gesamtscore, Baseline SGRQ-Gesamtscore*Visite bzw. Baseline SGRQ-I-Gesamtscore*Visite und zufälligem Effekt für Studienteilnehmer d: TS; OC Ein niedrigerer SGRQ- oder SGRQ-I-Gesamtscore steht für eine bessere subjektiv empfundene krankheitsspezifische Lebensqualität. Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.1.2, Tab. 15.2.2.2.2.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.1.2, Tab. 15.2.2.2.2.2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.6.1.1, Tab. 86.5.1.1</p>										

Der *SGRQ-Gesamtscore* von Baseline bis Woche 52 stieg jeweils in den INPULSIS Studien in beiden Untersuchungsgruppen, Nintedanib vs. Placebo, an (INPULSIS-1: 4,34 Punkte vs. 4,39 Punkte und INPULSIS-2: 2,80 Punkte vs. 5,48 Punkte). In den gepoolten Daten der INPULSIS Studien fiel der Anstieg im *SGRQ-Gesamtscore* unter der Behandlung mit Nintedanib geringer aus als in der Placebogruppe (3,53 Punkte vs. 4,96 Punkte). Dies deutet auf eine geringere Verschlechterung der Lebensqualität unter Nintedanib im Vergleich zu Placebo in dieser Zeitspanne hin. Die adjustierte Mittelwertdifferenz [95 % KI] der zwei Behandlungsgruppen beträgt in der INPULSIS-1 Studie -0,05 Punkte ([-2,50; 2,40]; p = 0,9657) und in der INPULSIS-2 Studie -2,69 Punkte ([-4,95; -0,43]; p = 0,0197). Für die gepoolten Daten der INPULSIS Studien ergab sich eine adjustierte Mittelwertdifferenz [95 % KI] von -1,43 Punkten ([-3,09; 0,23]; p = 0,0923) (Tabelle 4-24).

Eine negative Mittelwertdifferenz ist mit einer Veränderung der subjektiv empfundenen Lebensqualität zugunsten von Nintedanib assoziiert. Diese Ergebnisse belegen einen statistisch signifikanten Unterschied zugunsten von Nintedanib in der INPULSIS-2 Studie.

Der *SGRQ-I-Gesamtscore* stieg von Baseline bis Woche 52 in den Studien INPULSIS-1 (4,30 Punkte vs. 5,08 Punkte), INPULSIS-2 (2,72 Punkte vs. 5,84 Punkte) und den gepoolten Daten dieser beiden Studien (3,48 Punkte vs. 5,47 Punkte) in der Nintedanib-Gruppe weniger stark an als in der Placebo-Gruppe (Tabelle 4-24). Dies entspricht einer geringeren Abnahme der Lebensqualität unter Nintedanib im Vergleich zu Placebo. Die adjustierte MWD [95 % KI] für die Behandlungsgruppen beträgt in der INPULSIS-1 Studie -0,78 Punkte ([-3,31; 1,75]; $p = 0,5446$), in der INPULSIS-2 Studie -3,12 Punkte ([-5,46;-0,79]; $p = 0,0089$) und für die gepoolten Daten der INPULSIS Studien -1,99 Punkte ([-3,71; -0,27]; $p = 0,0230$). Diese negativen Mittelwertdifferenzen zeigen eine bessere subjektiv empfundene Lebensqualität zugunsten von Nintedanib an. Dieser Unterschied ist in der INPULSIS-2 Studie und der gepoolten Auswertung der INPULSIS Studien statistisch signifikant.

Tabelle 4-25: Ergebnisse für *SGRQ-(≤ -4 Punkte)-Responder* und *SGRQ-I-(≤ -4 Punkte)-Responder* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c		
	N ^a	n ^b (%)	N ^a	n ^b (%)	OR	95 % KI	p-Wert
SGRQ-(≤ -4 Punkte)-Responder							
INPULSIS (gepoolt) ^d	638	146 (22,88)	423	86 (20,33)	1,186	0,88; 1,61	0,2715
INPULSIS-1 ^d	309	63 (20,39)	204	49 (24,02)	0,840	0,55; 1,29	0,4298
INPULSIS-2 ^d	329	83 (25,23)	219	37 (16,89)	1,664	1,08; 2,57	0,0218
SGRQ-I-(≤ -4 Punkte)-Responder							
INPULSIS (gepoolt) ^d	638	150 (23,51)	423	83 (19,62)	1,284	0,95; 1,74	0,1091
INPULSIS-1 ^d	309	66 (21,36)	204	49 (24,02)	0,886	0,58; 1,36	0,5821
INPULSIS-2 ^d	329	84 (25,53)	219	34 (15,53)	1,877	1,20; 2,93	0,0056
<p>a: Anzahl analysierter Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: OR [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Termen für Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore OR [95 % KI] und p-Wert INPULSIS (gepoolt): berechnet mittels logistischer Regression mit Termen für Studie, Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore d: TS; OC</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.1.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.1.1; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.1.1.1, Tab. 86.1.1.2, Tab. 79.1.1.1; Tab. 79.1.1.2</p>							

In der INPULSIS-2 Studie (25,23 % vs. 16,89 % Responder) und in der gepoolten Analyse der Phase III-Studien (22,88 % vs. 20,33 % Responder) war der Anteil an Studienteilnehmern, die die klinische Relevanzschwelle einer Änderung von ≤ -4 Punkten im *SGRQ-Gesamtscore*, entsprechend einer klinisch relevanten Verbesserung der subjektiv empfundenen Lebensqualität, in der Nintedanib-Gruppe im Vergleich zur Placebogruppe größer (Tabelle 4-25). Das OR [95 % KI] zeigt für die INPULSIS-1 Studie einen Wert von 0,840 ([0,55; 1,29]; $p = 0,4298$), für die INPULSIS-2 Studie einen Wert von 1,664 ([1,08; 2,57]; $p = 0,0218$) und für die gepoolte Auswertung dieser beiden Studien einen Wert von 1,186 ([0,88; 1,61]; $p = 0,2715$). Der Unterschied zwischen den Behandlungsgruppen war in der INPULSIS-2 Studie statistisch signifikant.

Der Anteil an Studienteilnehmern, die eine Veränderung im *SGRQ-I-Gesamtscore* um ≤ -4 Punkte erreichten, war in der Studie INPULSIS 1 vergleichbar zwischen den beiden Behandlungsgruppen. In der INPULSIS-2 Studie (25,53 % vs. 15,53 %) und in der gepoolten Analyse der INPULSIS Studien (23,51 % vs. 19,62) war der Anteil der Responder in der Nintedanib-Gruppe größer als in der Placebogruppe. Das OR [95 % KI] zeigt für die INPULSIS-1 Studie einen Wert von 0,886 ([0,58; 1,36]; $p = 0,5821$), für die INPULSIS-2 Studie einen Wert von 1,877 ([1,20; 2,93]; $p = 0,0056$) und für die gepoolte Auswertung dieser beiden Studien einen Wert von 1,284 ([0,95; 1,74]; $p = 0,1091$) (Tabelle 4-25).

Ergebnisse zum SOBQTabelle 4-26: Ergebnisse für *Veränderung SOBQ-Gesamtscore* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ^c		
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SE)	MWD (SE)	95 % KI	p-Wert
Veränderung SOBQ-Gesamtscore (Punkte)										
INPULSIS (gepoolt) ^d	Nintedanib	587	32,85 (24,429)	481	37,51 (27,384)	382	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)	569	8,40 (0,945)			
INPULSIS-1 ^d	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 ^d	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)			
<p>a: Anzahl analysierter Studienteilnehmer b: nicht adjustierte Mittelwerte c: MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSIS-2 Studie: berechnet mittels MMRM mit festen Effekten für Behandlung, Behandlung*Visite, Baseline SOBQ-Gesamtscore, Baseline SOBQ-Gesamtscore*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, Baseline SOBQ-Gesamtscore, Baseline SOBQ-Gesamtscore*Visite und zufälligem Effekt für Studienteilnehmer d: TS; OC Ein niedrigerer SOBQ-Gesamtscore steht für eine geringere Beeinträchtigung der Lebensqualität durch Kurzatmigkeit. Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.3;(Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.2.3; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f):Tab. 86.10.1.1</p>										

Die Veränderung von Baseline bis Woche 52 für den *SOBQ-Gesamtscore* war in allen Studien für die Nintedanib-Behandlungsgruppe geringer als für die Placebo-Behandlungsgruppe (INPULSIS-1: 6,73 Punkte vs. 7,61 Punkte, INPULSIS-2: 6,69 Punkte vs. 9,07 Punkte und INPULSIS (gepoolt): 6,71 Punkte vs. 8,40 Punkte) (Tabelle 4-26). Damit hatte sich die subjektiv empfundene krankheitsspezifische Lebensqualität hinsichtlich der Kurzatmigkeit unter der Behandlung mit Nintedanib weniger stark verschlechtert als bei den Placebo-behandelten Studienteilnehmern. Die errechnete adjustierte Mittelwertdifferenz [95 % KI] für die Behandlungsgruppen betrug in der INPULSIS-1 Studie -0,88 Punkte ([-4,35; 2,60]; $p = 0,6203$), in der INPULSIS-2 Studie -2,38 Punkte [-5,68; 0,93]; $p = 0,1587$) und für die gepoolte Auswertung der INPULSIS Studien -1,69 Punkte ([-4,09; 0,70]; $p = 0,1648$).

Interessant ist vor diesem Hintergrund die Beobachtung, dass es in beiden INPULSIS Studien während der Beobachtungsperiode zu Veränderungen des Bedarfs an einer Begleittherapie mit Bronchodilatoren kam (siehe die diesbezüglichen detaillierten Ausführungen im Abschnitt 4.3.1.3.5, das Ergebnisse der Subgruppenanalysen zusammenfasst). Unter der Behandlung mit Studienmedikation fand sich eine Entwicklung hin zu einem höheren Bedarf an Bronchodilator-Begleittherapie in den Placebo-Gruppen sowie eine gegenläufige Entwicklung in den Nintedanib-Gruppen. Dies kann als indirekter Hinweis auf eine höhere subjektive Beeinträchtigung der Patienten durch Dyspnoe/Luftnot unter Placebo-Behandlung verglichen mit Nintedanib interpretiert werden, da Bronchodilatoren üblicherweise zur Behandlung des Symptoms Dyspnoe/Kurzatmigkeit eingesetzt werden. Es ist vorstellbar, dass durch diese Veränderungen in der Begleittherapie in den konkreten Studien möglicherweise die Sensitivität des *SOBQ*-Fragebogens, eine Beeinträchtigung der Lebensqualität durch Kurzatmigkeit nachzuweisen, reduziert wurde.

Ergebnisse zum CASA-Q

Tabelle 4-27: Ergebnisse für Veränderung CASA-Q (Husten-Symptome bzw. Husten-Impact) aus RCT mit dem zu bewertenden Arzneimittel

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ^c		
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SE)	MWD (SE)	95 % KI	P-Wert
Veränderung CASA-Q (Husten-Symptome) (Punkte)										
INPULSIS ^c (gepoolt)	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)			
INPULSIS-1 ^c	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 ^c	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)			
Veränderung CASA-Q (Husten-Impact) (Punkte)										
INPULSIS ^d (gepoolt)	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
	Placebo	423	75,66 (22,125)	345	73,42 (24,011)	417	-4,21 (0,865)			
INPULSIS-1 ^d	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 (22,344)	164	71,91 (25,920)	202	-4,00 (1,240)			
INPULSIS-2 ^d	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)			

Studie	Behandlung	Baseline		Woche 52		Veränderung		Behandlungseffekt Nintedanib vs. Placebo ^c		
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SE)	MWD (SE)	95 % KI	p-Wert
<p>a: Anzahl analysierter Studienteilnehmer b: nicht adjustierte Mittelwerte c: MWD [95 % KI] und p-Wert INPULSIS-1 und INPULSSI-2 Studie: berechnet mittels MMRM mit festen Effekten für Behandlung, Behandlung*Visite, Baseline CASA-Q Husten-Symptome bzw. CASA-Q Husten-Impact, Baseline CASA-Q Husten-Symptome*Visite bzw. CASA-Q Husten-Impact*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, Baseline CASA-Q Husten-Symptome bzw. CASA-Q Husten-Impact, Baseline CASA-Q Husten-Symptome*Visite bzw. CASA-Q Husten-Impact*Visite und zufälligem Effekt für Studienteilnehmer d: TS; OC Ein höherer Score ist mit weniger Husten-Symptomen bzw. weniger Husten-Impact assoziiert. Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.4:1, 15.2.2.2.2.4:2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.2.4:1, 15.2.2.2.2.4:2; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.7.1.1, Tab. 86.8.1.1</p>										

Für den Endpunkt *CASA-Q Husten-Symptome* war die Veränderung von Baseline bis Woche 52 in der INPULSIS-1 Studie in der Nintedanib-Gruppe größer als in der Placebo-Gruppe (-0,76 Punkte vs. -0,52 Punkte) (Tabelle 4-27). Hingegen war in der INPULSIS-2 Studie (-0,33 Punkte vs -2,38 Punkte) und für die gepoolte Analyse der INPULSIS Studien (-0,57 Punkte vs. -1,52 Punkte) die Veränderung dieses Scores unter Nintedanib im Vergleich zur Behandlung mit Placebo geringer. Dies spiegelt eine geringere Verschlechterung der gesundheitsbezogenen Lebensqualität hinsichtlich der Husten-Symptome unter Nintedanib wider. Die errechnete Mittelwertdifferenz [95 % KI] für die Behandlungsgruppen beträgt für die INPULSIS-1 Studie -0,24 Punkte ([-3,78; 3,30]; p = 0,8942), für die INPULSIS-2 Studie 2,05 Punkte ([-1,31; 5,41]; p = 0,2326) und für die gepoolten Daten dieser Studien 0,95 Punkte ([-1,49; 3,38]; p = 0,4450).

In den der Bewertung zugrunde liegenden Studien INPULSIS-1 und INPULSIS-2 sowie für die gepoolte Auswertung dieser Studien war die Veränderung von Baseline bis Woche 52 für den Endpunkt *CASA-Q Husten-Impact* in der Nintedanib-Gruppe geringer als in der Placebo-Gruppe (-2,36 Punkte vs. -4,00 Punkte; -2,58 Punkte vs. -4,39 Punkte, bzw. -2,48 Punkte vs. -4,21 Punkte) (Tabelle 4-27). Die errechnete Mittelwertdifferenz [95 % KI] für die Behandlungseffekte zwischen Nintedanib und Placebo beträgt in der INPULSIS-1 Studie 1,64 Punkte ([-1,49; 4,77]; p = 0,3042), in der INPULSIS-2 Studie 1,81 Punkte ([-1,26; 4,88]; p = 0,2475) sowie für die gepoolten Daten der Studien 1,73 Punkte ([-0,46; 3,92]; p = 0,1212). Diese Unterschiede zeigen die Tendenz, dass die Behandlung mit Nintedanib im Vergleich zu Placebo zu einer geringeren Verschlechterung der gesundheitsbezogenen Lebensqualität in Bezug auf den *Husten-Impact* führt.

Ergebnisse zum PGI-CTabelle 4-28: Ergebnisse für *PGI-C-Responder* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c		
	N ^a	n ^b (%)	N ^a	n ^b (%)	OR	95 % KI	p-Wert
PGI-C-Responder							
INPULSIS (gepoolt) ^d	638	391 (61,29)	423	230 (54,37)	1,328	1,04; 1,70	0,0254
INPULSIS-1 ^d	309	188 (60,84)	204	112 (54,90)	1,276	0,89; 1,83	0,1818
INPULSIS-2 ^d	329	203 (61,70)	219	118 (53,88)	1,379	0,98; 1,95	0,0690

a: Anzahl analysierter Studienteilnehmer
b: Anzahl Studienteilnehmer mit Ereignis
c: OR [95 % KI] und p-Werte INPULSIS-1 und INPULSIS-2 Studien: berechnet mittels logistischer Regression mit Term Behandlung
INPULSIS (gepoolt): OR [95 % KI] und p-Wert berechnet mittels logistischer Regression mit Termen Studien und Behandlung
d: TS, OC
PGI-C-Responder definiert als „sehr viel besser“, „viel besser“, „ein wenig besser“ oder „keine Veränderung“ der subjektiv empfundenen Lebensqualität ermittelt durch den PGI-C-Fragebogen.
Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.5; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.2.5; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.9.1.1

Der Anteil *PGI-C-Responder* an der gesamten Studienpopulation war in der Nintedanib-Gruppe größer als in der Placebo-Gruppe sowohl in den einzelnen Studien INPULSIS-1 (60,84 % vs. 54,90 % Responder) und INPULSIS-2 (61,70 % vs. 53,88 % Responder) als auch in der gepoolten Auswertung dieser beiden Studien (61,29 % vs. 54,37 % Responder) (Tabelle 4-28). Das OR [95 % KI] für den Behandlungseffekt von Nintedanib vs. Placebo beträgt in der INPULSIS-1 Studie 1,276 ([0,89; 1,83]; p = 0,1818), in der INPULSIS-2 Studie 1,379 ([0,98; 1,95]; p = 0,0690) und für die gepoolte Auswertung der INPULSIS Studien 1,328 ([1,04; 1,70]; p = 0,0254). In der gepoolten Analyse der INPULSIS Studien ist dieser Unterschied statistisch signifikant zwischen den Behandlungsgruppen. Der Anteil der Studienteilnehmer, die ihre gesundheitsbezogene Lebensqualität in Woche 52 im Vergleich zur Baseline als „sehr viel besser“, „viel besser“, „ein wenig besser“ oder „ohne Veränderung“ einschätzen, ist demnach unter Nintedanib höher als unter Placebo mit statistisch signifikantem Effekt.

Ergebnisse zum EQ-5D-VAS

Tabelle 4-29: Ergebnisse für EQ-5D-VAS aus RCT mit dem zu bewertenden Arzneimittel

Studie	Behandlung	Baseline		Woche 52		Veränderung	
		N ^a	MW ^b (SD)	N ^a	MW ^b (SD)	N ^a	MWD ^c (SD)
EQ-5D-VAS (Punkte)							
INPULSIS (gepoolt) ^d	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)
INPULSIS-1 ^d	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 ^d	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)

a: Anzahl analysierter Studienteilnehmer
b: nicht adjustierte Mittelwerte
c: nicht adjustierte Mittelwertdifferenz
d: TS; OC
Ein höherer Score steht für eine bessere subjektiv empfundene Lebensqualität.
Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c): Tab. 15.2.2.2.2.6; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d): Tab. 15.2.2.2.2.6; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.11.1.1

Die Veränderung von Baseline bis Woche 52 der EQ-5D-VAS war in den Studien INPULSIS-1 (-2,46 Punkte vs. -5,88 Punkte), INPULSIS-2 (-2,52 Punkte vs. -5,60 Punkte) und in der gepoolten Analyse der beiden Studien (-2,49 Punkte vs. -5,73 Punkte) in der Nintedanib-Gruppe geringer als in der Placebo-Gruppe (Tabelle 4-29). Im Verlauf der Studien war bei beiden Behandlungsgruppen die gesundheitsbezogene Lebensqualität, messbar mittels EQ-5D-VAS, (aufgrund des progressiven Charakters der Erkrankung erklärbar) schlechter als zu Studienbeginn. Jedoch gaben die Nintedanib-behandelten Studienteilnehmer gegenüber den Placebo-behandelten Studienteilnehmern eine geringere Verschlechterung der Lebensqualität an.

Die deskriptiv dargestellten Ergebnisse der EQ-5D-Domänen *Mobilität, Fähigkeit für sich selbst zu sorgen, alltägliche Aktivitäten, Schmerzen/körperliche Beschwerden* und *Angst/Niedergeschlagenheit* befinden sich in Modul 5.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits im Abschnitt 4.3.1.3.3.1 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4 Unerwünschte Ereignisse– RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Meta-Analyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des *Odds Ratios* und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (z. B. Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

4.3.1.3.4.1 Endpunkt Gesamtzahl UE

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-30: Operationalisierung von *Gesamtzahl UE*

Studie	Operationalisierung
INPULSIS 1	Definition: Ein UE war definiert als jegliches unerwünschtes medizinisches Vorkommnis (einschließlich einer Verschlechterung des bestehenden Krankheitszustandes) bei einem Studienteilnehmer, der Studienmedikation erhielt. Dabei musste das Ereignis nicht unbedingt in einem kausalen Zusammenhang mit dem angewandten Arzneimittel stehen. Ereignisse wurden registriert von Studienbeginn bis 28 Tage nach der letzten Arzneimitteleinnahme.
INPULSIS-2	Homogenes Studiendesign zu INPULSIS-1

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-31: Bewertung des Verzerrungspotenzials für den Endpunkt *Gesamtzahl UE* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die *Gesamtzahl UE* erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen hatten. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien aller Endpunkte zu den unerwünschter Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für die *Gesamtzahl UE* transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für den Endpunkt *Gesamtzahl UE* ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt Gesamtzahl unerwünschter Ereignisse für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-32: Ergebnisse für die *Gesamtzahl UE* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
Gesamtzahl UE								
INPULSIS (gepoolt)	638	609 (95,5)	423	379 (89,6)	1,07 [1,03; 1,11]	2,44 [1,50; 3,96]	5,86 [2,53; 9,18]	<0,001
INPULSIS-1	309	298 (96,4)	204	181 (88,7)	1,09 [1,03; 1,15]	3,44 [1,64; 7,23]	7,71 [2,91; 12,52]	<0,001
INPULSIS-2	329	311 (94,5)	219	198 (90,4)	1,05 [0,99; 1,10]	1,83 [0,95; 3,53]	4,12 [-0,49; 8,73]	0,066

a: Anzahl Studienteilnehmer
b: Anzahl Studienteilnehmer mit Ereignis
c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko, eine RD größer null ein höheres Risiko, ein UE zu erleiden unter Nintedanib im Vergleich zu Placebo an.
p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test
Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1, Tab. 23.1.2; Tab. 23.1.3

UE traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 95,5 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 89,6 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen ist statistisch signifikant (OR [95 % KI]: 2,44 [1,50; 3,96]; $p < 0,001$) (Tabelle 4-32).

Allerdings ist bei der Betrachtung jeglicher *UE* zu beachten, dass die klinische Relevanz solcher Ereignisse heterogen sein kann. Minimal beeinträchtigende Ereignisse werden gleichermaßen erfasst wie solche, die die Alltagsgestaltung erheblich kompromittieren können. Diese Heterogenität wird durch den zusammenfassenden Endpunkt der *Gesamtzahl UE* nicht differenziert abgebildet.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.4.1 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4.2 Endpunkt *schwerwiegende UE*

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-33: Operationalisierung von *SUE*

Studie	Operationalisierung
INPULSIS 1	Definition: 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, 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 registriert von Studienbeginn bis 28 Tage nach der letzten Arzneimitteleinnahme.
INPULSIS-2	Homogenes Studiendesign zu INPULSIS-1

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-34: Bewertung des Verzerrungspotenzials für die Endpunkte zu den *SUE* in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die *SUE* erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen hatten. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien aller Endpunkte zu den unerwünschter Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für die *SUE* transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für den Endpunkt *SUE* ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt SUE für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-35: Ergebnisse für die SUE aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
SUE								
INPULSIS (gepoolt)	638	194 (30,4)	423	127 (30,0)	1,01 [0,84; 1,22]	1,02 [0,78; 1,33]	0,38 [-5,26; 6,02]	0,894
INPULSIS-1	309	96 (31,1)	204	55 (27,0)	1,15 [0,87; 1,53]	1,22 [0,83; 1,81]	4,11 [-3,87; 12,09]	0,318
INPULSIS-2	329	98 (29,8)	219	72 (32,9)	0,91 [0,70; 1,17]	0,87 [0,60; 1,25]	-3,09 [-11,03; 4,86]	0,444
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko, eine RD größer null ein höheres Risiko, ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1, Tab. 23.1.2; Tab. 23.1.3</p>								

SUE traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 30,4 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, und somit mit vergleichbarer Häufigkeit wie bei den Placebo-behandelten Studienteilnehmern (30,0 %). Der Unterschied zwischen den Gruppen ist nicht statistisch signifikant (OR [95 % KI]: 1,02 [0,78; 1,33]; p = 0,894) (Tabelle 4-35).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.4.2 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4.3 Endpunkt UE, die zum Therapieabbruch führten

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-36: Operationalisierung von UE, die zum Therapieabbruch führten

Studie	Operationalisierung
INPULSIS 1	Definition: Jegliches berichtete UE, das als Grund für eine definitive Beendigung der Therapie mit der Studienmedikation angegeben wurde (keine Re-Exposition in voller oder reduzierter Dosierung).
INPULSIS-2	Homogenes Studiendesign zu INPULSIS-1

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-37: Bewertung des Verzerrungspotenzials für den Endpunkt UE, die zum Therapieabbruch führten, in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die UE, die zum Therapieabbruch führten, erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen hatten. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien aller Endpunkte zu den unerwünschter Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für den Endpunkt *UE, die zum Therapieabbruch führten*, transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für den Endpunkt *UE, die zum Therapieabbruch führten*, ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt UE, die zum Therapieabbruch führten, für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-38: Ergebnisse für UE, die zum Therapieabbruch führten, aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
UE, die zum Therapieabbruch führten								
INPULSIS (gepoolt)	638	123 (19,3)	423	55 (13,0)	1,48 [1,11; 1,99]	1,60 [1,13; 2,26]	6,28 [1,84; 10,71]	0,007
INPULSIS-1	309	65 (21,0)	204	22 (10,8)	1,95 [1,24; 3,06]	2,20 [1,31; 3,71]	10,25 [4,02; 16,48]	0,002
INPULSIS-2	329	58 (17,6)	219	33 (15,1)	1,17 [0,79; 1,73]	1,21 [0,76; 1,92]	2,56 [-3,72; 8,84]	0,430
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko, eine RD größer null ein höheres Risiko, ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1, Tab. 23.1.2; Tab. 23.1.3</p>								

UE, die zum Therapieabbruch führten, traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 19,3 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 13,0 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen ist statistisch signifikant (OR [95 % KI]: 1,60 [1,13; 2,26]; p = 0,007). Somit brachen in der Nintedanib-Gruppe häufiger Studienteilnehmer die Behandlung aufgrund eines UE ab als in der Placebo-Gruppe (Tabelle 4-38).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.4.3 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4.4 Endpunkt UE, die zum Tod führten

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-39: Operationalisierung von UE, die zum Tod führten

Studie	Operationalisierung
INPULSIS 1	Definition: Jegliches berichtete UE, das zum Tod führte. Als solches wurde ein Ereignis 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.
INPULSIS-2	Homogenes Studiendesign zu INPULSIS-1

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-40: Bewertung des Verzerrungspotenzials für den Endpunkt UE, die zum Tod führten, in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die UE, die zum Tod führten, erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen hatten. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien aller Endpunkte zu den unerwünschter Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für den Endpunkt *UE, die zum Tod führten*, transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für den Endpunkt *UE, die zum Tod führten*, ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt UE, die zum Tod führten, für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-41: Ergebnisse für *UE, die zum Tod führten*, aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
UE, die zum Tod führten								
INPULSIS (gepoolt)	638	37 (5,8)	423	31 (7,3)	0,79 [0,50; 1,25]	0,78 [0,48; 1,28]	-1,53 [-4,60; 1,55]	0,319
INPULSIS-1	309	12 (3,9)	204	10 (4,9)	0,79 [0,35; 1,80]	0,78 [0,33; 1,85]	-1,02 [-4,68; 2,64]	0,577
INPULSIS-2	329	25 (7,6)	219	21 (9,6)	0,79 [0,46; 1,38]	0,78 [0,42; 1,42]	-1,99 [-6,83; 2,85]	0,411

a: Anzahl Studienteilnehmer
b: Anzahl Studienteilnehmer mit Ereignis
c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko, eine RD größer null ein höheres Risiko, ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an.
p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test
Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1, Tab. 23.1.2; Tab. 23.1.3

Ein *UE, das zum Tod führte*, trat in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 5,8 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 7,3 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen war nicht statistisch signifikant (OR [95 % KI]: 0,78 [0,48; 1,28]; p = 0,319). Somit verstarb in der Nintedanib-Gruppe numerisch seltener ein Patient infolge eines UE als in der Placebo-Gruppe (Tabelle 4-41).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.4.4 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4.5 Endpunkt UE von besonderem Interesse

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-42: Operationalisierung von UE von besonderem Interesse

Studie	Operationalisierung
INPULSIS-1	Definition: Als UE von besonderem Interesse wurden gastrointestinale Störungen (Diarrhoe, Übelkeit, Erbrechen, Appetitlosigkeit, Gewichtsverlust, gastrointestinale Perforation), Leber- und Gallenerkrankungen (Leberenzymhöhung, Bilirubinerhöhung), kardiale Ereignisse, (darunter Myokardinfarkt), arterielle thromboembolische Ereignisse, respiratorische Ereignisse (Progression der IPF, Husten, Dyspnoe, Bronchitis, Nasopharyngitis), alle weiteren Störungen, die in mindestens einem Studienarm mit einer Häufigkeit von > 10 % der Studienteilnehmer auftraten, sowie Blutungsereignisse betrachtet. Ereignisse wurden registriert von Studienbeginn bis 28 Tage nach der letzten Arzneimittelleinnahme.
INPULSIS-2	Homogenes Studiendesign zu INPULSIS-1

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-43: Bewertung des Verzerrungspotenzials für den Endpunkt UE von besonderem Interesse in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
INPULSIS-1 ^a	niedrig	ja	ja	ja	ja	niedrig
INPULSIS-2 ^a	niedrig	ja	ja	ja	ja	niedrig

a: Das Verzerrungspotenzial der gepoolten Auswertung der INPULSIS Studien ergibt sich aus dem Verzerrungspotenzial der Einzelstudien.

Begründen Sie für jede Studie die abschließende Einschätzung.

- Verblindung Endpunkterheber

Die Datenerhebung für die UE von besonderem Interesse erfolgte doppelblind.

- Adäquate Umsetzung des ITT-Prinzips

Die Auswertung der Studien INPULSIS-1 und INPULSIS-2 erfolgte auf Basis des TS, definiert als randomisierte Studienteilnehmer, die mindestens eine Dosis der Studienmedikation eingenommen hatten. Das TS wick lediglich um zwei Studienteilnehmer in der INPULSIS-1 Studie und drei Studienteilnehmer in der INPULSIS-2 Studie, die keine Behandlung erhielten, von allen randomisierten Studienteilnehmern ab. Damit ist die Abweichung (< 1 % der randomisierten Population) als nicht ergebnisrelevant einzuschätzen. Die Auswertung erfolgte in der durch Randomisierung zugeordneten Behandlungsgruppe und erfüllte daher das ITT-Prinzip.

- Ergebnisunabhängige Berichterstattung

Die Auswertungskriterien aller Endpunkte zu den unerwünschter Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden.

- Sonstige Aspekte

Die Beschreibung der Methodik und Ergebnisse von Analysen im Studienbericht, statistischen Analyseplan und Studienprotokoll war für den Endpunkt *UE von besonderem Interesse* transparent.

- Verzerrungspotenzial auf Endpunktebene

Abschließend ergibt sich für den Endpunkt *UE von besonderem Interesse* ein niedriges Verzerrungspotenzial.

Stellen Sie die Ergebnisse für den Endpunkt UE von besonderem Interesse für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Als *UE von besonderem Interesse* wurden *gastrointestinale Störungen (Diarrhoe, Übelkeit, Erbrechen, Appetitlosigkeit, Gewichtsverlust, gastrointestinale Perforation), Leber- und Gallenerkrankungen (Leberenzymerrhöhung, Bilirubinerhöhung), kardiale Ereignisse (darunter Myokardinfarkt), respiratorische Ereignisse (Progression der IPF, Husten, Dyspnoe, Bronchitis, Nasopharyngitis, alle weiteren Störungen, die in mindestens einem Studienarm mit einer Häufigkeit von > 10 % der Studienteilnehmer auftraten, sowie Blutungsereignisse betrachtet.*

Unerwünschte gastrointestinale EreignisseTabelle 4-44: Ergebnisse für *UE – gastrointestinale Ereignisse* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
INPULSIS (gepoolt)	Diarrhoe (Gesamtzahl)							
	638	398 (62,4)	423	78 (18,4)	3,38 [2,74; 4,17]	7,33 [5,47; 9,84]	43,94 [38,67; 49,21]	< 0,001
	Diarrhoe (Intensität leicht ^d)							
	638	226 (35,4)	423	60 (14,2)	2,50 [1,93; 3,23]	3,32 [2,42; 4,56]	21,24 [16,26; 26,22]	< 0,001
	Diarrhoe (Intensität moderat ^d)							
	638	150 (23,5)	423	16 (3,8)	6,22 [3,77; 10,3]	7,82 [4,59; 13,3]	19,73 [15,97; 23,49]	< 0,001
	Diarrhoe (Intensität schwer ^d)							
	638	21 (3,3)	423	2 (0,5)	6,96 [1,64; 29,5]	7,16 [1,67; 30,7]	2,82 [1,29; 4,35]	0,002
	Diarrhoe (angegeben als SUE)							
	638	2 (0,3)	423	1 (0,2)	1,33 [0,12; 14,6]	1,33 [0,12; 14,7]	0,08 [-0,56; 0,71]	0,817
	Diarrhoe (angegeben als UE, das zum Therapieabbruch führte)							
	638	28 (4,4)	423	1 (0,2)	18,6 [2,54; 136]	19,4 [2,63; 143]	4,15 [2,50; 5,81]	< 0,001
	Übelkeit							
	638	156 (24,5)	423	28 (6,6)	3,69 [2,52; 5,42]	4,57 [2,99; 6,97]	17,83 [13,74; 21,92]	< 0,001
	Erbrechen							
	638	74 (11,6)	423	11 (2,6)	4,46 [2,40; 8,30]	4,91 [2,58; 9,38]	9,00 [6,09; 11,91]	< 0,001
	Appetitlosigkeit							
638	68 (10,7)	423	24 (5,7)	1,88 [1,20; 2,94]	1,98 [1,22; 3,21]	4,98 [1,73; 8,24]	0,005	
Gewichtsverlust								
638	62 (9,7)	423	15 (3,5)	2,74 [1,58; 4,75]	2,93 [1,64; 5,22]	6,17 [3,28; 9,07]	< 0,001	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
	Gastrointestinale Perforation							
	638	2 (0,3)	423	0 (0)	n.b. ^e	n.b. ^e	0,31 [-0,12; 0,75]	0,2491
INPULSIS-1	Diarrhoe (Gesamtzahl)							
	309	190 (61,5)	204	38 (18,6)	3,30 [2,45; 4,46]	6,97 [4,58; 10,6]	42,86 [35,25; 50,48]	< 0,001
	Diarrhoe (Intensität leicht^d)							
	309	103 (33,3)	204	29 (14,2)	2,34 [1,62; 3,40]	3,02 [1,91; 4,77]	19,12 [12,00; 26,23]	< 0,001
	Diarrhoe (Intensität moderat^d)							
	309	75 (24,3)	204	9 (4,4)	5,50 [2,82; 10,70]	6,94 [3,39; 14,20]	19,86 [14,31; 25,41]	< 0,001
	Diarrhoe (Intensität schwer^d)							
	309	11 (3,6)	204	0 (0,0)	n.b. ^e	n.b. ^e	3,56 [1,49; 5,63]	0,006
	Diarrhoe (angegeben als SUE)							
	309	1 (0,3)	204	0 (0,0)	n.b. ^e	n.b. ^e	0,32 [-0,31; 0,96]	0,416
	Diarrhoe (angegeben als UE, das zum Therapieabbruch führte)							
	309	14 (4,5)	204	0 (0,0)	n.b. ^e	n.b. ^e	4,53 [2,21; 6,85]	0,002
	Übelkeit							
	309	70 (22,7)	204	12 (5,9)	3,85 [2,14; 6,92]	4,69 [2,47; 8,90]	16,77 [11,10; 22,45]	< 0,001
	Erbrechen							
	309	40 (12,9)	204	4 (2,0)	6,60 [2,40; 18,2]	7,43 [2,62; 21,1]	10,98 [6,79; 15,18]	< 0,001
	Appetitlosigkeit							
309	26 (8,4)	204	14 (6,9)	1,23 [0,66; 2,29]	1,25 [0,63; 2,45]	1,55 [-3,10; 6,20]	0,521	
Gewichtsverlust								
309	25 (8,1)	204	13 (6,4)	1,27 [0,67; 2,42]	1,29 [0,65; 2,59]	1,72 [-2,81; 6,24]	0,467	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
INPULSIS-2	Diarrhoe (Gesamtzahl)							
	329	208 (63,2)	219	40 (18,3)	3,46 [2,58; 4,64]	7,69 [5,11; 11,60]	44,96 [37,65; 52,26]	< 0,001
	Diarrhoe (Intensität leicht ^d)							
	329	123 (37,4)	219	31 (14,2)	2,64 [1,85; 3,77]	3,62 [2,33; 5,63]	23,23 [16,26; 30,21]	< 0,001
	Diarrhoe (Intensität moderat ^d)							
	329	75 (22,8)	219	7 (3,2)	7,13 [3,35; 15,20]	8,94 [4,04; 19,80]	19,60 [14,50; 24,70]	< 0,001
	Diarrhoe (Intensität schwer ^d)							
	329	10 (3,0)	219	2 (0,9)	3,33 [0,74; 15,00]	3,40 [0,74; 15,70]	2,13 [-0,12; 4,37]	0,096
	Diarrhoe (angegeben als SUE)							
	329	1 (0,3)	219	1 (0,5)	0,67 [0,04; 10,60]	0,66 [0,04; 10,70]	-0,15 [-1,23; 0,92]	0,772
	Diarrhoe (angegeben als UE, das zum Therapieabbruch führte)							
	329	14 (4,3)	219	1 (0,5)	9,32 [1,23; 70,40]	9,69 [1,26; 74,20]	3,80 [1,44; 6,16]	0,008
	Übelkeit							
	329	86 (26,1)	219	16 (7,3)	3,58 [2,16; 5,93]	4,49 [2,55; 7,90]	18,83 [12,97; 24,70]	< 0,001
	Erbrechen							
	329	34 (10,3)	219	7 (3,2)	3,23 [1,46; 7,16]	3,49 [1,52; 8,02]	7,14 [3,11; 11,17]	0,002
	Appetitlosigkeit							
329	42 (12,8)	219	10 (4,6)	2,80 [1,43; 5,45]	3,06 [1,50; 6,24]	8,20 [3,66; 12,74]	0,001	
Gewichtsverlust								
329	37 (11,2)	219	2 (0,9)	12,3 [3,00; 50,6]	13,7 [3,28; 57,70]	10,33 [6,69; 13,97]	< 0,001	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. d: Für 1 Ereignis in der Nintedanib-Gruppe lag keine Bewertung der Intensität vor. Intensität leicht: Anzeichen oder Symptome sind spürbar, aber leicht zu tolerieren. Intensität moderat: Anzeichen oder Symptome führen zu Beeinträchtigungen im Alltag, ziehen aber keine Arbeitsunfähigkeit oder Einschränkung von Alltagsaktivitäten nach sich. Intensität schwer: Beeinträchtigungen führen zu Ermüdung und Arbeitsunfähigkeit bzw. Einschränkung der täglichen Aktivitäten. e: n.b.= nicht berechenbar p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1; Tab. 23.1.2; Tab. 23.1.3, Tab. Additional AEs</p>								

Unter den *gastrointestinalen Störungen* waren die Ereignisse *Diarrhoe, Übelkeit, Erbrechen, Appetitlosigkeit, Gewichtsverlust* und *gastrointestinale Perforation* von besonderem Interesse. *Diarrhoe* kam in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien am häufigsten vor (62,4% in der Nintedanib-Gruppe vs. 18,4% in der Placebo-Gruppe). Der Unterschied ist statistisch signifikant (OR [95 % KI]: 7,33 [5,47; 9,84]; $p < 0,001$) (Tabelle 4-44).

Diarrhoen traten überwiegend in den ersten drei Monaten der Behandlung auf und waren bei den meisten Studienteilnehmern von leichter bis moderater Intensität. Zu einer Einschränkung in der Ausübung von Alltagsaktivitäten oder zu Arbeitsunfähigkeit (Intensität „schwer“) kam es bei 5 % der Nintedanib-Studienteilnehmer mit *Diarrhoe* (entsprechend 3,3 % aller Studienteilnehmer der Nintedanib-Gruppe). Das Studienprotokoll der INPULSIS Studien gab Empfehlungen zum Management von *Diarrhoen*, die eine pharmakologische Intervention (z. B. mit dem Antidiarrhoikum Loperamid) und bei Persistenz der Beschwerden eine Dosisreduktion oder Unterbrechung der Behandlung vorsahen. Nach Abklingen der Beschwerden konnte die Arzneimittelgabe in normaler oder reduzierter Dosis wieder aufgenommen werden. Auf diese Weise ließ sich das UE *Diarrhoe* bei den meisten Studienteilnehmer der Nintedanib-Gruppe gut tolerieren: 78,6 % der Studienteilnehmer mit *Diarrhoe* konnten die Therapie mit der Normaldosis fortsetzen, weitere 14,3 % der Studienteilnehmer mit reduzierter Dosis. Lediglich bei 7,0 % der Studienteilnehmer mit *Diarrhoe* (entsprechend 4,4 % aller Studienteilnehmer der Nintedanib-Gruppe) kam es aufgrund dieses UE zu einem endgültigen Therapieabbruch (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014b). Bei den SUE gibt es hinsichtlich der *Diarrhoen* keinen statistisch signifikanten Unterschied zwischen den Untersuchungsarmen Nintedanib und Placebo.

Übelkeit wurde in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien im Nintedanib-Behandlungsarm (24,5 %) häufiger dokumentiert als im Placebo-Behandlungsarm (6,6 %). Der Unterschied ist statistisch signifikant: (OR [95 % KI]: 4,57 [2,99; 6,97]; $p < 0,001$). Auch *Erbrechen* wurde im Nintedanib-Behandlungsarm (11,6 %) häufiger berichtet als im Placebo-Behandlungsarm (2,6 %). Der beobachtete Effekt ist statistisch signifikant (OR [95 % KI]: 4,91 [2,58; 9,38]; $p < 0,001$) (Tabelle 4-44).

Die Fachinformation für das Produkt Ofev[®] gibt Empfehlungen zu Maßnahmen beim Auftreten von *Diarrhoen*, *Übelkeit* und *Erbrechen*, die auf dem Vorgehen in den beiden Phase III-Studien beruhen (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Appetitlosigkeit und *Gewichtsverlust* werden gemäß den beim Erfassen von UE in den INPULSIS Studien zugrunde gelegten „preferred terms“ des Medical Dictionary for Regulatory Activities (MedDRA) formal nicht der Kategorie der gastrointestinalen Ereignisse zugerechnet, könnten jedoch durch gastrointestinale Beschwerden beeinflusst sein und werden daher hier dargestellt. *Appetitlosigkeit* wurde in den INPULSIS Studien im Nintedanib-Behandlungsarm (10,7 %) häufiger dokumentiert als im Placebo-Behandlungsarm (5,7 %). Der Unterschied ist statistisch signifikant: (OR [95 % KI]: 1,98 [1,22; 3,21]; $p = 0,005$). *Gewichtsverlust* trat im Nintedanib-Behandlungsarm (9,7 %) häufiger auf als im Placebo-Behandlungsarm (3,5 %). Der beobachtete Effekt ist statistisch signifikant (OR [95 % KI]: 2,93 [1,64; 5,22]; $p < 0,001$).

Bedingt durch den Wirkmechanismus von Nintedanib könnten Patienten ein erhöhtes Risiko für eine *gastrointestinale Perforation* haben. In den INPULSIS Studien traten *gastrointestinale Perforationen* so selten auf (0 % in der Placebogruppe versus 0,3 % in der Nintedanib-Gruppe), dass keine Aussage zum Risiko einer *gastrointestinalen Perforation* unter einer Behandlung mit Nintedanib getroffen werden kann. Es wurde ein entsprechender Warnhinweis in die Fachinformation von Ofev[®] aufgenommen (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Zusammenfassend traten die *gastrointestinalen UE Diarrhoe*, *Übelkeit* und *Erbrechen* sowie die weiteren *UE Appetitlosigkeit* und *Gewichtsverlust* unter der Behandlung mit Nintedanib signifikant häufiger auf als in der Placebo-Gruppe. Häufigstes gastrointestinales Ereignis war die *Diarrhoe*. Die Mehrzahl der *Diarrhoen* waren von leichter bis moderater Intensität und ließen sich in der Regel mithilfe der Empfehlungen im Prüfplan zum Vorgehen beim Auftreten von *Diarrhoe* kontrollieren.

Unerwünschte Leberenzym- und Bilirubinveränderungen

Tabelle 4-45: Ergebnisse für UE – Leber- und Gallenerkrankungen (Leberenzym-, Bilirubin-, ALKP-Erhöhung) aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
INPULSIS (gepoolt)*	Leberenzymerrhöhung (AST, ALT)							
	≥ 3x ULN ^{d,e}							
	638	32 (5,0)	423	3 (0,7)	7,07 [2,18; 22,9]	7,39 [2,25; 24,3]	4,31 [2,43; 6,18]	< 0,001
	≥ 5x ULN ^{d,e}							
	638	14 (2,2)	423	1 (0,2)	9,28 [1,23; 70,3]	9,47 [1,24; 72,3]	1,96 [0,73; 3,19]	0,008
	≥ 8x ULN ^d							
	638	5 (0,8)	423	1 (0,2)	3,32 [0,39; 28,3]	3,33 [0,39; 28,6]	0,55 [-0,28; 1,37]	0,244
	Bilirubinerhöhung							
	≥ 1,5x ULN ^{d,e}							
	638	15 (2,4)	423	3 (0,7)	3,32 [0,97; 11,4]	3,37 [0,97; 11,7]	1,64 [0,22; 3,06]	0,043
	≥ 2x ULN ^d							
	638	3 (0,5)	423	2 (0,5)	0,99 [0,17; 5,93]	0,99 [0,17; 5,98]	-0,00 [-0,84; 0,84]	0,995
	ALKP-Erhöhung							
	≥ 1,5x ULN ^{d,e}							
	638	37 (5,8)	423	4 (0,9)	6,13 [2,20; 17,1]	6,45 [2,28; 18,2]	4,85 [2,82; 6,89]	< 0,001
≥ 2x ULN ^d								
638	17 (2,7)	423	1 (0,2)	11,3 [1,51; 84,4]	11,6 [1,53; 87,1]	2,43 [1,10; 3,76]	0,003	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p-Wert
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. d: ULN: Upper Limit of Normal (oberer Normgrenzwert). e: Kategorien sind kumulativ, d.h. Studienteilnehmer, die in die höheren Kategorien fallen, sind bei den niedrigeren Kategorien mit angegeben. p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test * Auf Grund zu geringer Zellenbesetzung werden die Ergebnisse auf Einzelstudienoberfläche im Dossier nicht dargestellt, sondern lediglich in der Anlage für Modul 5 geliefert. Grundsätzlich zeigen beide Studien ähnliche Ereignisraten in den jeweiligen Behandlungsarmen Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1</p>								

UE, welche das Leber- und Gallenwegssystem betrafen, wurden in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien häufiger unter Nintedanib als unter Placebo dokumentiert: Die Behandlung mit Nintedanib war im Vergleich zu Placebo häufiger verbunden mit einer Erhöhung der Leberenzyme (ALT, AST, ALKP) oder des Bilirubinwertes (Tabelle 4-45).

Erhöhungen der AST- und ALT-Werte um mindestens 3x ULN wurden bei 5 % der Nintedanib-behandelten Studienteilnehmer (gegenüber 0,7 % in der Placebo-Gruppe) registriert (OR [95 % KI]: 7,39 [2,25; 24,3]; p < 0,001). Bei 2,2 % (versus 0,2 %) erreichte die Erhöhung mindestens 5x ULN (OR [95 % KI]: 9,47 [1,24; 72,3]; p = 0,008). Eine Erhöhung um mindestens 8x ULN trat unter Nintedanib (0,8 %) jedoch nicht signifikant häufiger auf als in der Placebo-Gruppe (0,2 %) (OR [95 % KI]: 3,33 [0,39; 28,6]; p = 0,244).

Eine Erhöhung des Bilirubinwertes um mindestens 1,5x ULN fand sich in der gepoolten Analyse der INPULSIS Studien bei 2,4 % der Nintedanib-Studienteilnehmer (verglichen mit 0,7 % der Placebo-Studienteilnehmer) (OR [95 % KI]: 3,37 [0,97; 11,7]; p = 0,043). Eine Erhöhung des Bilirubinwertes um mindestens 2x ULN trat in der Nintedanib-Gruppe jedoch nicht häufiger auf als in der Placebo-Gruppe. Eine Erhöhung der ALKP um mindestens 1,5x ULN wurde in der gepoolten Analyse der INPULSIS Studien bei 5,8 % der Nintedanib-Studienteilnehmer (verglichen mit 0,9 % der Placebo-Studienteilnehmer) dokumentiert (OR [95 % KI]: 6,45 [2,28; 18,2]; p < 0,001). Eine Erhöhung der ALKP um mindestens 2,0x ULN wurde in der Nintedanib-Gruppe (2,7 %) ebenfalls häufiger registriert als in der Placebo-Gruppe (0,2 %) (OR [95 % KI]: 11,6 [1,53; 87,1]; p = 0,003).

UE aus der Gruppe der *Leber- und Gallenerkrankungen* waren überwiegend nicht-schwerwiegend. Als *SUE* wurden *Leber- und Gallenerkrankungen* bei sieben Studienteilnehmern der Nintedanib-Gruppe (1,1 % aller Studienteilnehmer) angegeben. Das Studienprotokoll der INPULSIS Studien legte ein standardisiertes Vorgehen bei einer Erhöhung von Leberenzymen fest, welches sich an der Höhe des Transaminasen-Anstiegs in Relation zum ULN sowie dem Vorhandensein von Zeichen einer schweren Leberschädigung ausrichtete. Unter Beachtung dieser Vorgaben führte ein *UE* aus der Gruppe der *Leber- und Gallenerkrankungen* bei sieben Studienteilnehmern der Nintedanib-Gruppe (1,1 % aller Studienteilnehmer) zu einem vorzeitigen Therapieabbruch. Zu einer dauerhaften Dosisreduktion auf 100 mg 2 x täglich kam es aufgrund eines *UE* aus der Gruppe der *Leber- und Gallenerkrankungen* bei weiteren sechs Studienteilnehmern (0,9 % aller Studienteilnehmer).

Die Fachinformation für das Produkt Ofev[®] gibt Empfehlungen zu Maßnahmen bei Leberenzym erhöhungen und zum Monitoring der Leberenzymwerte und enthält darüber hinaus einen Warnhinweis, wonach die Behandlung mit Nintedanib bei Studienteilnehmern mit bekannten mittelschweren oder schweren Leberfunktionsstörungen nicht empfohlen wird (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Unerwünschte kardiale EreignisseTabelle 4-46: Ergebnisse für *UE - kardiale Ereignisse* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p- Wert
INPULSIS (gepoolt)	Unerwünschte kardiale Ereignisse							
	638	64 (10,0)	423	45 (10,6)	0,94 [0,66; 1,35]	0,94 [0,63; 1,40]	-0,61 [-4,36; 3,14]	0,750
	Schwerwiegende unerwünschte kardiale Ereignisse							
	638	32 (5,0)	423	23 (5,4)	0,92 [0,55; 1,55]	0,92 [0,53; 1,59]	-0,42 [-3,17; 2,32]	0,762
	Tödliche kardiale Ereignisse							
	638	3 (0,5)	423	6 (1,4)	0,33 [0,08; 1,32]	0,33 [0,08; 1,32]	-0,95 [-2,19; 0,30]	0,099
	Arterielle thromboembolische Ereignisse							
	638	16 (2,5)	423	3 (0,7)	3,54 [1,04; 12,06]	3,60 [1,04; 12,44]	1,8 [0,35; 3,25]	0,0305
Myokardinfarkt / akuter Myokardinfarkt								
638	10 (1,6)	423	2 (0,5)	3,32 [0,73; 15,05]	3,35 [0,73; 15,37]	1,09 [-0,07; 2,26]	0,0988	
INPULSIS-1	Unerwünschte kardiale Ereignisse							
	309	30 (9,7)	204	19 (9,3)	1,04 [0,60; 1,80]	1,05 [0,57; 1,92]	0,40 [-4,78; 5,57]	0,882
	Schwerwiegende unerwünschte kardiale Ereignisse							
	309	14 (4,5)	204	11 (5,4)	0,84 [0,39; 1,81]	0,83 [0,37; 1,87]	-0,86 [-4,73; 3,01]	0,657
	Tödliche kardiale Ereignisse							
	309	1 (0,3)	204	2 (1,0)	0,33 [0,03; 3,62]	0,33 [0,03; 3,64]	-0,66 [-2,15; 0,84]	0,340
	Arterielle thromboembolische Ereignisse							
	309	8 (2,6)	204	1 (0,5)	5,25 [0,67; 41,91]	5,40 [0,67; 43,47]	2,1 [0,09; 4,11]	0,0764
Myokardinfarkt/akuter Myokardinfarkt								
309	5 (1,6)	204	1 (0,5)	3,30 [0,39; 28,05]	3,34 [0,39; 28,79]	1,13 [-0,57; 2,83]	0,2449	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR [95 % KI]	RD [95 % KI]	p- Wert
INPULSIS-2	Unerwünschte kardiale Ereignisse							
	329	34 (10,3)	219	26 (11,9)	0,87 [0,54; 1,41]	0,86 [0,50; 1,47]	-1,54 [-6,94; 3,86]	0,572
	Schwerwiegende unerwünschte kardiale Ereignisse							
	329	18 (5,5)	219	12 (5,5)	1,00 [0,49; 2,03]	1,00 [0,47; 2,12]	-0,01 [-3,90; 3,88]	0,997
	Tödliche kardiale Ereignisse							
	329	2 (0,6)	219	4 (1,8)	0,33 [0,06; 1,80]	0,33 [0,06; 1,81]	-1,22 [-3,18; 0,74]	0,179
	Arterielle thromboembolische Ereignisse							
	329	8 (2,4)	219	2 (0,9)	2,66 [0,57; 12,42]	2,70 [0,57; 12,86]	1,52 [-0,57; 3,61]	0,1933
Myokardinfarkt/akuter Myokardinfarkt								
329	5 (1,5)	219	1 (0,5)	3,33 [0,39; 28,29]	3,36 [0,39; 28,99]	1,06 [-0,53; 2,66]	0,2414	

a: Anzahl Studienteilnehmer

b: Anzahl Studienteilnehmer mit Ereignis

c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an.

p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test

Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1; Tab. 23.1.2; Tab. 23.1.3; Tab. Additional AEs

Bedingt durch den Wirkmechanismus von Nintedanib besteht die Möglichkeit, dass Patienten ein erhöhtes Risiko für *thromboembolische Ereignisse* haben. *Arterielle thromboembolische Ereignisse* wurden selten berichtet. In der gepoolten Analyse der INPULSIS-Studien bei 0,7 % der Patienten in der Placebogruppe und bei 2,5 % in der mit Nintedanib behandelten Gruppe (OR [95 % KI]: 3,60 [1,04; 12,44]; p = 0,0305) (Tabelle 4-46). *Unerwünschte Ereignisse*, die auf eine *ischämische Herzerkrankung* deuten, waren zwar zwischen der Nintedanib- und der Placebogruppe ausgeglichen, aber in der Nintedanib-Gruppe erlitt ein numerisch höherer Prozentsatz der Patienten (1,6 %) einen *Myokardinfarkt* als dies in der Placebogruppe (0,5 %) der Fall war (OR [95 % KI]: 3,35 [0,73; 15,37]; p = 0,099). Es wurde ein entsprechender Warnhinweis in die Fachinformation von Ofev[®] aufgenommen (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Bei Betrachtung aller *unerwünschten kardialen Ereignisse* wurden diese in den gepoolten INPULSIS Studien vergleichbar häufig unter Nintedanib und Placebo (10,0 % vs. 10,6 %) beobachtet. Außer den bereits oben genannten *Herzinfarkten* waren hierunter beispielsweise folgende Ereignisse subsummiert: Vorhofflimmern, koronare Herzerkrankung, Herzversagen, EKG-Auffälligkeiten wie Arrhythmien oder Schenkelblockbilder oder Cor pulmonale. Auch *schwerwiegende unerwünschte kardiale Ereignisse* traten in der Nintedanib-Gruppe (5,0 %) und Placebo-Gruppe (5,4 %) vergleichbar häufig auf. *Tödliche kardiale Ereignisse* traten numerisch seltener in der Nintedanib-Gruppe (0,5 %) als in der Placebo-Gruppe (1,4 %) auf (OR [95 % KI]: 0,33 [0,08; 1,32]; p = 0,099).

Unerwünschte respiratorische EreignisseTabelle 4-47: Ergebnisse für *UE - respiratorische Ereignisse* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR (95 % KI)	RD (95 % KI)	p-Wert
INPULSIS (gepoolt)	Progression der IPF							
	638	64 (10,0)	423	61 (14,4)	0,70 [0,50; 0,97]	0,66 [0,45; 0,96]	-4,39 [-8,47; -0,31]	0,030
	Husten							
	638	85 (13,3)	423	57 (13,5)	0,99 [0,72; 1,35]	0,99 [0,69; 1,42]	-0,15 [-4,34; 4,04]	0,943
	Dyspnoe							
	638	49 (7,7)	423	48 (11,3)	0,68 [0,46; 0,99]	0,65 [0,43; 0,99]	-3,67 [-7,33; -0,01]	0,042
	Bronchitis							
	638	67 (10,5)	423	45 (10,6)	0,99 [0,69; 1,41]	0,99 [0,66; 1,47]	-0,14 [-3,92; 3,64]	0,943
Nasopharyngitis								
638	87 (13,6)	423	68 (16,1)	0,85 [0,63; 1,14]	0,82 [0,58; 1,16]	-2,44 [-6,84; 1,96]	0,271	
INPULSIS-1	Progression der IPF							
	309	31 (10,0)	204	21 (10,3)	0,97 [0,58; 1,65]	0,97 [0,54; 1,74]	-0,26 [-5,61; 5,09]	0,923
	Husten							
	309	47 (15,2)	204	26 (12,7)	1,19 [0,76; 1,86]	1,23 [0,73; 2,06]	2,47 [-3,62; 8,55]	0,434
	Dyspnoe							
	309	22 (7,1)	204	23 (11,3)	0,63 [0,36; 1,10]	0,60 [0,33; 1,11]	-4,15 [-9,36; 1,05]	0,104
	Bronchitis							
	309	36 (11,7)	204	28 (13,7)	0,85 [0,54; 1,35]	0,83 [0,49; 1,41]	-2,08 [-8,00; 3,85]	0,486
Nasopharyngitis								
309	39 (12,6)	204	34 (16,7)	0,76 [0,50; 1,16]	0,72 [0,44; 1,19]	-4,05 [-10,36; 2,27]	0,199	

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR (95 % KI)	RD (95 % KI)	p-Wert
INPULSIS-2	Progression der IPF							
	329	33 (10,0)	219	40 (18,3)	0,55 [0,36; 0,84]	0,50 [0,30; 0,82]	-8,23 [-14,29; -2,17]	0,005
	Husten							
	329	38 (11,6)	219	31 (14,2)	0,82 [0,52; 1,27]	0,79 [0,48; 1,32]	-2,61 [-8,37; 3,16]	0,368
	Dyspnoe							
	329	27 (8,2)	219	25 (11,4)	0,72 [0,43; 1,20]	0,69 [0,39; 1,23]	-3,21 [-8,36; 1,94]	0,209
	Bronchitis							
	329	31 (9,4)	219	17 (7,8)	1,21 [0,69; 2,14]	1,24 [0,67; 2,29]	1,66 [-3,09; 6,41]	0,501
Nasopharyngitis								
329	48 (14,6)	219	34 (15,5)	0,94 [0,63; 1,41]	0,93 [0,58; 1,50]	-0,94 [-7,06; 5,19]	0,764	
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 23.1.1; Tab. 23.1.2; Tab. 23.1.3</p>								

Husten und Dyspnoe sind typische Symptome der Grunderkrankung IPF. Daher wurden Therapieeffekte auf die Beeinträchtigung der Lebensqualität durch diese Symptome mittels der spezifischen Fragebögen *CASA-Q* und *SOBQ* analysiert. Diese Messinstrumente erfragen explizit die durchschnittliche Beeinträchtigung der Lebensqualität durch die genannten Symptome über die Zeitspanne der letzten sieben Tage vor der jeweiligen Studienvsitede und bilden somit chronische Effekte der Therapie auf die Symptomintensität ab.

Husten und Dyspnoe können jedoch auch als einzeln berichtenswerte akute medizinische Vorkommnisse auftreten, die gegenüber der vom Studienteilnehmer als normal empfundenen Tagesvariabilität hervorstechen. Dies kann zudem auch außerhalb der Zeitspanne der letzten sieben Tage vor einer Studienvsitede der Fall gewesen sein. Daher lassen sich aus der Registrierung der Symptome *Husten* oder *Dyspnoe* als *UE* gegenüber den Fragebogeninstrumenten zusätzliche patientenrelevante Informationen gewinnen.

Als *respiratorische Ereignisse* wurden eine *Progression der IPF*, *Husten*, *Dyspnoe*, *Bronchitis* und *Nasopharyngitis* im Rahmen der Erfassung von *UE* angegeben (Tabelle 4-47).

Eine *Progression der IPF* wurde in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien in der Nintedanib-Gruppe von 10,0 % der Studienteilnehmer als *UE* angegeben, verglichen mit 14,4 % in der Placebo-Gruppe. Der Unterschied ist statistisch signifikant zugunsten von Nintedanib (OR [95 % KI]: 0,66 [0,45; 0,96]; $p = 0,030$). *Husten* als *UE* trat in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig auf (13,3 % vs. 13,5 %) (OR [95 % KI]: 0,99 ([0,72; 1,35]; $p = 0,943$). *Dyspnoe* als *UE* wurde in den gepoolten INPULSIS Studien signifikant seltener bei Nintedanib-behandelten Studienteilnehmern (7,7 %) als bei Placebo-behandelten Studienteilnehmern (11,3 %) erfasst (OR [95 % KI]: 0,65; [0,43; 0,99]; $p = 0,042$). *Bronchitis* und *Nasopharyngitis* als *UE* fanden sich in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig: *Bronchitis* in der Nintedanib-Gruppe bei 10,5 % der Studienteilnehmer vs. 10,6 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,99 [0,66; 1,47]; $p = 0,943$). *Nasopharyngitis* in der Nintedanib-Gruppe bei 13,6 % der Studienteilnehmer vs. 16,1 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,82 [0,58; 1,16]; $p = 0,271$).

Unerwünschte Blutungsereignisse

Aufgrund theoretischer – den Wirkmechanismus betreffender – Erwägungen ist eine Hemmung des vaskulären endothelialen Wachstumsfaktorrezeptors (VEGFR) möglicherweise mit einem erhöhten *Blutungsrisiko* verbunden (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b). Daher wurden Blutungsereignisse als UE von besonderem Interesse betrachtet.

Tabelle 4-48: Ergebnisse für *UE - Blutungsereignisse* aus RCT mit dem zu bewertenden Arzneimittel

Studie	Nintedanib		Placebo		Behandlungseffekt Nintedanib vs. Placebo ^c			
	N ^a	n ^b (%)	N ^a	n ^b (%)	RR [95 % KI]	OR (95 % KI)	RD (95 % KI)	p-Wert
Blutungsereignisse								
INPULSIS (gepoolt)	638	66 (10,3)	423	33 (7,8)	1,33 [0,89; 1,98]	1,36 [0,88; 2,11]	2,54 [-0,94; 6,02]	0,1631
INPULSIS-1	309	29 (9,4)	204	17 (8,3)	1,13 [0,64;2,00]	1,14 [0,61;2,13]	1,05 [-3,94;6,05]	0,6832
INPULSIS-2	329	37 (11,2)	219	16 (7,3)	1,54 [0,88;2,70]	1,61 [0,87;2,97]	3,94 [-0,91;8,79]	0,1264
<p>a: Anzahl Studienteilnehmer b: Anzahl Studienteilnehmer mit Ereignis c: RR [95 % KI], OR [95 % KI], RD [95 % KI] INPULSIS (gepoolt): Ein OR bzw. RR kleiner eins zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. Eine RD kleiner null zeigt ein geringeres Risiko ein Ereignis zu erleiden unter Nintedanib im Vergleich zu Placebo an. p-Wert INPULSIS (gepoolt): berechnet mittels Chi-Quadrat-Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. Additional AEs</p>								

In den INPULSIS Studien (gepoolt) traten *Blutungsereignisse* im Nintedanib-Behandlungsarm numerisch häufiger (10,3 %) als im Placebo-Arm (7,8 %) auf (OR [95 % KI]: 1,36 [0,88; 2,11]; p = 0,1631) (Tabelle 4-48). Nicht schwerwiegende Epistaxis und Kontusionen stellten die häufigsten Blutungsereignisse dar. Die Häufigkeit *schwerwiegender Blutungsereignisse* war in beiden Behandlungsgruppen gering und vergleichbar (Placebo: 1,4 %; Nintedanib: 1,3 %). Da jedoch mit der anti-angiogenen Aktivität von VEGFR-Inhibitoren ein mögliches Risiko von Blutungsereignissen assoziiert wurde, wurde ein entsprechender Warnhinweis in die Fachinformation von Ofev[®] aufgenommen (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b).

Unerwünschte Ereignisse mit einer Häufigkeit von > 10 %

Im Sinne der Vollständigkeit wurde geprüft, welche *UE* in den gepoolten INPULSIS Studien in mindestens einer der Gruppen mit einer Häufigkeit > 10 % auftraten. Hierbei wurden keine *UE* identifiziert, die nicht bereits in einer der oben genannten Kategorien nach Organsystemen von besonderem Interesse berücksichtigt wurden.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Aufgrund der Tatsache, dass es sich im vorliegenden Fall um eine IPD Meta-Analyse von zwei Studien handelt, wurden die statistischen Modelle wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf Forest Plots verzichtet. Bei den durchgeführten statistischen Analysen wurde *Studie* als Faktor in die Modelle aufgenommen, was einer Fixed-Effect Meta-Analyse auf der Basis von individuellen Patientendaten entspricht.

Die Ergebnisse der IPD Meta-Analyse wurden für „INPULSIS gepoolt“ bereits in Abschnitt 4.3.1.3.4.5 dargestellt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.5 Subgruppenanalysen – RCT

Für die tabellarische Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die tabellarische Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.3.1.3.1.

Beschreiben Sie die Ergebnisse von Subgruppenanalysen (einschließlich der Interaktionsterme). Stellen Sie dabei die Ergebnisse in den Subgruppen zunächst für jede einzelne Studie in tabellarischer Form dar. Diese Anforderung gilt sowohl für Subgruppenanalysen auf Basis individueller Patientendaten als auch für solche auf Basis aggregierter Daten. Begründen Sie die Wahl von Trennpunkten, wenn quantitative Merkmale kategorisiert werden. Verwenden Sie dabei nach Möglichkeit die in dem jeweiligen Gebiet gebräuchlichen Einteilungen und begründen Sie etwaige Abweichungen. Kennzeichnen Sie in einzelnen Studien a priori geplante Subgruppenanalysen.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Ergebnisse mithilfe einer Meta-Analyse quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analyse (als Forest-Plot) dar.

Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Meta-Analyse durchführen bzw. wenn Sie nicht alle Studien in die Meta-Analyse einschließen.

Generell sollte die Aussagekraft von Subgruppenanalysen, insbesondere auf Studienebene, kritisch beurteilt werden. Geringe Studienteilnehmerzahlen innerhalb der Subgruppenkategorien, ein Ungleichgewicht zwischen den Subgruppenkategorien und die steigende Anzahl statistischer Tests erhöhen die Wahrscheinlichkeit für falsch positive signifikante Ergebnisse (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013). Diese Aspekte müssen bei der Bewertung von Ergebnissen aus Subgruppenanalysen berücksichtigt werden.

Im Folgenden werden die Subgruppenanalysen der INPULSIS Studien für die Endpunkte zur Wirksamkeit sowie zu den UE berichtet.

Tabelle 4-49: Subgruppen, zu denen Interaktionstests bezüglich Wirksamkeit durchgeführt wurden

Subgruppe	Merkmale
Geschlecht	männlich, weiblich
Alter	< 65 Jahre, ≥ 65 Jahre
Ethnie	kaukasisch, asiatisch
FVC (% pred.)	≤ 70 %, > 70 %
SGRQ	≤ 40 Punkte, > 40 Punkte
Raucherstatus	Nichtraucher, Ex-Raucher/Raucher
systemische Kortikosteroide	ja, nein
Bronchodilator	ja, nein
Studie	1199.32, 1199.34

Verzerrungspotenzial des Subgruppenmerkmals Bronchodilator

Eine der präspezifizierten Subgruppenanalysen betraf das Merkmal Begleittherapie mit Bronchodilator (ja/nein). Die Subgruppenzuordnung erfolgte gemäß Studienprotokoll anhand der bei Studienbeginn angegebenen Begleittherapie. Dieses Vorgehen erwies sich jedoch als nicht zielführend: Die Analyse der Begleittherapie während der Studie ergab, dass die individuelle Ausprägung des Merkmals Bronchodilator über die Beobachtungsdauer der Studie bei einem relevanten Anteil der Studienteilnehmer nicht konstant war. In der INPULSIS-1 Studie erhielten bei Studienbeginn n = 61 Studienteilnehmer im Nintedanib-Arm (19,7 %) und n = 34 Studienteilnehmer im Placebo-Arm (16,7 %) begleitend Bronchodilatoren. Während der Beobachtungsdauer der Studie kehrte sich das Verhältnis jedoch um: Eine Bronchodilator-Begleittherapie erhielten n = 40 Studienteilnehmer im Nintedanib-Arm (12,9 %) bzw. n = 38 im Placebo-Arm (18,6 %). In der INPULSIS-2 Studie zeigten sich dazu analoge Verschiebungen der Merkmalsausprägung im Studienverlauf: Bei Studienbeginn erhielten n = 68 Studienteilnehmer im Nintedanib-Arm (20,7 %) und n = 38 Studienteilnehmer im Placebo-Arm (17,4 %) begleitend Bronchodilatoren. Während der Beobachtungsdauer der Studie kehrte sich das Verhältnis um: Mindestens einmal erhielten Bronchodilatoren n = 49 Studienteilnehmer im Nintedanib-Arm (14,9 %) bzw. n = 48 im Placebo-Arm (21,9 %).

Konsequenz dieser Änderungen bezüglich der individuellen Ausprägung des Subgruppenmerkmals Bronchodilator (ja/nein) war, dass die anhand der Baseline-Dokumentation (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015a) erfolgte Zuordnung eines relevanten Anteils der Studienteilnehmer im weiteren Studienverlauf nicht den tatsächlichen Gegebenheiten entsprach. Daher muss von einer erheblichen Verzerrung der Ergebnisse der Subgruppenanalysen zum Merkmal Bronchodilator (ja/nein) ausgegangen werden: Eine Begleittherapie kann sich nur insofern als potenzieller Effektmodifikator auswirken, wie sie auch tatsächlich besteht.

Die Tatsache, dass diese Veränderungen der Ausprägung für das Subgruppenmerkmal Bronchodilator (ja/nein) analog in beiden Studien beobachtet wurden, weist dabei auf einen systematischen Effekt hin. Es ist auffällig, dass die Veränderungen des Bronchodilator-Bedarfs unter der Behandlung mit Studienmedikation eine Entwicklung hin zu einem höheren Bedarf an Bronchodilator-Begleittherapie in den Placebo-Gruppen sowie eine gegenläufige Entwicklung in den Nintedanib-Gruppen anzeigen. Dies kann als indirekter Hinweis auf eine höhere subjektive Beeinträchtigung der Studienteilnehmer durch Dyspnoe/Luftnot unter Placebo-Behandlung verglichen mit Nintedanib interpretiert werden.

Aus den genannten Gründen werden nachfolgend Ergebnisse der Subgruppenanalyse zum Merkmal Bronchodilator (ja/nein) zwar dargestellt. Diese werden von BI jedoch aufgrund des hohen Verzerrungspotenzials als nicht schlussfolgerungsrelevant betrachtet.

Ergebnisse der Subgruppenanalysen

Im Folgenden (Abschnitt 4.3.1.3.5.1 bis Abschnitt 4.3.1.3.5.4) werden die Ergebnisse der Subgruppenanalysen zusammengefasst. Im Sinne der Übersichtlichkeit werden zunächst jeweils die p-Werte der Interaktionstests in einer Übersichtstabelle für die einzelnen Endpunkte zur Mortalität, Morbidität und Lebensqualität dargestellt. Einzelergebnisse der untersuchten Subgruppen werden für diejenigen Endpunkte und Subgruppen tabellarisch berichtet, bei denen sich mindestens ein Hinweis auf eine Effektmodifikation durch das jeweilige Merkmal ergab ($p < 0,2$). Im Text werden speziell die Subgruppen diskutiert, bei denen sich Belege für Effektmodifikationen ($p < 0,05$) durch den subgruppenbildenden Faktor ergeben haben. Bei Hinweisen auf eine Effektmodifikation werden für solche Endpunkte Subgruppenresultate im Text beschrieben, bei denen ein signifikanter Unterschied zwischen den Behandlungsarmen in der gesamten Zulassungspopulation vorlag.

Es wurden ausschließlich die Subgruppen-Analysen für die gepoolten Daten der INPULSIS Studien dargestellt. Wie bereits in Abschnitt 4.3.1.2 dargelegt, eignen sich die INPULSIS Studien aufgrund ihres identischen Studiendesigns für eine gepoolte Auswertung. Im vorliegenden Fall handelt es sich um IPD Meta-Analysen von zwei Studien. Die statistischen Modelle wurden wie in den Einzelstudien mit dem dazugehörigen Effektmaß gewählt und auf die Darstellung von Forest Plots wurde verzichtet. Zudem ergeben sich in diesen Analysen keine Hinweise auf subgruppenrelevante Effektmodifikationen bezüglich der patientenrelevanten Endpunkte, so dass auf eine zusätzliche, detailliertere Darstellung auf Einzelstudienebene verzichtet wird. Alle Ergebnisse der durchgeführten Subgruppen-Analysen sind zur Gewährleistung der notwendigen Transparenz in Modul 5 dargestellt (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f).

Ländereffekte wurden im Rahmen des Zulassungsverfahrens analysiert und mit der EMA diskutiert. Aufgrund der hieraus gewonnenen Erkenntnisse wurde auf die Durchführung von weiteren Interaktionstests verzichtet (Abschnitt 4.3.1.3.5.5).

Für die Endpunkte

- *Zeit bis zum Tod, unabhängig von der Ursache*
- *Zeit bis respiratorisch bedingtem Tod und*
- *Zeit bis zur ersten adjudizierten akuten Exazerbation*

wurden aufgrund der geringen Ereignisraten keine Subgruppenanalysen durchgeführt, da die niedrige Anzahl von Ereignissen in den Subgruppen zufallsbedingt ungleichmäßig verteilt sein können. Dies könnte möglicherweise zu statistisch signifikanten Ergebnissen führen, die jedoch kritisch betrachtet werden sollten. Überdies besteht schon allein aufgrund der hohen Anzahl an Subgruppenanalysen die Wahrscheinlichkeit, dass zufallsbedingt Unterschiede als statistisch signifikant erscheinen, obwohl es dafür keine medizinische Grundlage gibt. Die Ergebnisse der Endpunkte auf Studienebene sowie der gepoolten Analysen werden in Kapitel 4.3.1.3 berichtet.

4.3.1.3.5.1 Mortalität

Tabelle 4-50: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Mortalität

Subgruppe	Jährliche Abnahme der FVC (ml)	FVC-Responder (Schwellenwert > 10 %)	FVC-Responder (Schwellenwert > 5 %)
Geschlecht	0,1911	0,1746	0,8984
Alter	0,3832	0,3206	0,6954
Ethnie	0,7184	0,8936	0,9310
FVC (% pred.)	0,9505	0,3889	0,8616
SGRQ	0,0961	0,4478	0,5704
Raucherstatus	0,4728	0,8275	0,0166
systemische Kortikosteroide	0,9379	0,9498	0,6166
Bronchodilator	0,4864	0,3172	0,8008
Studie	0,4482	0,1432	0,9313
Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)			

4.3.1.3.5.1.1 Endpunkt *Forcierte Vitalkapazität*

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen (Tabelle 4-50) werden im Folgenden für den Endpunkt *jährliche Abnahme der FVC* die Subgruppen *Geschlecht* und *SGRQ-Gesamtscore* $\leq 40 / > 40$ Punkte berichtet.

Tabelle 4-51: Subgruppenergebnisse für die *jährliche Abnahme der FVC* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	adjustierte Rate ^b (SE)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				MWD (SE)	95 % KI
jährliche Abnahme der FVC (ml)					
Geschlecht				p-Wert Interaktionsterm ^c : 0,1911	
Männer	Nintedanib	507	118,79 (12,930)	115,74 (20,516)	75,46; 156,02
	Placebo	334	234,53 (15,924)		
Frauen	Nintedanib	131	92,98 (18,997)	90,65 (29,674)	32,11; 149,19
	Placebo	89	183,63 (22,745)		
SGRQ-Gesamtscore				p-Wert Interaktionsterm ^c : 0,0961	
≤ 40 Punkte	Nintedanib	323	100,53 (14,375)	100,56 (22,096)	57,14; 143,97
	Placebo	232	201,09 (16,756)		
> 40 Punkte	Nintedanib	301	130,53 (17,361)	125,04 (28,243)	69,52; 180,56
	Placebo	187	255,57 (22,243)		
a: Anzahl Studienteilnehmer in Subgruppe b: adjustierte Rate und MWD [95 % KI] INPULSIS (gepoolt): berechnet mittels Random Coeffizient-Regression mit festen Effekten für Studie, Behandlung, Alter, Größe und zufälligem Effekt für Studienteilnehmer, spezifischen Achsenabschnitt und Zeit c: berechnet mittels Random Coeffizient-Regression mit festen Effekten für Studie, Behandlung, Alter, Größe, Subgruppe, Wechselwirkung Behandlung*Zeit*Subgruppe und zufälligem Effekt für Studienteilnehmer, spezifischen Achsenabschnitt und Zeit Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.15.2.3, Tab. 86.15.2.15					

Die Interaktionstests für den Endpunkt *jährliche Abnahme der FVC* ergaben Hinweise auf mögliche Modifikationen der Behandlungseffekte durch die Subgruppen *Geschlecht* und *SGRQ-Gesamtscore* $\leq 40 / > 40$ Punkte (Tabelle 4-51).

In der Subgruppe *Geschlecht* war die Differenz der *jährlichen FVC-Abnahme* von Baseline bis Woche 52 zwischen den Behandlungsgruppen bei den männlichen Studienteilnehmern größer als bei den weiblichen. Jedoch wurde sowohl bei Männern als auch bei Frauen im Placebo-Arm eine größere *jährliche Abnahme der FVC* dokumentiert als im Nintedanib-Arm. Die Effekte waren somit gleichgerichtet. Zudem war der Unterschied zwischen den Behandlungsgruppen bei beiden Geschlechtern statistisch signifikant.

In der Subgruppe *SGRQ-Gesamtscore* $\leq 40 / > 40$ Punkte war die Differenz von Baseline bis Woche 52 zwischen den Behandlungsgruppen Nintedanib und Placebo bei Studienteilnehmern mit einem *SGRQ-Gesamtscore* von > 40 Punkten größer als bei Studienteilnehmern mit einem *SGRQ-Gesamtscore* < 40 Punkte, was auf eine stärkere *jährliche Abnahme der FVC* bei Studienteilnehmern mit einer bei Studienbeginn subjektiv schlechter empfundenen gesundheitsbezogenen Lebensqualität hinweist. Sowohl bei Studienteilnehmern mit einem *SGRQ-Gesamtscore* ≤ 40 Punkte als auch > 40 Punkte war im Placebo-Arm die *jährliche Abnahme der FVC* größer als im Nintedanib-Arm. Somit war in der Subgruppenanalyse *SGRQ-Gesamtscore* die Effektrichtung gleich. Der Unterschied zwischen den Behandlungsgruppen war sowohl bei einem *SGRQ-Gesamtscore* ≤ 40 Punkte als auch > 40 Punkte statistisch signifikant.

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen (Tabelle 4-50) werden im Folgenden für den Endpunkt *FVC-Responder* die Subgruppen *Geschlecht*, *Studie* und *Raucherstatus* berichtet.

Tabelle 4-52: Subgruppenergebnisse für *FVC-Responder* (Schwellenwert 10 % und 5 % vom Sollwert) für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N (%)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				OR	95 % KI
FVC-Responder (10 % vom Sollwert)					
Geschlecht p-Wert Interaktionsterm ^c : 0,1746					
Männer	Nintedanib	507	356 (70,22)	1,442	1,07; 1,94
	Placebo	334	210 (62,87)		
Frauen	Nintedanib	131	91 (69,47)	2,186	1,24; 3,84
	Placebo	89	46 (51,69)		
Studie^d p-Wert Interaktionsterm ^e : 0,1432					
1199.32	Nintedanib	309	218 (70,55)	1,914	1,32; 2,79
	Placebo	204	116 (56,86)		
1199.34	Nintedanib	329	229 (69,60)	1,286	0,89; 1,86
	Placebo	219	140 (63,93)		
FVC-Responder (5 % vom Sollwert)					
Raucherstatus p-Wert Interaktionsterm ^c : 0,0166					
Nicht-Raucher	Nintedanib	174	77 (44,25)	1,122	0,70; 1,80
	Placebo	122	51 (41,80)		
Ex-Raucher/ Raucher	Nintedanib	464	261 (56,25)	2,216	1,64; 2,99
	Placebo	301	113 (37,54)		

a: Anzahl Studienteilnehmer in Subgruppe
 b: OR [95 % KI] INPULSIS (gepoolt): berechnet mittels logistischer Regression mit Termen für Studie, Behandlung, Alter, Geschlecht, Größe und Baseline FVC (% pred.) eingeschlossen
 c: berechnet mittels logistischer Regression mit Termen für Studie, Behandlung, Alter, Geschlecht, Größe, Subgruppe und Wechselwirkung Behandlung*Subgruppe
 d: Bei den Subgruppenanalysen für die gepoolten INPULSIS Studien wurde grundsätzlich die Kovariate „Studie“ bei den Analysemodellen berücksichtigt, um ggf. vorhandene Studieneffekte auszugleichen.
 e: berechnet mittels logistischer Regression mit Termen Behandlung, Alter, Geschlecht, Größe, Subgruppe und Wechselwirkung Behandlung*Subgruppe
 Studienteilnehmer waren FVC-Responder, wenn sie keinen absoluten Abfall der FVC von mehr als 10 % oder 5 % der FVC (% vom Sollwert) hatten und ein Wert zu Woche 52 vorlag.
 Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.14.2.3, Tab. 86.13.2.8, Tab. 86.14.1.2

Für den Endpunkt *Anteil FVC-Responder* (Schwellenwert $<^{\circ}10\%$ vom Sollwert) ergab der Interaktionstest einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *Geschlecht* (Tabelle 4-52). Die Effekte in den Subgruppen waren gleichgerichtet, da sowohl bei Männern als auch bei Frauen die Anteile *FVC-Responder* (Schwellenwert $<^{\circ}10\%$ vom Sollwert) bei den Nintedanib-behandelten Studienteilnehmern nach 52^o Wochen Beobachtungszeit größer waren als bei den Placebo-behandelten. Die Unterschiede zwischen den Behandlungsgruppen Nintedanib und Placebo waren bei beiden Geschlechtern statistisch signifikant.

Der Interaktionstest für den Endpunkt *FVC-Responder* (Schwellenwert $<^{\circ}5\%$) ergab einen Beleg für eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *Raucherstatus* (Tabelle 4-52). Die Effektrichtung war gleichgerichtet, da sowohl bei Nicht-Rauchern als auch bei Ex-Rauchern/Rauchern der Anteil *FVC-Responder* (Schwellenwert $<^{\circ}5\%$) im Nintedanib-Arm größer war als im Placebo-Arm. Somit handelt es sich um eine quantitative, nicht jedoch qualitative Effektmodifikation. In der Kategorie Ex-Raucher/Raucher waren die Unterschiede zwischen den Behandlungsgruppen Nintedanib und Placebo statistisch signifikant.

Fazit

Bei den nach formalen Kriterien identifizierten Hinweisen auf bzw. dem Beleg für eine Effektmodifikation durch die Subgruppenmerkmale *Geschlecht*, *Baseline-SGRQ* und *Raucherstatus* handelt es sich nicht um qualitative sondern lediglich quantitative Modifikationen des Effekts von Nintedanib auf FVC-Surrogatendpunkte für die *Mortalität*. Diese haben keine Konsequenz für die Schlussfolgerungen bezüglich des Patientennutzens, da das Vorhandensein eines ausgeprägten Therapieeffektes von Nintedanib gegenüber Placebo im Hinblick auf die *jährliche FVC-Abnahme* sowie die *FVC-Responder* für alle betrachteten Subgruppen außer Frage steht.

4.3.1.3.5.2 Morbidität

Tabelle 4-53: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Morbidität

Subgruppe	Zeit bis zur ersten akuten Exazerbation
Geschlecht	0,8336
Alter	0,2325
Ethnie	0,6752
FVC (% pred.)	0,1747
SGRQ	0,6142
Raucherstatus	0,5739
systemische Kortikosteroide	0,1425
Bronchodilator	0,2014
Studie	0,0320
Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a), (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)	

4.3.1.3.5.2.1 Endpunkt Zeit bis zur ersten akuten Exazerbation

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen (Tabelle 4-50) werden im Folgenden für den Endpunkt *Zeit bis zur ersten akuten Exazerbation* die Subgruppen *FVC*, *systemischer Kortikosteroide* und *Studie* berichtet.

Tabelle 4-54: Subgruppenergebnisse für die *Zeit bis zur ersten akuten Exazerbation* die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N (%)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				HR	95 % KI
Zeit bis zur ersten akuten Exazerbation					
FVC (% pred.)				p-Wert Interaktionsterm ^c : 0,1747	
≤ 70 %	Nintedanib	207	16 (7,73)	0,52	0,28; 0,99
	Placebo	154	23 (14,94)		
> 70 %	Nintedanib	431	15 (3,48)	1,00	0,44; 2,30
	Placebo	269	9 (3,35)		
systemische Kortikosteroide				p-Wert Interaktionsterm ^c : 0,1425	
ja	Nintedanib	136	6 (4,41)	0,34	0,12; 0,94
	Placebo	89	11 (12,36)		
nein	Nintedanib	502	25 (4,98)	0,82	0,46; 1,47
	Placebo	334	21 (6,29)		
Studie^d				p-Wert Interaktionsterm ^c : 0,0320	
1199.32	Nintedanib	309	19 (6,1)	1,15	0,54; 2,42
	Placebo	204	11 (5,4)		
1199.34	Nintedanib	329	12 (3,6)	0,38	0,19; 0,77
	Placebo	219	21 (9,6)		

a: Anzahl Studienteilnehmer in Subgruppe

b: HR [95 % KI] berechnet mittels Cox-Regression mit Termen für Behandlung, Geschlecht, Alter, Größe und Studie.

c: berechnet mittels Cox-Regression mit Termen für Behandlung, Geschlecht, Alter, Größe, Studie, Subgruppe und Wechselwirkung Behandlung*Subgruppe

d: Bei den Subgruppenanalysen für die gepoolten INPULSIS Studien wurde grundsätzlich die Kovariate „Studie“ bei den Analysemodellen berücksichtigt, um ggfs vorhandene Studieneffekte auszugleichen.

Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014a): Tab. 4.2.2.2.8; Tab. 4.2.2.2.16; (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.16.1.3

Für den Endpunkt *Zeit bis zur ersten akuten Exazerbation* ergab der Interaktionstest einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppen *FVC* ($\leq 70\%$ / $> 70\%$) und *systemischer Kortikosteroide* (Tabelle 4-54). Die Effekte in den Subgruppen waren gleichgerichtet. Patienten unter der Behandlung mit Nintedanib hatten nach 52 Wochen Beobachtungszeit seltener eine Exazerbation im Vergleich zu Placebo-behandelten Studienteilnehmern. Lediglich in der Subgruppe *FVC* $>70\%$ zeigte sich nominal kein Unterschied zwischen den Behandlungsarmen. Ein statistisch signifikanter Effekt ergab sich für den Endpunkt *Zeit bis zur ersten akuten Exazerbation* in der Subgruppe *FVC* $\leq 70\%$ und Verwendung *systemische Kortikosteroide*.

Fazit

Für die Subgruppen *FVC* und Verwendung *systemischer Kortikosteroide* wird ein Hinweis auf eine Interaktion zwischen den Subgruppenausprägungen gezeigt. Dabei handelt es sich um quantitative Effekt-Modifikationen von Nintedanib auf die *Zeit bis zur ersten akuten Exazerbation*. Diese haben keine Konsequenz für die Schlussfolgerungen bezüglich des Patientennutzens.

4.3.1.3.5.3 Gesundheitsbezogene Lebensqualität

Tabelle 4-55: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zur Lebensqualität

Subgruppe	SGRQ-Gesamt-score	SGRQ-I-Gesamt-score	SGRQ-(≤ -4 Punkte)-Responder	SGRQ-I-(≤ -4 Punkte)-Responder	SOBQ	CASA-Q Husten-Symptome	CASA-Q Husten-Impact	PGI-C
Geschlecht	0,0591	0,1267	0,9095	0,9617	0,6121	0,6729	0,1151	0,9424
Alter	0,2232	0,1848	0,0839	0,3068	0,2029	0,2318	0,0994	0,5522
Ethnie	0,7292	0,4855	0,8388	0,7679	0,8864	0,3784	0,7846	0,2534
FVC (% pred.)	0,0631	0,0256	0,0293	0,1058	0,2132	0,3644	0,2234	0,0757
SGRQ	0,4785	0,8811	0,7544	0,5631	0,7292	0,7429	0,4575	0,5437
Raucherstatus	0,9753	0,6958	0,9076	0,4283	0,6280	0,8299	0,5731	0,9953
systemische Kortikosteroide	0,6876	0,4328	0,7410	0,8819	0,9190	0,1314	0,3648	0,8626
Bronchodilator	0,1131	0,1983	0,0002	0,0026	0,9944	0,2090	0,0699	0,7940
Studie	0,2081	0,3285	0,0277	0,0164	0,6483	0,3627	0,9667	0,7607
Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)								

4.3.1.3.5.3.1 Endpunkte zum SGRQ, SGRQ-I, CASA-Q und PGI-C

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen (Tabelle 4-55) werden im Folgenden für die Endpunkte *Veränderung des SGRQ-Gesamtscores* und *Veränderung des SGRQ-I Gesamtscores* die Subgruppen *Geschlecht*, *FVC >70%/≤70%* und *Verwendung eines Bronchodilators* sowie zusätzlich die Subgruppe *Alter* beim Endpunkt *SGRQ-I* berichtet.

Tabelle 4-56: Subgruppenergebnisse für die *Veränderung des SGRQ-Gesamtscores* und *SGRQ-I-Gesamtscores* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	N ^b	MWD ^c (SE)	Behandlungseffekt Nintedanib vs. Placebo ^c	
					MWD (SE)	95 % KI
Veränderung SGRQ-Gesamtscore						
Geschlecht p-Wert Interaktionsterm ^d : 0,0591						
Männer	Nintedanib	507	487	3,88 (0,600)	-0,42 (0,950)	-2,29; 1,44
	Placebo	334	328	4,31 (0,737)		
Frauen	Nintedanib	131	122	2,02 (1,233)	-5,38 (1,868)	-9,05; -1,71
	Placebo	89	85	7,41 (1,407)		
FVC (% pred.) p-Wert Interaktionsterm ^d : 0,0631						
≤ 70 %	Nintedanib	207	199	3,52 (0,973)	-3,34 (1,506)	-6,29; -0,38
	Placebo	154	150	6,86 (1,151)		
> 70 %	Nintedanib	431	410	3,56 (0,643)	-0,34 (1,018)	-2,34; 1,65
	Placebo	269	263	3,90 (0,789)		
Verwendung Bronchodilator p-Wert Interaktionsterm ^d : 0,1131						
ja	Nintedanib	129	122	2,86 (1,132)	-6,41 (1,979)	-10,29; -2,52
	Placebo	72	68	9,27 (1,612)		
nein	Nintedanib	509	487	3,60 (0,611)	-0,71 (0,942)	-2,56; 1,14
	Placebo	351	345	4,31 (0,717)		
Veränderung SGRQ-I-Gesamtscore						
Geschlecht p-Wert Interaktionsterm ^d : 0,1267						
Männer	Nintedanib	507	487	3,88 (0,624)	-1,12 (0,988)	-3,05; 0,82
	Placebo	334	327	4,99 (0,766)		
Frauen	Nintedanib	131	123	1,81 (1,254)	-5,40 (1,897)	-9,12; -1,67
	Placebo	89	86	7,21 (1,427)		

Endpunkt/ Subgruppe	Behandlung	N ^a	N ^b	MWD ^c (SE)	Behandlungseffekt Nintedanib vs. Placebo ^c	
					MWD (SE)	95 % KI
Alter					p-Wert Interaktionsterm ^d : 0,1848	
< 65 Jahre	Nintedanib	144	249	3,87 (0,834)	-0,41 (1,371)	-3,10; 2,28
	Placebo	145	144	4,28 (1,088)		
≥ 65 Jahre	Nintedanib	380	361	3,20 (0,748)	-2,93 (1,141)	-5,16; -0,69
	Placebo	278	269	6,12 (0,862)		
FVC (% pred.)					p-Wert Interaktionsterm ^d 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)		
Verwendung Bronchodilator					p-Wert Interaktionsterm ^d 0,1983	
ja	Nintedanib	129	123	2,30 (1,152)	-6,33 (2,018)	-10,30; -2,37
	Placebo	72	68	8,63 (1,647)		
nein	Nintedanib	509	487	3,68 (0,636)	-1,34 (0,978)	-3,25; 0,58
	Placebo	351	345	5,02 (0,743)		
<p>a: Anzahl Studienteilnehmer in Subgruppe b: Anzahl analysierter Studienteilnehmer in Subgruppe c: MWD [95 % KI] INPULSIS (gepoolt): berechnet mittels MMRM mit festen Effekten für Studie, Behandlung, Visit, Behandlung*Visit, (Baseline SGRQ oder SGRQ-I-Gesamtscore), (Baseline SGRQ- oder SGRQ-I-Gesamtscore)*Visit und zufälligem Effekt für Studienteilnehmer d: berechnet mittels MMRM mit festen Effekten für Studie, Behandlung, Visit, Behandlung*Visit, (Baseline SGRQ oder SGRQ-I-Gesamtscore), (Baseline SGRQ- oder SGRQ-I-Gesamtscore)*Visit, Subgruppe, Behandlung*Subgruppe und zufälligem Effekt für Studienteilnehmer</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.6.2.3, Tab. 86.6.2.12, Tab. 86.6.2.24, Tab. 86.5.2.3, Tab. 86.5.2.6, Tab. 86.5.2.12, Tab. 86.5.2.24</p>						

Ergebnisse zum SGRO

Für die gepoolte Auswertung der INPULSIS Studien ergaben die Interaktionstests für den Endpunkt *Veränderung SGRQ-Gesamtscores* Hinweise auf mögliche Modifikationen der Behandlungseffekte durch die Subgruppen *Geschlecht*, *FVC* und *Verwendung Bronchodilator*, wobei sich in allen Subgruppen gleichgerichtete Effekte zugunsten von Nintedanib zeigten (Tabelle 4-56).

Sowohl bei den männlichen als auch bei den weiblichen Studienteilnehmern stieg der *SGRQ-Gesamtscore* von Baseline bis Woche 52 im Placebo-Arm stärker an als im Nintedanib-Arm. Dies zeigt, dass die Studienteilnehmer unter Placebo-Behandlung im Vergleich zur Nintedanib-Behandlung im Krankheitsverlauf eine schlechtere gesundheitsbezogene Lebensqualität berichteten. Der Unterschied in der *Veränderung des SGRQ-Gesamtscores* im Zeitverlauf zwischen den Behandlungsgruppen war in der Subgruppe der Frauen im Vergleich zu den Männern größer. Der Unterschied zwischen den Behandlungsgruppen Nintedanib und Placebo war bei den weiblichen Studienteilnehmern statistisch signifikant und übertraf zudem für die Mittelwertdifferenz die MCID von weniger als 4 Punkten.

Für die Subgruppe *FVC* als Maß für die Krankheitsschwere bei Studienbeginn war die Zunahme des *SGRQ-Gesamtscores* von Baseline bis Woche 52 sowohl bei Studienteilnehmern mit einer größeren Krankheitsschwere ($FVC \leq 70\%$) als auch bei denen mit einer geringeren Krankheitsschwere ($FVC > 70\%$) im Placebo-Arm größer als im Nintedanib-Arm. Die Effektrichtung war somit gleich. Die Differenz zwischen den Behandlungsgruppen war in der Kategorie Studienteilnehmer mit einer $FVC \leq 70\%$ größer als bei Studienteilnehmern mit einer $FVC > 70\%$. Die Betrachtung der Subgruppenanalyse weist darauf hin, dass Studienteilnehmer mit einer höheren Krankheitsschwere unter Placebo im Verlauf der Studie eine Verschlechterung der subjektiv empfundenen gesundheitsbezogenen Lebensqualität (6,86 Punkte) angaben als Studienteilnehmer mit einer leichten IPF (3,90 Punkte). Der Unterschied zwischen den Behandlungsgruppen Nintedanib und Placebo war bei Studienteilnehmern mit einer größeren Krankheitsschwere ($FVC \leq 70\%$) statistisch signifikant (MWD [95 % KI]: -3,34 [-6,29; -0,38]).

Für die Subgruppe *Verwendung von Bronchodilatoren* waren die Behandlungseffekte gleichgerichtet. Sowohl bei Studienteilnehmern mit als auch bei solchen ohne eine Begleittherapie mit einem Bronchodilator war die *Veränderung des SGRQ-Gesamtscores* von Baseline bis Woche 52 bei den Placebo-behandelten Studienteilnehmern größer als bei den Nintedanib-behandelten Studienteilnehmern. Die Differenz zwischen den Behandlungsgruppen war bei den Studienteilnehmern mit Bronchodilator-Therapie größer als bei denen ohne diese Therapie. Der Unterschied zwischen den Behandlungsgruppen Nintedanib und Placebo war bei Studienteilnehmern mit Bronchodilator-Therapie statistisch signifikant.

Wie einleitend zum Abschnitt Subgruppenanalysen beschrieben, betrachtet BI die Subgruppenanalyse zum Merkmal *Verwendung Bronchodilator* allerdings aufgrund der Veränderungen der Merkmalsausprägung unter der Behandlung mit Studienmedikation als hochgradig verzerrt und somit nicht valide.

Ergebnisse zum SGRQ-I

Die Interaktionstests für den Endpunkt *Veränderung des SGRQ-I-Gesamtscores* ergaben Hinweise auf mögliche Modifikationen der Behandlungseffekte durch die Subgruppen *Geschlecht*, *Alter* und *Verwendung Bronchodilator* sowie einen Beleg für eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *FVC (% vom Sollwert)* als Maß für die Krankheitsschwere. Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen werden im Folgenden für die Endpunkte *Veränderung des SGRQ-Gesamtscores* und *Veränderung des SGRQ-I-Gesamtscores* die Subgruppen *Geschlecht*, *FVC ($\leq 70\%$ / $> 70\%$)* und *Verwendung Bronchodilator* sowie zusätzlich die Subgruppe *Alter* beim Endpunkt *SGRQ-I* berichtet (Tabelle 4-56).

Sowohl bei männlichen als auch bei weiblichen Studienteilnehmern war die Zunahme des *SGRQ-I-Gesamtscores* von Baseline bis Woche 52 unter Placebo-Behandlung größer als unter Nintedanib-Behandlung, entsprechend einer ausgeprägteren Verschlechterung der HRQoL unter Placebo-Behandlung im Vergleich zur Behandlung mit Nintedanib. Die Differenz zwischen den Behandlungsgruppen war dabei bei den weiblichen Studienteilnehmern größer als bei den männlichen. Der Unterschied zwischen den Behandlungsgruppen war bei Frauen statistisch signifikant und klinisch relevant zum Schwellenwert von -4 Punkten.

In der Subgruppenanalyse der Studienteilnehmer jünger als 65 Jahre und mindestens 65 Jahre zeigt sich für beide Alters-Strata eine größere Abnahme der Lebensqualität durch einen stärkeren Anstieg im *SGRQ-I-Gesamtscore* von Baseline bis Woche 52 im Placebo-Arm im Vergleich zum Nintedanib-Arm. Im Zeitverlauf gaben die Placebo-behandelten Studienteilnehmer gegenüber den Nintedanib-behandelten somit eine größere Verschlechterung der subjektiv empfundenen Lebensqualität an. Die Differenz zwischen den Behandlungsgruppen war bei den älteren Studienteilnehmern größer als bei den jüngeren Studienteilnehmern. Die Unterschiede zwischen den Behandlungsgruppen waren bei Älteren statistisch signifikant.

Sowohl bei Studienteilnehmern mit einer größeren Krankheitsschwere (*FVC $\leq 70\%$*) als auch bei denen mit einer geringeren Krankheitsschwere (*FVC $> 70\%$*) fiel die Zunahme im *SGRQ-I-Gesamtscore* von Baseline bis Woche 52 im Placebo-Arm größer aus als im Nintedanib-Arm. Die Differenz zwischen den Behandlungsgruppen war in der Studienpopulation mit *FVC $\leq 70\%$* , d. h. mit moderater oder schwerer IPF-Erkrankung, größer als bei der Studienpopulation mit leichter IPF (*FVC $> 70\%$*). Diese Unterschiede waren bei Studienteilnehmern mit *FVC $\leq 70\%$* statistisch signifikant.

Für die Subgruppe *Verwendung von Bronchodilator* waren die Behandlungseffekte ebenfalls gleichgerichtet. Sowohl bei Studienteilnehmern, die einen *Bronchodilator*, als auch bei denen, die keinen *Bronchodilator* verwendeten, war die Zunahme im *SGRQ-I-Gesamtscore* von Baseline bis Woche 52 unter Placebo größer als im Vergleich zu Nintedanib. Die Differenz zwischen den Behandlungsgruppen Nintedanib und Placebo war bei den Studienteilnehmern, die *Bronchodilatoren* verwendeten, größer, als bei denen, die keine verwendeten (-6,33 Punkte vs. -1,34 Punkte). Dieser Unterschied war bei Studienteilnehmern mit Bronchodilator-Therapie statistisch signifikant. Wie jedoch einleitend zum Abschnitt Subgruppenanalysen beschrieben, betrachtet BI die Subgruppenanalyse zum Merkmal *Bronchodilator* aufgrund der Veränderungen der Merkmalsausprägung unter der Behandlung mit Studienmedikation als hochgradig verzerrt und somit nicht valide.

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen werden im Folgenden für den Endpunkt *SGRQ-(≤ -4 Punkte)-Responder* die Subgruppen *Alter*, *FVC* und *Verwendung von Bronchodilator* und für den Endpunkt *SGRQ-I-(≤ -4 Punkte)-Responder* die Subgruppen *FVC* und *Verwendung von Bronchodilator* berichtet.

Tabelle 4-57: Subgruppenergebnisse für *SGRQ-(≤ -4 Punkte)-Responder* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Responder N (%)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				OR	95 % KI
SGRQ-(≤ -4 Punkte)-Responder					
Alter				p-Wert Interaktionsterm ^c : 0,0839	
< 65 Jahre	Nintedanib	258	58 (22,48)	0,843	0,52; 1,36
	Placebo	145	38 (26,21)		
≥ 65 Jahre	Nintedanib	380	88 (23,16)	1,460	0,98; 2,17
	Placebo	278	48 (17,27)		
FVC (% pred.)				p-Wert Interaktionsterm ^c : 0,0293	
≤ 70 %	Nintedanib	207	51 (24,64)	1,934	1,11; 3,38
	Placebo	154	23 (14,94)		
> 70 %	Nintedanib	431	95 (22,04)	0,917	0,63; 1,33
	Placebo	269	63 (23,42)		
Verwendung Bronchodilator				p-Wert Interaktionsterm ^c : 0,0002	
Ja	Nintedanib	129	32 (24,81)	7,375	2,16; 25,17
	Placebo	72	3 (4,17)		
Nein	Nintedanib	509	114 (22,40)	0,972	0,70; 1,35
	Placebo	351	83 (23,65)		

Endpunkt/ Subgruppe	Behandlung	N ^a	Responder N (%)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				OR	95 % KI
Studie^d				p-Wert Interaktionsterm ^e : 0,0277	
1199.32	Nintedanib	309	63 (20,39)	0,840	0,55;1,29
		204	49 (24,02)		
1199.34	Nintedanib	329	83 (25,23)	1,664	1,08;2,57
		219	37 (16,89)		
SGRQ-I(≤ -4 Punkte)-Responder					
FVC (% pred.)				p-Wert Interaktionsterm ^e : 0,1058	
≤ 70 %	Nintedanib	207	53 (25,60)	1,829	1,06; 3,15
	Placebo	154	25 (16,23)		
> 70 %	Nintedanib	431	97 (22,51)	1,057	0,73; 1,54
	Placebo	269	58 (21,56)		
Verwendung Bronchodilator				p-Wert Interaktionsterm ^e : 0,0026	
Ja	Nintedanib	129	37 (28,68)	4,301	1,71; 10,84
	Placebo	72	6 (8,33)		
Nein	Nintedanib	509	113 (22,20)	1,061	0,76; 1,48
	Placebo	351	77 (21,94)		
Studie^d				p-Wert Interaktionsterm ^e : 0,0164	
1199.32	Nintedanib	309	66 (21,36)	0,886	0,58; 1,36
	Placebo	204	49 (24,02)		
1199.34	Nintedanib	329	84 (25,53)	1,877	1,20; 2,93
	Placebo	219	34 (15,53)		
<p>a: Anzahl Studienteilnehmer in Subgruppe</p> <p>b: OR [95 % KI] INPULSIS (gepoolt): berechnet mittels logistische Regression mit Termen für Studie, Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore</p> <p>c: berechnet mittels logistische Regression mit Termen für Studie, Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore, Subgruppe, Wechselwirkung Behandlung*Subgruppe</p> <p>d: Bei den Subgruppenanalysen für die gepoolten INPULSIS Studien wurde grundsätzlich die Kovariate „Studie“ bei den Analysemodellen berücksichtigt, um ggfs vorhandene Studieneffekte auszugleichen.</p> <p>e: berechnet mittels logistische Regression mit Termen für Behandlung, Baseline SGRQ-Gesamtscore bzw. SGRQ-I-Gesamtscore, Subgruppe, Wechselwirkung Behandlung*Subgruppe</p> <p>Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.1.2.4, Tab. 86.1.2.6, Tab. 86.1.2.10, Tab. 79.1.2.6, Tab. 79.1.2.10, Tab. 86.1.1.2, Tab. 79.1.1.2</p>					

Ergebnisse zu SGRQ-(≤ -4 Punkte)-Responder

Für die gepoolte Auswertung der INPULSIS Studien ergaben die Interaktionstests für den Endpunkt *SGRQ-(≤ -4 Punkte)-Responder* einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *Alter* und Belege für mögliche Modifikationen der Behandlungseffekte durch die Subgruppen *FVC* als Maß für die Krankheitsschwere sowie *Verwendung von Bronchodilator* (Tabelle 4-56).

Die Subgruppenanalyse für das *Alter* ergab, dass bei Studienteilnehmern jünger als 65 Jahre der Anteil *SGRQ-(≤ -4 Punkte)-Responder* im Placebo-Arm größer war als im Nintedanib-Arm. Bei Studienteilnehmern älter als 65 Jahre hingegen war der Anteil *SGRQ-(≤ -4 Punkte)-Responder* im Nintedanib-Arm größer als im Placebo-Arm. Die Effektrichtung war somit gegenläufig. Die Unterschiede zwischen den Behandlungsgruppen waren in beiden Altersgruppen nicht statistisch signifikant.

Für die Subgruppe *FVC (≤ 70 %/ > 70 %)* als Maß für die Krankheitsschwere war der Anteil *SGRQ-(≤ -4 Punkte)-Responder* bei Studienteilnehmern mit einer größeren Krankheitsschwere (*FVC ≤ 70 %*) im Nintedanib-Arm circa doppelt so groß wie im Placebo-Arm. Hingegen war der Anteil *SGRQ-(≤ -4 Punkte)-Responder* bei Studienteilnehmern mit einer geringeren Krankheitsschwere (*FVC > 70 %*) ähnlich. Die Unterschiede zwischen den Behandlungsgruppen waren für die Subgruppe *FVC ≤ 70 %* statistisch signifikant zu Gunsten von Nintedanib.

Für die Subgruppe *Verwendung von Bronchodilator* waren die Behandlungseffekte ebenfalls gegenläufig. Bei Nintedanib-behandelten Studienteilnehmern, die einen *Bronchodilator* verwendeten, war der Anteil *SGRQ-(≤ -4 Punkte)-Responder* größer als bei Placebo-behandelten Studienteilnehmern mit *Bronchodilator*. Hingegen war bei den Studienteilnehmern, die keinen Bronchodilator verwendeten, der Anteil *SGRQ-(≤ -4 Punkte)-Responder* in den verschiedenen Behandlungsgruppen unter Placebo etwas höher. Die Unterschiede zwischen den Behandlungsgruppen Nintedanib und Placebo waren unter Bronchodilator-Therapie statistisch signifikant. Wie einleitend zum Abschnitt Subgruppenanalysen beschrieben, betrachtet BI die Subgruppenanalyse zum Merkmal Bronchodilator allerdings aufgrund der Veränderungen der Merkmalsausprägung unter der Behandlung mit Studienmedikation als hochgradig verzerrt und somit nicht valide.

Ergebnisse zu SGRQ-I-(≤ -4 Punkte)-Responder

Für die gepoolte Auswertung der INPULSIS Studien ergaben die Interaktionstests für den Endpunkt *SGRQ-I-(≤ -4 Punkte)-Responder* einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *FVC* als Maß für die Krankheitsschwere und einen Beleg für eine mögliche Modifikation der Behandlungseffekte durch die Subgruppe *Verwendung von Bronchodilator*. Die Effekte waren in beiden Subgruppenanalysen jeweils für beide Strata gleichgerichtet (Tabelle 4-57).

Für die Subgruppenanalyse nach *FVC* als Maß für die Krankheitsschwere war der Anteil *SGRQ-I-(≤ -4 Punkte)-Responder* bei Studienteilnehmern mit einer größeren Krankheitsschwere ($FVC \leq 70\%$) im Nintedanib-Arm circa doppelt so groß wie im Placebo-Arm. Der Anteil *SGRQ-I-(≤ -4 Punkte)-Responder* bei Studienteilnehmern mit einer geringeren Krankheitsschwere ($FVC > 70\%$) war unter Nintedanib etwas größer als unter Placebo. Die Effekte waren somit gleichgerichtet. Die Unterschiede zwischen den Behandlungsgruppen Nintedanib und Placebo waren bei Studienteilnehmern mit $FVC \leq 70\%$ statistisch signifikant, bei einer $FVC > 70\%$ nicht statistisch signifikant.

Für die Subgruppe *Verwendung von Bronchodilator* waren die Behandlungseffekte ebenfalls gleichgerichtet. Bei Nintedanib-behandelten Studienteilnehmern, die einen *Bronchodilator* verwendeten, war der Anteil *SGRQ-I-(≤ -4 Punkte)-Responder* circa viermal größer als bei Placebo-behandelten Studienteilnehmern mit *Bronchodilator*. Bei den Studienteilnehmern, die keinen Bronchodilator verwendeten, war der Anteil *SGRQ-I-(≤ -4 Punkte)-Responder* unter Nintedanib etwas größer als unter Placebo. Die Unterschiede zwischen den Behandlungsgruppen Nintedanib und Placebo waren unter Bronchodilator-Therapie statistisch signifikant. Wie einleitend zum Abschnitt Subgruppenanalysen beschrieben, betrachtet BI die Subgruppenanalyse zum Merkmal *Verwendung von Bronchodilator* allerdings aufgrund der Veränderungen der Merkmalsausprägung unter der Behandlung mit Studienmedikation als hochgradig verzerrt und somit als nicht valide.

Fazit

Für den *SGRQ-I Gesamtscore* und für die *SGRQ-Responderanalyse* wurde jeweils für die Subgruppe *FVC ($\leq 70\%/> 70\%$)* ein Beleg für eine Interaktion ($p < 0,05$) gefunden. Die Ergebnisse der Subgruppenanalyse könnten sich dahingehend interpretieren lassen, dass die Patienten mit einem Baselinewert von $FVC \leq 70\%$ stärker auf die Behandlung mit Nintedanib im Vergleich zu Patienten mit einem höheren FVC zur Baseline ansprechen.

Ergebnisse zu CASA-Q

Für die gepoolte Auswertung der gepoolten INPULSIS Studien ergaben die Interaktionstests (Tabelle 4-55) für den Endpunkt *Veränderung CASA-Q Husten-Symptome* und *Husten-Impact* Hinweise auf mögliche Modifikationen der Behandlungseffekte durch die Subgruppen *systemische Kortikosteroide*, *Geschlecht*, *Alter* und *Verwendung von Bronchodilator*. Die vollständigen Ergebnisse finden sich jedoch in Modul 5.

Tabelle 4-58: Subgruppenergebnisse für CASA-Q für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	N ^b	MWD ^c (SE)	Behandlungseffekt Nintedanib vs. Placebo ^c	
					MWD (SE)	95 % KI
CASA-Q Husten-Symptome						
systemische Kortikosteroide				p-Wert Interaktionsterm ^d :		0,1314
ja	Nintedanib	131	136	-5,11 (1,839)	-3,69 (2,978)	-9,53; 2,16
	Placebo	89	86	-1,42 (2,340)		
nein	Nintedanib	502	494	0,56 (0,865)	2,09 (1,362)	-0,58; 4,76
	Placebo	334	331	-1,52 (1,052)		
CASA-Q Husten-Impact						
Geschlecht				p-Wert Interaktionsterm ^d :		0,1151
Männer	Nintedanib	507	495	-2,77 (0,766)	0,87 (1,221)	-1,52; 3,27
	Placebo	334	329	-3,64 (0,951)		
Frauen	Nintedanib	131	129	-1,41 (1,741)	4,64 (2,674)	-0,61; 9,90
	Placebo	89	88	-6,05 (2,033)		
Alter				p-Wert Interaktionsterm ^d :		0,0994
< 65 Jahre	Nintedanib	258	253	-3,53 (1,094)	-0,24 (1,815)	-3,81; 3,32
	Placebo	145	144	-3,29 (1,450)		
≥ 65 Jahre	Nintedanib	380	371	-1,66 (0,923)	3,10 (1,418)	0,32; 5,88
	Placebo	278	273	-4,75 (1,076)		
Verwendung Bronchodilator				p-Wert Interaktionsterm ^d :		0,0699
ja	Nintedanib	129	126	-1,92 (1,495)	7,89 (2,618)	2,84; 13,12
	Placebo	72	70	-9,90 (2,149)		
nein	Nintedanib	509	498	-2,51 (0,796)	0,79 (1,234)	-1,63; 3,21
	Placebo	351	347	-3,31 (0,944)		

a: Anzahl Studienteilnehmer in Subgruppe

b: Anzahl analysierter Studienteilnehmer in Subgruppe

c: MWD [95 % KI] INPULSIS (gepoolt): berechnet mittels MMRM mit festen Effekten für Studie, Behandlung, Visit, Behandlung*Visit, Behandlung*Visit, (Baseline SGRQ oder SGRQ-I-Gesamtscore), (Baseline SGRQ- oder SGRQ-I-Gesamtscore)*Visit und zufälligem Effekt für Studienteilnehmer

d: berechnet mittels MMRM mit festen Effekten für Studie, Behandlung, Visit, Behandlung*Visit, (Baseline SGRQ oder SGRQ-I-Gesamtscore), (Baseline SGRQ- oder SGRQ-I-Gesamtscore)*Visit, Subgruppe, Behandlung*Subgruppe und zufälligem Effekt für Studienteilnehmer

Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.7.2.21, Tab. 86.8.2.3, Tab. 86.8.2.6, Tab. 86.8.2.24

Für den Endpunkt *CASA-Q Husten-Symptome* ergab der Interaktionstest einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppen *systemische Kortikosteroide* (Tabelle 4-58). Die Effektrichtung war gegenläufig, jedoch lag keine statistische Signifikanz für die Ergebnisse der Endpunkte vor.

Für den Endpunkt *CASA-Q Husten-Impact* ergab der Interaktionstest jeweils Hinweise auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppen *Alter*, *Geschlecht* und *Verwendung eines Bronchodilators* (Tabelle 4-58). Patienten unter der Behandlung mit Nintedanib berichteten nach 52 Wochen Beobachtungszeit im Vergleich zu Placebo-behandelten Studienteilnehmern eine bessere Lebensqualität. Lediglich in der Subgruppe *Alter < 65 Jahre* zeigte sich nominal ein Unterschied zu Gunsten von Placebo-behandelten Studienteilnehmern. Statistisch signifikante Effekte ergaben sich für den Endpunkt *CASA-Q Husten-Impact Score* in der Subgruppe *Verwendung eines Bronchodilators* zu Gunsten von Patienten unter der Behandlung mit Nintedanib.

Fazit

Bei den nach formalen Kriterien identifizierten Hinweisen für eine Effektmodifikation durch die Subgruppenmerkmale *Alter* und *Geschlecht* handelt es sich nicht um qualitative sondern um quantitative Effekt-Modifikationen von Nintedanib auf die *Lebensqualität*. Diese haben keine Konsequenz für die Schlussfolgerungen bezüglich des Patientennutzens.

Ergebnisse zu PGI-CTabelle 4-59: Subgruppenergebnisse für *PGI-C-Responder* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Responder N (%)	Behandlungseffekt Nintedanib vs. Placebo ^b	
				OR	95 % KI
PGI-C-Responder					
FVC (% pred.)				p-Wert Interaktionsterm ^c : 0,0757	
≤ 70 %	Nintedanib	207	121 (58,45)	1,774	1,16; 2,71
	Placebo	154	68 (44,16)		
> 70 %	Nintedanib	431	270 (62,65)	1,106	0,81; 1,51
	Placebo	269	162 (60,22)		
a: Anzahl Studienteilnehmer in Subgruppe b: OR [95 % KI] INPULSIS (gepoolt): berechnet mittels logistische Regression mit Termen für Studie, Behandlung, Alter, Geschlecht, Größe und Baseline PGI-C eingeschlossen c: berechnet mittels logistische Regression mit Termen für Studie, Behandlung, Alter, Geschlecht, Größe, Subgruppe, Wechselwirkung Behandlung*Subgruppe Responder zu Woche 52 definiert als sehr viel besser, viel besser, ein wenig besser oder keine Veränderung Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 86.9.2.6					

Der Interaktionstest für den Endpunkt *PGI-C-Responder* ergab einen Hinweis auf eine mögliche Modifikation des Behandlungseffekts durch die Subgruppe *FVC* als Maß für die Krankheitsschwere, wobei der Effekt bei Studienteilnehmern mit $FVC \leq 70\%$ und $FVC > 70\%$ gleichgerichtet war (Tabelle 4-59).

Für die Subgruppe *FVC* als Maß für die Krankheitsschwere war der Anteil *PGI-C-Responder* sowohl bei Studienteilnehmern mit einer größeren Krankheitsschwere ($FVC \leq 70\%$) als auch bei denen mit einer geringeren Krankheitsschwere ($FVC > 70\%$) im Nintedanib-Arm größer als im Placebo-Arm. Der Unterschied zwischen den Behandlungsgruppen Nintedanib und Placebo war bei schwerer erkrankten Studienteilnehmern statistisch signifikant.

Fazit

Eine Effektmodifikation ist auf Basis dieser Daten nicht zu erwarten. Es kann von einem konstanten Therapieeffekt im Hinblick auf die *PGI-C-Responder* über die analysierten Subgruppen hinweg ausgegangen werden.

4.3.1.3.5.4 Unerwünschte Ereignisse

Im Folgenden werden die Subgruppenanalysen der INPULSIS Studien für die Endpunkte zu UE berichtet. Dabei erfolgt die Darstellung der Subgruppen zu den UE auf Basis der Analysen zu den präspezifizierten Subgruppen gemäß Studienberichten (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c, Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d). Abweichend von den präspezifizierten Subgruppenanalysen zu den Wirksamkeitsendpunkten waren für die Endpunkte zu UE die Subgruppenmerkmale *FVC (% vom Sollwert)*, *SGRQ*, *systemische Kortikosteroide* und *Verwendung von Bronchodilator* nicht präspezifiziert. Als zusätzliche Subgruppenmerkmale für die Arzneimittelsicherheit waren die potenziell die Pharmakokinetik beeinflussenden Faktoren *Gewicht* und Vorliegen einer *Nierenfunktionseinschränkung*, definiert als Kreatinin-Clearance von < 90 mL/min bestimmt über die Cockcroft–Gault-Formel, präspezifiziert. Darüber hinaus umfasste bei der Arzneimittelsicherheit die Subgruppenanalyse nach *Alter* drei Strata (< 65 Jahre, 65–74 Jahre, ≥ 75 Jahre), wohingegen bei den Wirksamkeitsendpunkten nur zwei Alters-Strata (< 65 Jahre, ≥ 65 Jahre) präspezifiziert waren. Somit umfassten die Subgruppenanalysen zu UE die folgenden a priori definierten Subgruppenmerkmale:

Tabelle 4-60: Subgruppen, zu denen Interaktionstests bezüglich *unerwünschter Ereignisse* durchgeführt wurden

Subgruppe	Merkmale
Geschlecht	männlich, weiblich
Alter	< 65 Jahre, 65-74 Jahre, ≥ 75 Jahre
Ethnie	kaukasisch, asiatisch
Nierenfunktionseinschränkung zur Baseline	ja, nein
Raucherstatus	Nichtraucher, Ex-Raucher/Raucher
Gewicht	< 65 kg, ≥ 65 kg

Übersicht zu den p-Werten der Interaktionstests

Tabelle 4-61: Übersicht zu den p-Werten der für die Endpunkte zu unerwünschten Ereignissen

Subgruppe	Gesamtzahl UE	Gesamtzahl SUE	UE, die zum Therapieabbruch führten	UE, die zum Tod führten
Geschlecht	0,541	0,523	0,025	0,249
Alter	0,949	0,778	0,189	0,031
Ethnie	0,944	0,256	0,962	0,805
Nierenfunktionseinschränkung zur Baseline	0,695	0,856	0,499	0,149
Raucherstatus	0,785	0,797	0,114	0,133
Gewicht	0,015	0,619	0,194	0,808

Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)

4.3.1.3.5.4.1 Endpunkt *Gesamtzahl UE*

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen (Tabelle 4-61) werden im Folgenden für die *Gesamtzahl UE* die Ergebnisse der Subgruppenanalyse nach *Gewicht* berichtet.

Tabelle 4-62: Subgruppenergebnisse für die *Gesamtzahl UE* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N (%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Gesamtzahl UE					
Gewicht				p-Wert Interaktionsterm ^b :	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 (94,7)	1,82	1,07;3,10
	Placebo	347	315 (90,8)		
a: Anzahl Studienteilnehmer in Subgruppe b: berechnet mittels Breslow-Day Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 2.1.1.1.1					

Für den Endpunkt *Gesamtzahl UE* fand sich bezüglich des Subgruppenmerkmals *Gewicht* ein Beleg für eine Effektmodifikation (Tabelle 4-62). 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. Dabei wurde auch in der letzteren Gruppe eine statistisch signifikant höhere *Gesamtzahl der UE* unter Nintedanib gegenüber Placebo dokumentiert. Somit handelt es sich hierbei um eine quantitative Effektmodifikation und eine Patientengruppe ohne Zunahme der *Gesamtzahl UE* lässt sich anhand der Subgruppenanalyse nach *Gewicht* nicht identifizieren.

4.3.1.3.5.4.2 Endpunkt UE, die zum Therapieabbruch führen

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen werden im Folgenden für die UE, die zum Therapieabbruch führen, die Subgruppen *Geschlecht*, *Alter*, *Raucherstatus* und *Gewicht* berichtet (Tabelle 4-61).

Tabelle 4-63: Subgruppenergebnisse für die UE, die zum Therapieabbruch führen für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
UE, die zum Therapieabbruch führen					
Geschlecht p-Wert Interaktionsterm ^b : 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)		
Alter p-Wert Interaktionsterm ^c : 0,189					
< 65 Jahre	Nintedanib	258	33 (12,8)	1,10	0,59; 2,06
	Placebo	145	17 (11,7)		
65-74 Jahre	Nintedanib	263	52 (19,8)	1,59	0,97; 2,61
	Placebo	216	29 (13,4)		
≥ 75 Jahre	Nintedanib	117	38 (32,5)	2,83	1,27; 6,34
	Placebo	62	9 (14,5)		
Raucherstatus p-Wert Interaktionsterm ^b : 0,114					
Nicht-Raucher	Nintedanib	174	38 (21,8)	2,56	1,28; 5,14
	Placebo	122	12 (9,8)		
Ex-Raucher/ Raucher	Nintedanib	464	85 (18,3)	1,35	0,90; 2,01
	Placebo	301	43 (14,3)		

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Gewicht				p-Wert Interaktionsterm ^b : 0,194	
< 65 kg	Nintedanib	126	39 (31,0)	2,39	1,16; 4,93
	Placebo	76	12 (15,8)		
≥ 65 kg	Nintedanib	512	84 (16,4)	1,39	0,93; 2,06
	Placebo	347	43 (12,4)		
a: Anzahl Studienteilnehmer in Subgruppe b: berechnet mittels Breslow-Day Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 2.1.1.1.1					

Für den Sicherheits-Endpunkt *UE, die zum Therapieabbruch führen*, fand sich ein Beleg für eine Effektmodifikation durch das Subgruppenmerkmal *Geschlecht* (Tabelle 4-63). Das OR für das Auftreten eines solchen UE war bei Frauen höher als bei Männern. Bei isolierter Betrachtung der Subgruppe der Männer fand sich keine statistisch signifikante Zunahme der Wahrscheinlichkeit des Auftretens von *UE, die zum Therapieabbruch führen*. Allerdings lag der Punktschätzer für das OR auch in der Subgruppe der Männer mit 1,28 über 1, sodass sich das Subgruppenmerkmal *Geschlecht* nicht als qualitativer sondern als quantitativer Effektmodifikator auswirkt.

Bezüglich des Subgruppenmerkmals *Alter* fand sich für *UE, die zum Therapieabbruch führen*, ein Hinweis auf eine Effektmodifikation (Tabelle 4-63) Bei den über 75-jährigen Patienten war das OR für das Auftreten eines solchen Ereignisses unter Nintedanib am stärksten erhöht, und für die beiden jüngeren Alters-Strata wurde jeweils keine statistisch signifikante Zunahme der Wahrscheinlichkeit des Auftretens von *UE, die zum Therapieabbruch führen*, registriert. Die Punktschätzer für das OR in den beiden jüngeren Subgruppen lagen mit 1,10 bzw. 1,59 über 1, sodass der angegebene p-Wert im Interaktionstest auf eine quantitative Effektmodifikation hinweist.

Für das Subgruppenmerkmal *Raucherstatus* ergaben die Analysen der *UE, die zum Therapieabbruch führten*, einen Hinweis auf eine Effektmodifikation (Tabelle 4-63), wobei *Nichtraucher* eine stärkere Erhöhung des OR für solche *UE* aufwiesen als die Subgruppe der *Ex-Raucher/Raucher*. Für letztere isoliert betrachtet fand sich keine statistisch signifikante Erhöhung des OR unter einer Behandlung mit Nintedanib verglichen mit Placebo. Der Punktschätzer für das OR wurde für die Subgruppe der *Ex-Raucher/Raucher* jedoch ebenfalls als größer 1 ermittelt (1,35), sodass für das Subgruppenmerkmal *Raucherstatus* ein Hinweis auf eine quantitative Effektmodifikation vorliegt.

Bezüglich des Subgruppenmerkmals Gewicht fand sich für den Endpunkt *UE, die zum Therapieabbruch führten*, ein Hinweis auf eine Effektmodifikation (Tabelle 4-63). 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 der Gruppe mit höherem Gewicht fand sich keine statistisch signifikante Zunahme der *UE, die zum Therapieabbruch führen*, unter Nintedanib verglichen mit Placebo, doch lag auch hier der Punktschätzer für das OR mit 1,39 über 1. Somit handelt es sich hierbei um einen Hinweis auf eine quantitative Effektmodifikation.

4.3.1.3.5.4.3 Endpunkt UE, die zum Tod führten

Entsprechend der Ergebnisse zu den Interaktionstests innerhalb der Subgruppen werden im Folgenden für die UE, die zum Tod führten, die Subgruppen *Alter*, *Niereninsuffizienz* und *Raucherstatus* berichtet (Tabelle 4-61).

Tabelle 4-64: Subgruppenergebnisse für die UE, die zum Tod führten, für die gepoolte Auswertung der INPULSIS Studien

Endpunkt/ Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
UE, die zum Tod führten					
Alter p-Wert Interaktionsterm^b: 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)		
Raucherstatus p-Wert Interaktionsterm^b: 0,133					
Nicht-Raucher	Nintedanib	174	11 (6,3)	1,58	0,53; 4,67
	Placebo	122	5 (4,1)		
Ex-Raucher/ Raucher	Nintedanib	464	26 (5,6)	0,63	0,36; 1,10
	Placebo	301	26 (8,6)		
Nierenfunktionseinschränkung p-Wert Interaktionsterm^b: 0,149					
ja	Nintedanib	331	18 (5,4)	0,58	0,30; 1,11
	Placebo	244	22 (9,0)		
nein	Nintedanib	306	19 (6,2)	1,24	0,55; 2,81
	Placebo	178	9 (5,1)		
a: Anzahl Studienteilnehmer in Subgruppe b: berechnet mittels Breslow-Day Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 2.1.1.1.1					

Bezüglich der UE, die zum Tod führten, fand sich in der Subgruppenanalyse nach *Alter* ein Beleg für eine Effektmodifikation (Tabelle 4-64). Da die Punktschätzer für die verschiedenen Alters-Strata teilweise unter 1 und teilweise über 1 liegen, liegt hier eine qualitative Effektmodifikation vor. In der Subgruppe der über 75-jährigen Patienten fand 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. Auch in der jüngsten

Subgruppe fand sich ein OR unter 1 (0,77). In der mittleren Altersgruppe fand sich hingegen ein OR für *UE, die zum Tod führten*, von 1,37. In den beiden jüngeren Altersgruppen isoliert betrachtet lag somit kein statistisch signifikanter Unterschied zwischen einer Behandlung mit Nintedanib oder Placebo vor.

Die absoluten Zahlen der Todesfälle, auf denen der statistisch signifikante Vorteil für Nintedanib in der höchsten Altersgruppe beruht, ist mit 5 versus 10 Fällen niedrig, sodass Schlussfolgerungen aus diesem Befund zurückhaltend gezogen werden sollten (Tabelle 4-64). Es fällt jedoch auf, dass die Gruppe der Patienten ≥ 75 Jahre unter einer Therapie mit Nintedanib einerseits die höchste Rate an *UE, die zum Therapieabbruch führten*, andererseits jedoch eine statistisch signifikant niedrigere Rate an *UE, die zum Tod führen*, aufweist. Pharmakokinetische Untersuchungen dokumentierten, dass unter einer Behandlung mit Nintedanib tendenziell höhere Plasmaspiegel in dieser Altersgruppe auftraten (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015b). Diese könnten sowohl häufigere Therapieabbrüche als auch eine niedrigere Mortalität unter der Therapie erklären.

Bezüglich des Subgruppenmerkmals *Raucherstatus* fand sich für *UE, die zum Tod führten*, ein Hinweis auf eine Effektmodifikation (Tabelle 4-64). Bei *Ex-Rauchern/Rauchern* fand sich unter Nintedanib verglichen mit Placebo ein OR für solche Ereignisse < 1 , während bei *Nichtrauchern* das OR > 1 lag. Somit weist der Interaktionstest auf eine qualitative Effektmodifikation durch das Merkmal *Raucherstatus* hin, wobei keine Subgruppe definiert werden kann, für die isoliert betrachtet ein statistisch signifikanter Effekt dokumentiert wurde.

Für das Subgruppenmerkmal *Nierenfunktionseinschränkung* (ja: Kreatinin-Clearance < 90 ml/min; nein: Kreatinin-Clearance ≥ 90 ml/min) zeigt der Endpunkt *UE, die zum Tod führten*, einen Hinweis auf eine Effektmodifikation (Tabelle 4-64). Bei Patienten, die bei Studienbeginn eine *Nierenfunktionseinschränkung* aufwiesen, lag das OR < 1 , wohingegen bei den Patienten mit *normaler Nierenfunktion* das OR > 1 lag. Da in beiden Strata kein statistisch signifikanter Effekt gezeigt wird, ergibt sich somit ein Hinweis auf eine quantitative Effektmodifikation, ohne dass eine Subgruppe mit einem statistisch signifikanten Vor- oder Nachteil abgeleitet werden kann.

4.3.1.3.5.4.4 Endpunkt *UE von besonderem Interesse*Tabelle 4-65: Übersicht zu den p-Werten der Interaktionstests für die Endpunkte zu *UE von besonderem Interesse*

Subgruppe	Diarrhoe	Übelkeit	Erbrechen	Appetit- verlust	kardiale Ereignisse	Progression IPF	Husten	Dyspnoe	Bronchitis	Naso- pharyngitis	Blutungen
Geschlecht	0,172	0,348	0,063	0,005	0,745	0,956	0,944	0,439	0,078	0,805	0,669
Alter	0,886	0,888	0,657	0,508	0,647	0,521	0,444	0,106	0,516	0,364	0,826
Ethnie	0,941	0,066	0,233	0,647	0,320	0,932	0,312	0,853	0,369	0,281	0,347
Nierenfunktions- einschränkung zur Baseline	0,196	0,843	0,222	0,808	0,960	0,420	0,739	0,756	0,436	0,459	0,038
Raucherstatus	0,008	0,473	0,623	0,045	0,206	0,822	0,335	0,717	0,770	0,897	0,284
Gewicht	0,699	0,325	0,690	0,154	0,477	0,274	0,657	0,525	0,915	0,270	0,823
Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f)											

Tabelle 4-66: Subgruppenergebnisse für die *UE von besonderem Interesse* für die gepoolte Auswertung der INPULSIS Studien

Endpunkt /Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Diarrhoe					
Geschlecht p-Wert Interaktionsterm ^b : 0,172					
männlich	Nintedanib	507	314 (61,9)	8,25	5,87; 11,60
	Placebo	334	55 (16,5)		
weiblich	Nintedanib	131	84 (64,1)	5,13	2,83; 9,29
	Placebo	89	23 (25,8)		
Nierenfunktionseinschränkung p-Wert Interaktionsterm ^b : 0,196					
ja	Nintedanib	331	206 (62,2)	6,24	4,26; 9,12
	Placebo	244	51 (20,9)		
nein	Nintedanib	306	191 (62,4)	9,29	5,80; 14,87
	Placebo	178	27 (15,2)		
Raucherstatus p-Wert Interaktionsterm ^b : 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	301	51 (16,9)		
Übelkeit					
Ethnie p-Wert Interaktionsterm ^b : 0,066					
kaukasisch	Nintedanib	360	97 (26,9)	3,61	2,21; 5,88
	Placebo	248	23 (9,3)		
asiatisch	Nintedanib	194	35 (18,0)	13,87	3,27; 58,77
	Placebo	128	2 (1,6)		
Erbrechen					
Geschlecht p-Wert Interaktionsterm ^b : 0,063					
männlich	Nintedanib	507	36 (7,1)	3,11	1,43; 6,79
	Placebo	334	8 (2,4)		
weiblich	Nintedanib	131	38 (29,0)	11,71	3,49; 39,34
	Placebo	89	3 (3,4)		

Endpunkt /Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Appetitverlust					
Geschlecht p-Wert Interaktionsterm ^b : 0,063					
männlich	Nintedanib	507	47 (9,3)	1,38	0,82; 2,32
	Placebo	334	23 (6,9)		
weiblich	Nintedanib	131	21 (16,0)	16,80	2,22; 127,35
	Placebo	89			
Raucherstatus p-Wert Interaktionsterm ^b : 0,045					
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	301	22 (7,3)		
Gewicht p-Wert Interaktionsterm ^b : 0,154					
< 65 kg	Nintedanib	126	19 (15,1)	4,32	1,23; 15,13
	Placebo	76	3 (3,9)		
≥ 65 kg	Nintedanib	512	49 (9,6)	1,64	0,97; 2,79
	Placebo	347	21 (6,1)		
Dyspnoe					
Alter p-Wert Interaktionsterm ^b : 0,106					
< 65 kg	Nintedanib	258	18 (7,0)	1,13	0,50; 2,59
	Placebo	145	9 (6,2)		
65-74 Jahre	Nintedanib	263	19 (7,2)	0,43	0,24; 0,78
	Placebo	216	33 (15,3)		
≥ 65 kg	Nintedanib	117	12 (10,3)	1,07	0,38; 2,99
	Placebo	62	6 (9,7)		
Bronchitis					
Geschlecht p-Wert Interaktionsterm ^b : 0,078					
männlich	Nintedanib	507	53 (10,5)	1,23	0,76; 1,98
	Placebo	334	29 (8,7)		
weiblich	Nintedanib	131	14 (10,7)	0,55	0,25; 1,18
	Placebo	89	16 (18,0)		

Endpunkt /Subgruppe	Behandlung	N ^a	Anzahl Ereignisse N(%)	Behandlungseffekt Nintedanib vs. Placebo	
				OR	95 % KI
Blutungen					
Geschlecht p-Wert Interaktionsterm ^b : 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)		
a: Anzahl Studienteilnehmer in Subgruppe b: berechnet mittels Breslow-Day Test Quellen: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f): Tab. 2.1.1.1.1					

Für *Diarrhoe* fanden sich ein Beleg für eine Effektmodifikation durch das Subgruppenmerkmal *Raucherstatus* sowie Hinweise auf eine Effektmodifikation durch die Subgruppenmerkmale *Geschlecht* und *Nierenfunktionseinschränkung* (Tabelle 4-66). In diesen drei Subgruppenanalysen wiesen jedoch alle Strata für sich genommen ein OR > 1 auf, sodass jeweils ein statistisch signifikanter Nachteil anzunehmen ist und es sich insgesamt um quantitative Effektmodifikationen handelt.

Bezüglich *Übelkeit* zeigen die Subgruppenanalysen einen Hinweis auf eine quantitative Effektmodifikation durch das Subgruppenmerkmal *Ethnie* (Tabelle 4-66), wobei kaukasische und asiatische Patienten unter einer Behandlung mit Nintedanib statistisch signifikant häufiger *Übelkeit* aufwiesen als unter Placebo. Das OR lag bei asiatischen Patienten höher als bei kaukasischen Patienten.

Für *Erbrechen* fand sich ein Hinweis auf eine quantitative Effektmodifikation durch das Subgruppenmerkmal *Geschlecht* (Tabelle 4-66). Das OR bei Frauen war höher als bei Männern, wobei in beiden Subgruppen *Erbrechen* unter der Behandlung mit Nintedanib statistisch signifikant häufiger auftrat als unter Placebo.

Die Subgruppenanalysen für *Appetitlosigkeit* ergaben Belege für eine Effektmodifikation durch die Subgruppenmerkmale *Geschlecht* (Tabelle 4-66) und *Raucherstatus*. Darüber hinaus zeigt sich ein Hinweis auf eine Effektmodifikation durch das Subgruppenmerkmal *Gewicht*. Dabei wiesen die Subgruppen der *Männer*, der *Ex-Raucher/Raucher* und der Patienten mit *Gewicht* ≥ 65 kg jeweils für sich genommen keine statistisch signifikante Erhöhung des Auftretens dieses *UE* auf, während in den jeweils anderen Strata der entsprechenden Subgruppenmerkmale statistisch signifikante Erhöhungen des Auftretens dokumentiert wurden. Da dennoch jeweils der Punktschätzer für das OR > 1 lag, ist von einer quantitativen Effektmodifikation durch das *Geschlecht* und den *Raucherstatus* bzw. von einem Hinweis auf eine quantitative Effektmodifikation durch das *Gewicht* auszugehen.

Für *Dyspnoe* fand sich ein Hinweis auf eine Effektmodifikation durch das Subgruppenmerkmal *Alter* (Tabelle 4-66). In der Gruppe der 65- bis 74-jährigen Patienten lag ein statistisch signifikant geringerer Nachteil unter Nintedanib gegenüber Placebo vor. In den Gruppen der < 65-jährigen und der ≥ 75 -jährigen Patienten lagen die Punktschätzer für das OR > 1 , allerdings ohne signifikanten Unterschied zwischen den Behandlungsarmen. Hier liegt formal ein Hinweis auf eine qualitative Effektmodifikation vor.

Für *Bronchitis* ergab sich für das Subgruppenmerkmal *Geschlecht* ein Hinweis auf eine Effektmodifikation (Tabelle 4-66). Dabei fand sich unter den Frauen mit einem OR von 0,55 eine numerische Reduktion der entsprechenden Ereignisse, wohingegen diese bei Männern mit einem OR von 1,23 unter Nintedanib tendenziell etwas häufiger auftraten als unter Placebo. Somit ergibt sich ein Hinweis auf eine qualitative Effektmodifikation durch das *Geschlecht*.

Für *Blutungen* zeigte die Analyse einen Beleg für eine Effektmodifikation durch das Subgruppenmerkmal *Nierenfunktionseinschränkung* (Tabelle 4-66): Während bei Patienten mit einer *Nierenfunktionseinschränkung* bei Studienbeginn mit einem OR von 2,13 ein statistisch signifikanter Nachteil unter einer Therapie mit Nintedanib verglichen mit Placebo vorlag, wurde ein solcher Nachteil bei Patienten ohne *Nierenfunktionseinschränkung* mit einem OR von 0,84 nicht dokumentiert. Somit handelt es sich hier um eine qualitative Effektmodifikation.

Fazit

Bei den nach formalen Kriterien identifizierten Hinweisen auf bzw. Beleg für Effektmodifikationen hinsichtlich *UE* handelt es sich zum größten Teil um quantitative Effektmodifikationen, die nicht dazu führen, dass einzelne Subgruppen von einem für die Gesamtpopulation angenommenen Nachteil ausgenommen werden.

Eine Ausnahme bilden *UE, die zum Tod führten*: Hier fand sich ein Beleg für eine qualitative Effektmodifikation durch das Subgruppenmerkmal *Alter*, wobei die Analyse für die Altersgruppe der ≥ 75 -jährigen Patienten einen statistisch signifikant geringeren Nachteil unter der Behandlung mit Nintedanib versus Placebo ergab. Für *Raucherstatus* und *Nierenfunktionseinschränkung* fanden sich für *UE, die zum Tod führten*, Hinweise auf qualitative Effektmodifikationen, ohne dass jedoch für einzelne Subgruppen statistisch signifikante Effekte auffielen.

Für das *UE von besonderem Interesse Dyspnoe* fand sich ein Hinweis auf eine qualitative Effektmodifikation durch das Subgruppenmerkmal *Alter*, wobei die Subgruppe der 65- bis 74-jährigen Patienten einen statistisch signifikant geringeren Nachteil unter Nintedanib gegenüber Placebo aufwies. Bei *Bronchitis* (wurde ein Hinweis auf eine qualitative Effektmodifikation durch das Subgruppenmerkmal *Geschlecht* dokumentiert (OR < 1 für Frauen, OR > 1 für Männer), ohne dass jedoch für eine der beiden Subgruppen ein statistisch signifikanter Effekt im Gruppenvergleich resultierte. Bezüglich *Blutungen* zeigte sich ein Beleg für eine qualitative Effektmodifikation durch das Subgruppenmerkmal *Nierenfunktionseinschränkung*. Ein statistisch signifikanter Nachteil unter Nintedanib gegenüber Placebo fand sich ausschließlich in der Subgruppe der Patienten mit *Nierenfunktionseinschränkung*, während Patienten ohne eine *Nierenfunktionseinschränkung* mit einem OR von 0,84 nicht von einer erhöhten Rate an Blutungsereignissen betroffen waren.

4.3.1.3.5.5 Ländereffekte

Zentrums- und Ländereffekte wurden in den statistischen Analysemodellen der INPULSIS Studien gemäß Studienprotokoll nicht berücksichtigt.

Da es sich bei der IPF um eine seltene Erkrankung handelt, war die Anzahl an Studienzentren, die für die Rekrutierung in den INPULSIS-1 und INPULSIS-2 Studien herangezogen wurden relativ hoch (INPULSIS-1: 98 Zentren in 13 Ländern, INPULSIS-2: 107 Zentren in 17 Ländern). Die Anzahl der Studienteilnehmer je Studienzentrum war somit entsprechend gering. Eine Auswertung unter Berücksichtigung potenzieller Zentrumseffekte wurde vor diesem Hintergrund als methodisch nicht sinnvoll erachtet.

Ländereffekte wurden im Rahmen des Zulassungsverfahrens mit der EMA jedoch diskutiert, und zusätzlich zu den a priori geplanten Subgruppenanalysen zur Ethnizität der Studienteilnehmer (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f) wurden weitere länderabhängige Auswertungen zum primären Endpunkt *jährliche FVC-Abnahme*, zur *Veränderung des SGRQ* und zur *Inzidenz akuter Exazerbationen* Post hoc durchgeführt (European Medicines Agency, 2014).

Zusammenfassend kann festgehalten werden, dass die im Rahmen der EMA-Bewertung durchgeführten Post hoc-Analysen für alle drei Endpunkte keinen Hinweis auf eine Effektmodifikation durch die der Bewertung zugrunde liegenden Länder aufweisen. Zusätzlich wird darauf hingewiesen, dass auch die Subgruppen zur Ethnizität keinen Hinweis auf eine Effektmodifikation zeigten (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014f). Daher wurde an dieser Stelle auf zusätzliche Auswertungen in Form der gemäß Verfo des G-BA vorgesehenen Interaktionstests verzichtet.

4.3.1.3.6 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen aus randomisierten kontrollierten Studien geben. Die Zusammenfassung soll Aussagen zu allen in Abschnitt 4.3.1.3 präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Meta-Analysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen aus randomisierten kontrollierten Studien zusammen.

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden) (European Commission (EC)). Somit gilt der medizinische Zusatznutzen von Nintedanib gemäß der gesetzlichen Vorgaben (§ 35a Abs. 1 Satz 10 SGB V) mit der Zulassung als belegt. Daher wird im Dossier insgesamt wie auch in der folgenden Zusammenfassung der Ergebnisse aus den berichteten RCTs explizit das Ausmaß des Zusatznutzens betrachtet.

Mortalität

Die Überlebenswahrscheinlichkeit, dargestellt durch die Endpunkte *Zeit bis Tod, unabhängig von der Ursache* und *Zeit bis respiratorisch bedingtem Tod*, ist unter Nintedanib höher als unter Placebo. Die Kaplan-Meier-Kurven zeigen, dass mit zunehmender Beobachtungszeit die Kurvenverläufe der verstorbenen Studienteilnehmer der beiden Behandlungsgruppen zugunsten von Nintedanib auseinandergehen (Abschnitt 4.3.1.3.1.1). Im Studienverlauf hatten Nintedanib-behandelte Studienteilnehmer eine geringere *jährliche Abnahme der FVC* als Placebo-behandelte Studienteilnehmer. Ebenfalls war der Anteil *FVC-Responder* unter Nintedanib höher als unter Placebo. Die genannten Ergebnisse zeigen Unterschiede mit statistisch signifikantem Effekt zugunsten von Nintedanib. Dies deutet ebenfalls auf eine Verlängerung des Gesamtüberlebens bei einer Behandlung von IPF-Patienten mit Nintedanib hin (Abschnitt 4.3.1.3.1.2).

Zeit bis zum Tod

Die *Zeit bis zum Tod, unabhängig von der Ursache* wurde über die Überlebenswahrscheinlichkeit, berechnet anhand einer Cox-Regression. Diese ist unter Nintedanib höher als unter Placebo (HR [95 % KI]: 0,70 [0,43; 1,12]; $p = 0,1399$) (Tabelle 4-14). Ein vergleichbarer Vorteil wird auch für die *Zeit bis respiratorisch bedingtem Tod* beobachtet (HR [95 % KI]: 0,74 [0,41; 1,34]; $p = 0,3435$). Die Kaplan-Meier-Kurven zeigen jedoch deutlich, dass mit zunehmender Beobachtungszeit die Kurvenverläufe der verstorbenen Studienteilnehmer zwischen den Behandlungsgruppen Nintedanib und Placebo auseinandergehen (Abbildung 4-6, Abbildung 4-7).

Forcierte Vitalkapazität

Die *jährliche FVC-Abnahme* (ml) ist in der gepoolten Analyse der INPULSIS Studien unter Nintedanib mit 113,59 ml um ca. 50 % geringer als unter Placebo (223,53 ml) mit klinisch relevantem und statistisch signifikantem Effekt. Somit blieb bei Studienteilnehmern, welche mit Nintedanib behandelt wurden, verglichen mit Studienteilnehmern, die Placebo erhielten, die forcierte Vitalkapazität signifikant länger erhalten. Der Unterschied in der adjustierten Mittelwertdifferenz ist zwischen den beiden Behandlungsgruppen mit 109,94 ml (MWD [95 % KI]: 109,94 ml [75,85; 144,03]; $p < 0001$) ebenfalls statistisch signifikant (Tabelle 4-17).

Die Responder-Analyse definierte Studienteilnehmer mit einem FVC-Schwellenwert $< 10\%$. Unter Nintedanib sind mehr Studienteilnehmer *FVC-Responder* als unter Placebo. Das OR beträgt 1,577 (OR [95 % KI]: 1,577 [1,21; 2,05]; $p < 0,0007$) zugunsten von Nintedanib mit einem statistisch signifikanten Effekt zwischen den Behandlungsgruppen (Tabelle 4-18).

Die Responder-Analyse mit einem FVC-Schwellenwert von 5 % zeigt ebenfalls einen statistisch signifikanten Unterschied zwischen den Behandlungsgruppen. Auf Nintedanib sprachen gemäß dieser Analyse mehr Studienteilnehmer an als auf Placebo (OR [95 % KI]: 1,84 [1,43; 2,36]; $p < 0,0001$) (Tabelle 4-18).

Morbidität

Die Ergebnisse aller Endpunkte zur Morbidität zeigen einen Vorteil der Nintedanib-Therapie im Vergleich zu Placebo. Speziell für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* ist der Unterschied statistisch signifikant zugunsten von Studienteilnehmern unter der Behandlung mit Nintedanib (Abschnitt 4.3.1.3.2.1).

Exazerbationen

Adjudizierte akute Exazerbationen traten in den gepoolten INPULSIS Studien unter Nintedanib seltener auf als unter Placebo (1,9 % vs. 5,7 %). Die Wahrscheinlichkeit, keine Exazerbation zu erleiden (ausgedrückt durch den Kaplan-Meier-Schätzer), beträgt in der Nintedanib-Gruppe 0,980 und in der Placebo-Gruppe 0,940. Auch das HR von 0,32 (HR [95 % KI]: 0,32 [0,16; 0,65]; $p = 0,0010$) für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* zeigt einen Vorteil zugunsten von Nintedanib mit statistisch signifikantem Effekt (Tabelle 4-21). Dies entspricht einer relativen Risikoreduktion von 68 % für das Auftreten *adjudizierter akuter Exazerbationen* unter der Behandlung mit Nintedanib.

Auch der Endpunkt *Zeit bis zur ersten akuten Exazerbation*, bewertet durch den Prüfarzt, zeigt für die gepoolte Auswertung der INPULSIS Studien (HR [95 % KI]: 0,64 [0,39; 1,05]; $p = 0,0823$) einen Vorteil zugunsten von Nintedanib (Tabelle 4-21). Auch für diesen Endpunkt zeigen die Kaplan-Meier-Kurven deutlich, dass mit zunehmender Beobachtungszeit die Kurven hinsichtlich der *Zeit bis zur ersten akuten Exazerbation* zwischen den Behandlungsgruppen Nintedanib und Placebo kontinuierlich auseinandergehen (Abbildung 4-10).

Gesundheitsbezogene Lebensqualität

Die Ergebnisse aller Endpunkte zur Lebensqualität zeigen einen Vorteil in Form einer geringeren Verschlechterung der Lebensqualität innerhalb des Studienzeitraums unter Nintedanib im Vergleich zu Placebo. Für die *Veränderung des SGRQ-I* und die *PGI-C-Responder* war der Unterschied statistisch signifikant (Abschnitt 4.3.1.3.3).

Endpunkte zum SGRQ und SGRQ-I

Die *Veränderung des SGRQ-Gesamtscores* über den Studienverlauf zeigt bei der gepoolten Auswertung der INPULSIS Studien einen Vorteil von Nintedanib gegenüber Placebo hinsichtlich der subjektiven krankheitsspezifischen Lebensqualität (MWD [95 % KI]: -1,43 Punkte [-3,09; 0,23]; $p = 0,092$), jedoch ohne statistische Signifikanz. (Tabelle 4-24).

Die Veränderung der Lebensqualität von Baseline bis Woche 52, ausgedrückt durch eine Zunahme des *SGRQ-I-Gesamtscores*, war in den gepoolten INPULSIS Studien unter Nintedanib insgesamt geringer als bei den Placebo-behandelten Studienteilnehmern (3,48 Punkte vs. 5,84 Punkte). Dies entspricht einem Vorteil hinsichtlich einer besseren subjektiv empfundenen Lebensqualität für Nintedanib-Behandelte im Vergleich zu Studienteilnehmern der Placebo-Gruppe. Der beobachtete Unterschied zwischen den Behandlungsgruppen ist statistisch signifikant (MWD [95 % KI]: -1,99 Punkte [-3,71; -0,27]; $p = 0,0230$) (Tabelle 4-24).

Die dichotomisierte Analyse der *SGRQ-(≤ -4 Punkte)-Responder* der gepoolten INPULSIS Studien zeigt, dass unter Nintedanib mehr Studienteilnehmer eine klinisch relevante Verbesserung der Lebensqualität angeben als unter Placebo (22,88 % vs. 20,33 % Responder) (OR [95 % KI]: 1,186 [0,88; 1,61]; $p = 0,2715$) (Tabelle 4-25).

Die Analyse der *SGRQ-I-(≤ -4 Punkte)-Responder* der gepoolten INPULSIS Studien zeigt, dass unter der Behandlung mit Nintedanib ebenfalls mehr Studienteilnehmer eine klinisch relevante Verbesserung der Lebensqualität angeben als unter Placebo (23,51% vs. 19,62 %) (OR [95 % KI]: 1,186 [0,88; 1,61]; $p = 0,2715$) (Tabelle 4-25).

Weitere Endpunkte zur krankheitsspezifischen Lebensqualität

Die Veränderung des *SOBQ-Score*, der die subjektiv empfundene krankheitsspezifische Lebensqualität hinsichtlich Kurzatmigkeit misst, zeigt eine geringere Verschlechterung unter Nintedanib im Vergleich zu Placebo in den gepoolten INPULSIS Studien (6,71 Punkte vs. 8,40 Punkte; ein höherer Wert entspricht einer stärkeren Verschlechterung der Kurzatmigkeit). Der Unterschied ist nicht statistisch signifikant (MWD [95 % KI]: -1,69 Punkte [-4,09; 0,70]; $p = 0,165$) (Tabelle 4-26).

Auch die Ergebnisse zu den *CASA-Q-Scores Husten-Symptome* und *Husten Impact* aus den gepoolten INPULSIS Studien zeigen eine geringere Verschlechterung der subjektiv empfundenen Lebensqualität hinsichtlich des Symptoms *Husten* und dessen Impact (*Symptome*: MWD^o[95 % KI]: 0,95 Punkte [-1,49; 3,38]; $p = 0,445$); *Impact*: (MWD [95 % KI]: 1,73 Punkte [-0,46; 3,92]; $p = 0,121$). Dieser Unterschied ist nicht statistisch signifikant, bestätigt jedoch in der Tendenz die Ergebnisse der anderen Auswertungen zur krankheitsspezifischen Lebensqualität (Tabelle 4-27).

Endpunkte zur allgemeinen Lebensqualität

Der Anteil *PGI-C-Responder*, definiert als Studienteilnehmer, die ihre gesundheitsbezogene Lebensqualität in Woche 52 im Vergleich zur Baseline als „sehr viel besser“, „viel besser“, „ein wenig besser“ oder „ohne Veränderung“ einschätzen, ist in den gepoolten INPULSIS Studien unter Nintedanib höher als unter Placebo (61,29 % vs. 54,37 %). Der Unterschied ist statistisch signifikant (OR [95 % KI]: 1,328 [1,04; 1,70]; $p = 0,025$) (Tabelle 4-28).

Auch die *Veränderung der EQ-5D-VAS* von Baseline bis Woche 52 zeige in den gepoolten INPULSIS Studien eine geringere Verschlechterung der Lebensqualität unter Nintedanib im Vergleich zu Placebo (1,69 Punkte vs. 5,73 Punkte). Eine vergleichende Statistik wurde gemäß Studienprotokoll nicht angewendet, so dass eine Aussage zur Signifikanz dieses Unterschieds nicht möglich ist (Tabelle 4-29).

Unerwünschte Ereignisse

Nintedanib ist ein Wirkstoff, welcher neben einer sehr guten Effektivität zur Behandlung der IPF im Allgemeinen ein gutes Verträglichkeitsprofil zeigt. Gastrointestinale Ereignisse traten vermehrt auf, dabei handelte es sich in den meisten Fällen um milde bis moderate Ereignisse, die nur mit geringer bis mäßiger Beeinträchtigung der Studienteilnehmer verbunden waren und durch Dosisreduktion oder medikamentöse Behandlungen gut handhabbar waren. Insgesamt brachen in den gepoolten INPULSIS Studien 19,3 % der Studienteilnehmer unter Nintedanib im Vergleich zu 13,0 % unter Placebo die Teilnahme wegen unerwünschter Ereignisse ab (Abschnitt 4.3.1.3.4).

Gesamtzahl UE

UE traten in der gepoolten Analyse der INPULSIS-Studien in der Nintedanib-Gruppe mit 95,5 % der Studienteilnehmer statistisch signifikant häufiger auf als in der Placebo-Gruppe (89,6 %) (OR [95 % KI]: 2,44 [1,50; 3,96]; $p < 0,001$). Hierbei handelte es sich allerdings um Ereignisse von heterogener klinischer Relevanz.

Schwerwiegende UE

SUE traten in der gepoolten Analyse der INPULSIS Studien bei 30,4 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf und bei 30 % der Placebo-Studienteilnehmer (Tabelle 4-35). Der Unterschied zwischen den Gruppen war nicht statistisch signifikant (OR [95 % KI]: 1,02 [0,78; 1,33]; $p = 0,894$).

UE, die zum Therapieabbruch führten

UE, die zum Therapieabbruch führten, traten in der gepoolten Analyse der INPULSIS Studien bei 19,3 % der Studienteilnehmer, die mit Nintedanib behandelt wurden verglichen mit 13,0 % der Placebo-behandelten Studienteilnehmer auf (Tabelle 4-38). Der Unterschied zwischen den Gruppen ist statistisch signifikant (OR [95 % KI]: 1,60 [1,13; 2,26]; $p = 0,007$). Somit brachen in der Nintedanib-Gruppe häufiger Studienteilnehmer die Behandlung aufgrund eines UE ab als in der Placebo-Gruppe.

UE, die zum Tod führten

Ein UE, das zum Tod führte, trat in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien seltener unter der Behandlung mit Nintedanib im Vergleich zu Placebo-behandelten Studienteilnehmern auf (5,8 % vs. 7,3 %) (Tabelle 4-41). Der Unterschied zwischen den Gruppen war nicht statistisch signifikant (OR [95 % KI]: 0,78 [0,48; 1,28]; $p = 0,319$). Somit verstarb in der Nintedanib-Gruppe numerisch seltener ein Patient infolge eines UE als in der Placebo-Gruppe.

UE von besonderem Interesse

Gastrointestinale Störungen

Diarrhoe

Häufigste *gastrointestinale UE* waren *Diarrhoen*, die in den gepoolten INPULSIS Studien unter Nintedanib häufiger auftraten als unter Placebo (62,4 % vs. 18,4 %) mit statistisch signifikantem Unterschied ($p < 0,001$) (Tabelle 4-44). *Diarrhoen* traten überwiegend in den ersten drei Monaten der Behandlung auf und waren bei den meisten Studienteilnehmern von leichter bis moderater Intensität. Zu einer Einschränkung in der Ausübung von Alltagsaktivitäten oder zu Arbeitsunfähigkeit (Intensität „schwer“) kam es bei 5 % der Nintedanib-Studienteilnehmer mit *Diarrhoe* (entsprechend 3,3 % aller Studienteilnehmer der Nintedanib-Gruppe). 78,6 % der Studienteilnehmer mit *Diarrhoe* konnten die Therapie mit der Normaldosis fortsetzen, weitere 14,3 % der Studienteilnehmer mit reduzierter Dosis. Lediglich bei 7,0 % der Studienteilnehmer mit *Diarrhoe* (entsprechend 4,4 % aller Studienteilnehmer der Nintedanib-Gruppe) kam es aufgrund dieses *UE* zu einem endgültigen Therapieabbruch (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014b).

Übelkeit

Übelkeit wurde in den gepoolten INPULSIS Studien unter der Behandlung mit Nintedanib häufiger berichtet als bei Placebo (24,5 % vs. 6,6 %). Der Unterschied ist statistisch signifikant (OR [95 % KI]: 4,57 [2,99; 6,97]; $p < 0,001$) (Tabelle 4-44).

Erbrechen

Auch *Erbrechen* wurde im Nintedanib-Behandlungsarm (11,6 %) häufiger berichtet als im Placebo-Behandlungsarm (2,6 %). Der beobachtete Effekt ist statistisch signifikant (OR [95 % KI]: 4,91 [2,58; 9,38]; $p < 0,001$) (Tabelle 4-44).

Appetitlosigkeit

Appetitlosigkeit wurde in den INPULSIS Studien im Nintedanib-Behandlungsarm (10,7 %) häufiger dokumentiert als im Placebo-Behandlungsarm (5,7 %). Der Unterschied ist statistisch signifikant (OR [95 % KI]: 1,98 [1,22; 3,21]; $p = 0,005$) (Tabelle 4-44).

Gewichtsverlust

Gewichtsverlust trat im Nintedanib-Behandlungsarm (9,7 %) häufiger auf als im Placebo-Behandlungsarm (3,5 %). Der beobachtete Effekt ist statistisch signifikant (OR [95 % KI]: 2,93 [1,64; 5,22]; $p < 0,001$) (Tabelle 4-44).

Gastrointestinale Perforation

Gastrointestinale Perforationen traten insgesamt so selten auf, dass keine Aussage zu einem möglichen Risiko vorgenommen werden kann (Tabelle 4-44).

Unerwünschte Leberenzym- und Bilirubinerhöhungen

UE, welche das Leber- und Gallenwegssystem betrafen, wurden in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien häufiger unter Nintedanib als unter Placebo dokumentiert (Tabelle 4-45). Die Behandlung mit Nintedanib war im Vergleich zu Placebo häufiger verbunden mit einer Erhöhung der Leberenzyme (ALT, AST, ALKP) oder des Bilirubin-Wertes.

- AST- und ALT-Werte um mindestens 3x ULN erhöht (Nintedanib vs. Placebo): 5 % vs. 0,7 % (OR [95 % KI]: 7,39 [2,25; 24,3]; $p < 0,001$)
- AST- und ALT-Werte um mindestens 5x ULN erhöht (Nintedanib vs. Placebo): 2,2 % vs. 0,2 % (OR [95 % KI]: 9,47 [1,24; 72,3]; $p = 0,008$)
- AST- und ALT-Werte um mindestens 8x ULN erhöht (Nintedanib vs. Placebo): 0,8 % vs. 0,2 % (OR [95 % KI]: 3,33 [0,39; 28,6]; $p = 0,244$)
- Bilirubinwert um mindestens 1,5x ULN erhöht (Nintedanib vs. Placebo): 2,4 % vs. 0,7 % (OR [95 % KI]: 3,37 [0,97; 11,7]; $p = 0,043$)
- Bilirubinwert um mindestens 2x ULN erhöht (Nintedanib vs. Placebo): 0,5 % vs. 0,5 % (OR [95 % KI]: 0,99 [0,17; 5,98]; $p = 0,995$)
- ALKP-Wert um mindestens 1,5x ULN erhöht (Nintedanib vs. Placebo): 5,8 % vs. 0,9 % (OR [95 % KI]: 6,45 [2,28; 18,2]; $p < 0,001$)
- ALKP-Wert um mindestens 2x ULN erhöht (Nintedanib vs. Placebo): 2,7 % vs. 0,2 % (OR [95 % KI]: 11,6 [1,53; 87,1]; $p = 0,003$)

Unerwünschte kardiale Ereignisse

Unerwünschte kardiale Ereignisse wurden in den gepoolten INPULSIS Studien vergleichbar häufig unter Nintedanib und Placebo (10,0 % vs. 10,6 %) beobachtet (Tabelle 4-46). Auch *schwerwiegende unerwünschte kardiale Ereignisse* traten in der Nintedanib-Gruppe (5,0 %) und Placebo-Gruppe (5,4 %) vergleichbar häufig auf. *Tödliche kardiale Ereignisse* traten numerisch seltener in der Nintedanib-Gruppe (0,5 %) als in der Placebo-Gruppe (1,4 %) auf (OR [95 % KI]: 0,33 [0,08; 1,32]; $p = 0,099$).

Arterielle thromboembolische Ereignisse wurden in den gepoolten INPULSIS Studien häufiger unter Nintedanib als unter Placebo (2,5 % vs. 0,7 %) beobachtet (Tabelle 4-46). *Myokardinfarkte* traten numerisch häufiger unter Nintedanib als unter Placebo (1,6 % vs. 0,5 %) auf.

Unerwünschte respiratorische Störungen

Respiratorische Störungen traten in der gepoolten Analyse der INPULSIS Studien seltener unter der Behandlung mit Nintedanib auf im Vergleich zu Placebo (3,1 % vs. 6,6%) (Tabelle 4-47). Unter *respiratorischen Störungen* sind z. B. *Progression der IPF*, *Husten*, *Dyspnoe*, *Bronchitis* etc. subsummiert. Dabei wurde eine *Progression der IPF* als häufigstes Ereignis in den der Bewertung zugrunde liegenden Studien dokumentiert.

Unerwünschte Blutungsereignisse

Blutungsereignisse fanden sich bei 10,3 % der Nintedanib- und 7,8 % der Placebo-behandelten Patienten (Tabelle 4-48). Der Gruppenunterschied war nicht statistisch signifikant. *Nicht schwerwiegende Epistaxis* und *Kontusionen* stellten die häufigsten *Blutungsereignisse* dar. Die Häufigkeit *schwerwiegender Blutungsereignisse* war in beiden Behandlungsgruppen gering und vergleichbar (Placebo: 1,4 %; Nintedanib: 1,3 %).

Unerwünschte Ereignisse >°10 %

Es wurden keine *UE* identifiziert, die nicht bereits in einer der oben genannten Kategorien nach Organsystemen von besonderem Interesse berücksichtigt wurden.

Subgruppenanalysen

Es kann von einem konstanten Therapieeffekt im Hinblick auf die dargestellten Endpunkte über die analysierten Subgruppen hinweg ausgegangen werden (Abschnitt 4.3.1.3.5).

4.3.2 Weitere Unterlagen

4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn indirekte Vergleiche als Nachweis für einen Zusatznutzen herangezogen werden sollen. Das ist dann möglich, wenn keine direkten Vergleichsstudien für das zu bewertende Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu Studien für indirekte Vergleiche. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie

- Studien des pharmazeutischen Unternehmers
- Studien aus der bibliografischen Literaturrecherche
- Studien aus der Suche in Studienregistern
- Resultierender Studienpool aus den einzelnen Suchschritten

Nicht zutreffend

4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche

Charakterisieren Sie nachfolgend die Studien, die für indirekte Vergleiche herangezogen wurden, und bewerten Sie deren Verzerrungspotenzial. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

Nicht zutreffend

4.3.2.1.3 Ergebnisse aus indirekten Vergleichen

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus indirekten Vergleichen beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-67: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

Studie	<Mortalität>	<Gesundheits- bezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
<Studie 1>	nein	ja	ja	ja	nein

4.3.2.1.3.1 <Endpunkt xxx> – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. **Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.1.**

Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche. Beispielhaft wäre folgende Darstellung denkbar:

Tabelle 4-68: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Studie	Intervention	<Vergleichs- therapie 1>	<Vergleichs- therapie 2>	<Vergleichs- therapie 3>
1	<Studie 1>	•		•	•
2	<Studie 2>	•		•	
	<Studie 3>	•		•	
1	<Studie 4>		•	•	•
etc.	etc.	etc.	etc.		

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Nicht zutreffend

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-69: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-70: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>

Begründen Sie für jede Studie die abschließende Einschätzung.

Nicht zutreffend

Stellen Sie die Ergebnisse für den Endpunkt xxx für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-71: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche

Studie	Tabellarische Präsentation in geeigneter Form (Anforderungen siehe Erläuterung in Abschnitt 4.3.1.3.1)
<Studie 1>	

Nicht zutreffend

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt, für den ein indirekter Vergleich vorgenommen wird, fortlaufend in einem eigenen Abschnitt dar.

4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.5.

Nicht zutreffend

4.3.2.2 Nicht randomisierte vergleichende Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn nicht randomisierte vergleichende Studien als Nachweis für einen Zusatznutzen herangezogen werden sollen.

4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie

- Studien des pharmazeutischen Unternehmers
- Studien aus der bibliografischen Literaturrecherche
- Studien aus der Suche in Studienregistern
- Resultierender Studienpool aus den einzelnen Suchschritten

Nicht zutreffend

4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien

Charakterisieren Sie nachfolgend die nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2. und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Beschreibung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-72: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
			Patient	Behandelnde Personen		
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Studienebene.

Nicht zutreffend

4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-73: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Beschreiben Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-F.

Tabelle 4-74: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Studie	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
<Studie 1>	ja / nein / unklar	ja / nein / unklar	ja / nein / unklar	ja / nein

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Endpunktebene.

Nicht zutreffend

Stellen Sie die Ergebnisse der nicht randomisierten vergleichenden Studien gemäß den Anforderungen des TREND- bzw. des STROBE-Statements dar. Machen Sie dabei auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus nicht randomisierten vergleichenden Studien fortlaufend in einem eigenen Abschnitt dar.

4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien

*Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus nicht randomisierten vergleichenden Studien. **Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.5.***

Nicht zutreffend

4.3.2.3 Weitere Untersuchungen

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn über die in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 genannten Studien hinausgehende Untersuchungen als Nachweis für einen Zusatznutzen herangezogen werden sollen.

4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie

- Studien des pharmazeutischen Unternehmers
- Studien aus der bibliografischen Literaturrecherche
- Studien aus der Suche in Studienregistern
- Resultierender Studienpool aus den einzelnen Suchschritten

Nicht zutreffend

4.3.2.3.2 Charakteristika der weiteren Untersuchungen

Charakterisieren Sie nachfolgend die weiteren Untersuchungen und bewerten Sie deren Verzerrungsaspekte.

Ergebnisse nicht randomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2. und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

Nicht zutreffend

4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen

4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-75: Operationalisierung von <Endpunkt xxx> - weitere Untersuchungen

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt. Ergebnisse nichtrandomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Nicht zutreffend

Stellen Sie die Ergebnisse der weiteren Untersuchungen gemäß den jeweils gültigen Standards für die Berichterstattung dar. Begründen Sie dabei die Auswahl des Standards für die Berichterstattung. Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

4.3.2.3.2 Subgruppenanalysen – weitere Untersuchungen

*Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. **Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.5.***

Nicht zutreffend

4.3.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse aus weiteren Unterlagen (Abschnitte 4.3.2.1, 4.3.2.2 und 4.3.2.3) geben. Die Zusammenfassung soll Aussagen zu allen in diesen Abschnitten präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Meta-Analysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse aus weiteren Unterlagen zusammen.

Nicht zutreffend

4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens

Zusammenfassung des Abschnitts 4.4

Mit den beiden Phase III-Studien (INPULSIS-1 und INPULSIS-2) liegen RCTs mit niedrigem Verzerrungspotenzial auf Studienebene vor, welche der höchsten Evidenzstufe zuzuordnen sind. Auf Basis der gepoolten Analyse der INPULSIS Studien lassen sich Hinweise auf einen **beträchtlichen** Zusatznutzen ableiten. Die patientenrelevanten Endpunkte weisen ein niedriges Verzerrungspotenzial auf.

Ein Zusatznutzen mit beträchtlichem Ausmaß wird nach AM-NutzenV dann ausgewiesen, wenn eine gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens im Sinne von § 2 Absatz 3 erreicht wird, insbesondere eine Abschwächung schwerwiegender Symptome, eine moderate Verlängerung der Lebensdauer, eine für die Patientinnen und Patienten spürbare Linderung der Erkrankung, eine relevante Vermeidung schwerwiegender Nebenwirkungen oder eine bedeutsame Vermeidung anderer Nebenwirkungen. Mit Nintedanib (Ofev[®]) wurde im Rahmen der oben beschriebenen RCTs eine klinisch relevante Reduktion des jährlichen FVC-Verlustes um ca. 50 % sowie eine signifikante Senkung des Risikos adjudizierter akuter Exazerbationen um 68 % erreicht. Hieraus ergibt sich ein Hinweis für einen beträchtlichen Zusatznutzen, der sich in der klinisch bedeutsamen Verbesserung des mortalitäts-assoziierten Endpunktes *FVC* sowie der Reduktion der *adjudizierten akuten Exazerbationen* bei einem guten Verträglichkeitsprofil begründet.

Fazit

Die Behandlung mit Nintedanib (Ofev[®]) stellt eine sehr wirksame und im Allgemeinen gut verträgliche Therapieoption für Patienten mit idiopathischer Lungenfibrose dar.

Mit der Verfügbarkeit des Tyrosinkinaseinhibitors Nintedanib (Ofev[®]) ist ein relevanter Fortschritt in der Behandlung von Patienten mit IPF erreicht worden. Darüber hinaus ist Nintedanib (Ofev[®]) derzeit das einzige Medikament, das in Deutschland zur Behandlung von IPF Patienten über alle Schweregrade hinweg zugelassen ist.

Inbesondere auch unter Berücksichtigung der limitierten Therapieoptionen ist der Zusatznutzen von Nintedanib (Ofev[®]) nach der oben dargestellten Anforderung der AM-NutzenV als **beträchtlich** einzustufen.

4.4.1 Beurteilung der Aussagekraft der Nachweise

Legen Sie für alle im Dossier eingereichten Unterlagen die Evidenzstufe dar. Beschreiben Sie zusammenfassend auf Basis der in den Abschnitten 4.3.1 und 4.3.2 präsentierten Ergebnisse die Aussagekraft der Nachweise für einen Zusatznutzen unter Berücksichtigung der Studienqualität, der Validität der herangezogenen Endpunkte sowie der Evidenzstufe.

Ofev[®] (Nintedanib) ist ein Arzneimittel zur Behandlung einer seltenen Erkrankung (Anerkennung des Orphan Drug Status¹ in der Europäischen Union (EU) am 19.01.2015 (EU/1/14/979) basierend auf Designierung EU/3/13/1123, nach der Verordnung (EG) Nr.141/2000 des Europäischen Parlaments und Rates am 16.12.1999 über Arzneimittel für seltene Leiden) (European Commission (EC)). Das Ausmaß des medizinischen Zusatznutzens ist nach § 5 Abs. 7 VerfO als Verbesserung des patientenrelevanten Nutzens im Sinne von § 3 Absatz 1 VerfO zu bestimmen. Die Bewertung des medizinischen Zusatznutzens ist bei Orphan Drugs ausschließlich auf Grundlage der Zulassungsstudien durchzuführen. Grundlage der Bewertung des medizinischen Zusatznutzens von Nintedanib im Anwendungsgebiet sind die Studien INPULSIS-1 und INPULSIS-2. Hierbei handelte es sich um multizentrische, doppelblinde, randomisierte und kontrollierte Phase III-Studien im Parallelgruppendesign, die die Wirksamkeit von täglich 2 x 150 mg Nintedanib gegenüber Placebo bei Studienteilnehmern mit IPF untersuchten.

Die INPULSIS Studien weisen sowohl endpunktübergreifend als auch in allen patientenrelevanten Endpunkten ein niedriges Verzerrungspotenzial und somit eine hohe Aussagesicherheit auf. Da es sich entsprechend um randomisierte, kontrollierte Studien (RCTs) handelt, werden die INPULSIS Studien dem höchsten Evidenzgrad der Evidenzklassifizierung des IQWiG (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2013) zugeordnet. Die vorhandene Evidenz ist daher methodisch geeignet, Belege für einen Zusatznutzen zu generieren. Es war a priori geplant, beide INPULSIS Studien auf Patientenebene gepoolt im Rahmen von IPD Meta-Analysen auszuwerten. Die Begründung für das Ausmaß des Zusatznutzens im vorliegenden Dossier wurde somit auf Basis dieser gepoolten Analysen abgeleitet. Entsprechend wird die Wahrscheinlichkeit für das Ausmaß des Zusatznutzens, gemäß dem Methodenpapier des IQWiG mit einem Hinweis eingeschätzt.

Die Untersuchung des Zusatznutzens wurde anhand der folgenden Endpunkte durchgeführt:

- Mortalität
 - *Zeit bis zum Tod (unabhängig von der Ursache bzw. respiratorisch bedingt)*
 - *Jährliche FVC-Abnahme (ml) und FVC-Responder*
- Morbidität
 - *Zeit bis zur ersten akuten Exazerbation (bewertet durch den Prüfarzt) und Zeit bis zur ersten adjudizierten akuten Exazerbation (bewertet durch ein unabhängiges Bewertungskomitee)*
- Gesundheitsbezogene Lebensqualität
 - *SGRQ- Veränderung und $-(\leq -4 \text{ Punkte})$ -Responder-Analysen*
 - *SGRQ-I-Veränderung und $-(\leq -4 \text{ Punkte})$ -Responder-Analysen*
 - *Veränderung des SOBQ*
 - *Veränderung des CASA-Q*
 - *PGI-C-Responder*
 - *Veränderung der EQ-5D-VAS*
- Unerwünschte Ereignisse
 - *Gesamtübersicht unerwünschter Ereignisse (UE, SUE, UEA, UE, die zum Tod führten)*
 - *UE von besonderem Interesse (gastrointestinale Störungen, Leberenzymwerte und Bilirubinerhöhungen, kardiovaskuläre Ereignisse, respiratorische Störungen, Blutungsereignisse, weitere häufige UE, > 10 %)*

Mortalität

Die Bestimmung der FVC ist ein validiertes Verfahren, welches die Lungenfunktion der IPF-Patienten analysiert. FVC ist ein wichtiger klinisch relevanter Parameter, der Auskunft hinsichtlich der klinischen Progredienz der IPF-Erkrankung und der Mortalität gibt (Abschnitt 4.5.4).

Morbidität

Die Endpunkte akute Exazerbationen und adjudizierte akute Exazerbationen sind insofern patientenrelevant, da Exazerbationen mit einem Krankheitsprogress und einer erhöhten Sterblichkeit einhergehen. Kriterien zur Diagnose akuter Exazerbationen waren im Studienprotokoll definiert.

Lebensqualität

Die gesundheitsbezogene Lebensqualität wurde mittels validierter Messinstrumente von den Studienteilnehmern direkt vor jedem Arztbesuch bewertet. Aufgrund der Verblindung der Studienteilnehmer und behandelnden Personen kann von einer unverzerrten Messung der Symptomatik und der HRQoL ausgegangen werden. Der Zusatznutzen für diese Endpunkte wurde zusätzlich anhand der Responderanalysen abgeleitet.

Unerwünschte Ereignisse

Die Auswertungskriterien aller Endpunkte zu den unerwünschten Ereignissen waren im Prüfplan a priori definiert. Die Ergebnisse wurden gemäß Prüfplan ausgewertet und transparent berichtet. Eine ergebnisgesteuerte Publikation kann daher ausgeschlossen werden. Alle UE wurden in den berücksichtigten Studien routinemäßig bei jeder Studienvsiste ab der Randomisierung bis 28 Tage nach der letzten Einnahme der Studienmedikation erfasst.

Subgruppenanalysen

Statistisch signifikante Unterschiede zwischen den Subgruppen wurden im Dossier beschrieben und diskutiert.

4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß

Führen Sie die in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse zum Zusatznutzen auf Ebene einzelner Endpunkte zusammen und leiten Sie ab, ob sich aus der Zusammenschau der Ergebnisse zu den einzelnen Endpunkten insgesamt ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie ergibt. Berücksichtigen Sie dabei auch die Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext. Liegt ein Zusatznutzen vor, beschreiben Sie, worin der Zusatznutzen besteht.

Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, d. h., beschreiben und begründen Sie unter Berücksichtigung der in Abschnitt 4.4 dargelegten Aussagekraft der Nachweise die Ergebnissicherheit der Aussage zum Zusatznutzen.

Beschreiben Sie außerdem das Ausmaß des Zusatznutzens unter Verwendung folgender Kategorisierung (in der Definition gemäß AM-NutzenV):

- *erheblicher Zusatznutzen*
- *beträchtlicher Zusatznutzen*
- *geringer Zusatznutzen*
- *nicht quantifizierbarer Zusatznutzen*
- *kein Zusatznutzen belegbar*
- *der Nutzen des zu bewertenden Arzneimittels ist geringer als der Nutzen der zweckmäßigen Vergleichstherapie*

Berücksichtigen Sie bei den Aussagen zum Zusatznutzen ggf. nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen.

Die Bewertung des medizinischen Zusatznutzens erfolgte in Bezug auf die patientenrelevanten Nutzendimensionen *Mortalität, Morbidität, gesundheitsbezogene Lebensqualität* und *unerwünschte Ereignisse*:

Die Phase III-Studien INPULSIS-1 und INPULSIS-2 haben einen homogenen Studienaufbau und sind für die gepoolte Auswertung geeignet bzw. waren für eine solche Auswertung bei der Studienplanung bereits vorgesehen. Auf Basis der gepoolten Analyse wird daher eine wesentlich höhere Ergebnissicherheit im Vergleich zu den einzelnen Studien erreicht. Deshalb wurde in der Ergebnisdarstellung und Zusatznutzenbewertung auf die Ergebnisse der gepoolten Auswertung fokussiert.

Beschreibung des Ausmaßes des Zusatznutzens

Mortalität

Die IPF ist eine chronische, rasch progrediente unheilbare Erkrankung. Die Mortalitätsrate ist dabei höher als bei den meisten Krebserkrankungen (Douglas et al., 2000). Ein wichtiges Therapieziel ist daher eine Verlängerung des Überlebens der Studienteilnehmer bei gleichzeitigem Erhalt der Lebensqualität.

Die Überlebenswahrscheinlichkeit ist unter Nintedanib höher als unter Placebo (Abbildung 4-1). Das HR für die *Zeit bis zum Tod unabhängig von der Ursache* beträgt 0,70, d. h. die Wahrscheinlichkeit im Verlauf des Beobachtungszeitraums zu versterben lag bei Nintedanib behandelten Studienteilnehmern um 30 % niedriger als für Studienteilnehmer im Placebo-Arm. Ein vergleichbares Bild zeigt sich auch bei der *Zeit bis respiratorisch bedingtem Tod*. Aufgrund der fehlenden statistischen Signifikanz wird von einem *nicht quantifizierbaren* Zusatznutzen für Nintedanib ausgegangen.

Der Anteil *FVC-Responder* (definiert als Anteil der Studienteilnehmer, deren Abfall der FVC weniger als 10 % (bzw. 5 %) des Sollwertes im beobachteten Zeitraum beträgt) war unter Nintedanib deutlich höher als unter Placebo (OR [95 %KI] 1,577 [1,21; 2,05]; p = 0,0007). Somit konnte gezeigt werden, dass bei Studienteilnehmern, die mit Nintedanib behandelt wurden, der Abfall der jährlichen FVC statistisch und klinisch signifikant aufgehalten werden kann. Die Ergebnisse der Studien zeigen ebenfalls, dass der Abfall der FVC um ca. 50 % geringer unter der Therapie mit Nintedanib ausfällt im Vergleich zu Placebo (Tabelle 4-76).

Der Zusatznutzen für den Endpunkt *FVC* kann vor diesem Hintergrund als *beträchtlich* und dessen Wahrscheinlichkeit als *Hinweis* eingestuft werden. Als validiertes Surrogat für die Mortalität deutet dies ebenfalls auf eine Verlängerung des Gesamtüberlebens bei Behandlung von IPF-Patienten mit Nintedanib hin.

Für die Mortalität ergibt sich hieraus ein Hinweis auf einen <i>beträchtlichen</i> Zusatznutzen.

Tabelle 4-76: Bewertung des Zusatznutzens mit Erläuterung - Mortalität

Ergebnisse aus gepoolten INPULSIS Studien	<p><u>Jährliche FVC- Abnahme (ml)</u> MWD [95 % KI]: 109,94 ml [75,85;144,03]; p = 0,0001</p> <p><u>FVC-Responder (10 % vom Sollwert)</u> OR^b [95 % KI]: 1,577 [1,21; 2,05]; p = 0,0007</p> <p><u>FVC-Responder (5 % vom Sollwert)</u> OR^b [95 % KI]: 1,835 [1,43; 2,36]; p < 0,0001</p>
Erläuterungen	Nach dem SGB V – beispielsweise in der AM-NutzenV operationalisiert – wird ein patientenrelevanter Nutzen durch Verlängerung des Überlebens bestimmt.
Bewertung des Zusatznutzens	<p>Bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens im Sinne einer moderaten Verringerung in der Abnahme der FVC als Prädiktor für ein verringertes Mortalitätsrisiko.</p> <ul style="list-style-type: none"> • Hinweis auf einen beträchtlichen Zusatznutzen von Nintedanib im Vergleich zu Placebo aufgrund einer bisher nicht erreichten Verbesserung des Überlebens der mit Nintedanib behandelten Patienten.
<p>a: Ein HR < 1 zeigt einen Behandlungseffekt zugunsten der Nintedanib-Therapie an. b: Ein OR > 1 zeigt einen Behandlungseffekt zugunsten der Nintedanib-Therapie an.</p>	

Morbidität

IPF ist eine Erkrankung mit einer hohen Krankheitslast. Die Krankheitssymptomatik ist insbesondere gekennzeichnet durch eine zunehmende Luftnot (Dyspnoe) der Patienten mit erheblichen Auswirkungen auf die körperliche Leistungs- und Bewegungsfähigkeit. Diese sind meist schleichend progredient. Hinzu können *akute Exazerbationen* (respiratorische Dekompensationen) kommen, welche akut lebensbedrohlich sind, einen maßgeblichen Einfluss auf die Progression der Erkrankung haben und zu einer weiteren rapiden Verschlechterung des Gesundheitszustands der Betroffenen führen. Solche Ereignisse sind mit einer hohen Sterblichkeit assoziiert (medianes Überleben nach *akuter Exazerbation* 2,2 Monate) (Song et al., 2011).

Adjudizierte akute Exazerbationen traten in den gepoolten INPULSIS Studien unter Nintedanib deutlich seltener auf als unter Placebo (1,9 % vs. 5,7 %). Das HR von 0,32 für die *Zeit bis zur ersten adjudizierten akuten Exazerbation* zeigt mit einer 68 %-igen Reduktion der Exazerbationen einen statistisch signifikanten Vorteil zugunsten der Behandlung mit Nintedanib (HR [95 % KI]: 0,32 [0,16; 0,65]; p = 0,0010) (Tabelle 4-77). Die Kaplan-Meier-Graphen zeigen ebenfalls, dass mit zunehmender Beobachtungszeit die Kurven für den Endpunkt *Zeit bis zur ersten adjudizierten akuten Exazerbation* zwischen den Behandlungsgruppen Nintedanib und Placebo kontinuierlich auseinandergehen.

Bezüglich der Morbidität ergibt sich ein Hinweis auf einen **beträchtlichen** Zusatznutzen von Nintedanib.

Tabelle 4-77: Bewertung des Zusatznutzens mit Erläuterung - Morbidität

Ergebnisse aus gepoolten INPULSIS Studien	<u>Zeit bis zur ersten adjudizierten Exazerbation</u> HR [95 % KI]: 0,32 [0,16; 0,65]; p = 0,0010
Erläuterungen	Nach dem SGB V – beispielsweise in der AM-NutzenV operationalisiert – wird ein patientenrelevanter Nutzen durch Verlängerung des Überlebens bestimmt. Vor dem Hintergrund, dass akute Exazerbationen, Ereignisse sind, die mit einer hohen Sterblichkeit assoziiert sind (medianes Überleben nach akuter Exazerbation: 2,2 Monate) wird durch die Vermeidung des Auftretens akuter Exazerbationen implizit eine Verlängerung des Überlebens erreicht (Song et al., 2011).
Bewertung des Zusatznutzens	<p>Die Ergebnisse <u>aller</u> Endpunkte zur Morbidität zeigen einen Vorteil der Nintedanib-Therapie im Vergleich zu Placebo.</p> <p>Ein beträchtlicher Zusatznutzen, wird vor allem aber basierend auf den Ergebnissen hinsichtlich der Vermeidung adjudizierter akuter Exazerbationen abgeleitet.</p> <p>Gemäß AM-Nutzen V liegt ein erheblicher Zusatznutzen vor, wenn unter anderem eine langfristige Freiheit von schwerwiegenden Symptomen oder eine weitgehende Vermeidung schwerwiegender Nebenwirkungen oder eine Verlängerung des Überlebens erreicht wird.</p> <ul style="list-style-type: none"> • Hinweis auf einen beträchtlichen Zusatznutzen von Nintedanib (Ofev®) im Vergleich zu Placebo bezüglich einer Verbesserung schwerwiegender Symptome oder eine weitgehende Vermeidung schwerwiegender Nebenwirkungen bzw. eine Verlängerung des Überlebens <p>Dies trifft hier aufgrund der erzielten Vermeidung der potenziell lebensbedrohlichen (adjudizierten) akuten Exazerbationen zu.</p>
a: Ein HR < 1 zeigt einen Behandlungseffekt zugunsten der Nintedanib-Therapie an.	

Gesundheitsbezogene Lebensqualität

IPF führt zu erheblichen Einschränkungen in der Lebensqualität betroffener Patienten. Symptome wie Dyspnoe, Erschöpfung sowie eingeschränkte Mobilität beeinträchtigen die Teilhabe der Patienten am gesellschaftlichen und sozialen Leben. Aufgrund der kontinuierlichen Progression der IPF besteht ein wichtiges Therapieziel in der Erhaltung der Lebensqualität im Krankheitsverlauf.

Die Verschlechterung der Lebensqualität bis Woche 52, ausgedrückt durch eine *Zunahme des SGRQ-I-Gesamtscore*, war in den gepoolten INPULSIS Studien unter Nintedanib signifikant geringer als bei den Placebo-behandelten Studienteilnehmern. Darüber hinaus fand sich eine höhere Anzahl der *PGI-C-Responder* unter Behandlung mit Nintedanib verglichen mit Placebo (Tabelle 4-78).

Die gesundheitsbezogene Lebensqualität, gemessen anhand weiterer Instrumente wie *SGRQ*, *SOBQ*, *CASA-Q* oder *EQ-5D-VAS*, zeigt einen konsistenten Trend hinsichtlich einer besseren Lebensqualität zugunsten einer Behandlung mit Nintedanib im Vergleich zu Placebo. Insgesamt wird für den patientenrelevanten Endpunkt *Lebensqualität* ein Anhaltspunkt auf einen *nicht quantifizierbaren* Zusatznutzen abgeleitet.

In Bezug auf die gesundheitsbezogene Lebensqualität der IPF-Patienten zeigen die Ergebnisse insgesamt einen Anhaltspunkt auf einen **nicht quantifizierbaren** Zusatznutzen von Nintedanib.

Tabelle 4-78: Bewertung des Zusatznutzens mit Erläuterung - Lebensqualität

Ergebnisse aus gepoolten INPULSIS Studien	<u>Veränderung des SGRO-I-Gesamtscore</u> MWD [95 % KI]: -1,99 [-3,71;-0,27]; p = 0,0230 <u>PGI-C-Responder:</u> OR ^a [95 % KI]: 1,328 [1,04; 1,70]; p = 0,0254
Erläuterungen	Zum Nachweis der therapeutischen Verbesserung sind nach Verfo und AM-NutzenV vorrangig klinische Studien zu berücksichtigen, die unter anderem insbesondere die Lebensqualität als patientenrelevanten Endpunkt darstellen (Bundesministerium der Justiz und für Verbraucherschutz, 2010, Gemeinsamer Bundesausschuss, 2014).
Bewertung des Zusatznutzens	<ul style="list-style-type: none"> Anhaltspunkt für einen nicht quantifizierbaren Zusatznutzen von Nintedanib (Ofev[®]) im Vergleich zu Placebo bezüglich einer höheren Lebensqualität.
a: Ein OR > 1 zeigt einen Behandlungseffekt zugunsten der Nintedanib-Therapie an.	

Unerwünschte Ereignisse

UE traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 95,5 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 89,6 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen ist statistisch signifikant (OR [95 % KI]: 2,44 [1,50; 3,96]; p < 0,001). Daraus ergibt sich ein Hinweis auf einen Nachteil für Nintedanib gegenüber Placebo für diesen Endpunkt.

SUE traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 30,4 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, und somit mit vergleichbarer Häufigkeit wie bei den Placebo-behandelten Studienteilnehmern (30,0 %). Der Unterschied zwischen den Gruppen war nicht statistisch signifikant (OR [95 % KI]: 1,02 [0,78; 1,33]; p = 0,894). Für diesen Endpunkt liegt somit weder ein Zusatznutzen noch ein Nachteil von Nintedanib gegenüber Placebo vor.

UE, die zum Therapieabbruch führten, traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 19,3 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 13,0 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen ist statistisch signifikant (OR [95 % KI]: 1,60 [1,13; 2,26]; p = 0,007). Somit brachen in der Nintedanib-Gruppe häufiger Studienteilnehmer die Behandlung aufgrund eines *UE* ab als in der Placebo-Gruppe. Daraus ergibt sich ein Hinweis auf einen Nachteil für Nintedanib gegenüber Placebo für diesen Endpunkt.

UE, die zum Tod führten, traten in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien bei 5,8 % der Studienteilnehmer, die mit Nintedanib behandelt wurden, auf, verglichen mit 7,3 % der Placebo-behandelten Studienteilnehmer. Der Unterschied zwischen den Gruppen war nicht statistisch signifikant (OR [95 % KI]: 0,78 [0,48; 1,28]; $p = 0,319$). Somit verstarb in der Nintedanib-Gruppe numerisch seltener ein Patient infolge eines *UE* als in der Placebo-Gruppe. Für diesen Endpunkt liegt somit weder ein Zusatznutzen noch ein Nachteil von Nintedanib vor.

Als *UE von besonderem Interesse* wurden *gastrointestinale Störungen (Diarrhoe, Übelkeit, Erbrechen, Appetitlosigkeit, Gewichtsverlust, gastrointestinale Perforation), Leber- und Gallenerkrankungen (Leberenzymhöhung, Bilirubinerhöhung), kardiale Ereignisse, respiratorische Ereignisse (Progression der IPF, Husten, Dyspnoe, Bronchitis, Nasopharyngitis), Blutungsereignisse* sowie alle weiteren Störungen, die in mindestens einem Studienarm mit einer Häufigkeit von > 10 % der Studienteilnehmer auftraten, betrachtet.

Gastrointestinale Ereignisse

Die gastrointestinalen *UE Diarrhoe, Übelkeit und Erbrechen* sowie die weiteren *UE Appetitlosigkeit und Gewichtsverlust* traten unter der Behandlung mit Nintedanib signifikant häufiger auf als in der Placebo-Gruppe. *Gastrointestinale Perforationen* traten insgesamt so selten auf, dass keine Aussage zu einem möglichen Risiko vorgenommen werden kann. (Tabelle 4-44).

Leberenzym- und Bilirubinerhöhungen

UE, welche das Leber- und Gallenwegssystem betrafen, wurden in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien häufiger unter Nintedanib als unter Placebo dokumentiert: Die Behandlung mit Nintedanib war im Vergleich zu Placebo häufiger verbunden mit einer *Erhöhung der Leberenzyme (ALT, AST, ALKP) oder des Bilirubinwertes*.

Es ist jedoch zu beachten, dass sich die symptomatischen *unerwünschten Ereignisse* insgesamt gut behandeln lassen. Eine *Diarrhoe* sollte möglichst frühzeitig erkannt, durch pharmakologische Intervention (z. B. Loperamid) behandelt und nachbeobachtet werden. Außerdem lassen sich die unerwünschten Ereignisse durch eine Unterbrechung der Behandlung und Dosisreduktion häufig reduzieren bzw. beseitigen.

Kardiale Ereignisse

Unerwünschte kardiale Ereignisse wurden in den gepoolten INPULSIS Studien vergleichbar häufig unter Nintedanib und Placebo (10,0 % vs. 10,6 %) beobachtet. Hierunter waren beispielsweise folgende Ereignisse subsummiert: *Vorhofflimmern, Myokardinfarkt, koronare Herzerkrankung, Herzversagen, EKG-Auffälligkeiten wie Arrhythmien oder Schenkelblockbilder oder Cor pulmonale*. Auch *schwerwiegende unerwünschte kardiale Ereignisse* traten in der Nintedanib-Gruppe (5,0 %) und Placebo-Gruppe (5,4 %) vergleichbar häufig auf.

Tödliche kardiale Ereignisse traten numerisch seltener in der Nintedanib-Gruppe (0,5 %) als in der Placebo-Gruppe (1,4 %) auf (OR [95 % KI]: 0,33 [0,08; 1,32]; p = 0,099). *Arterielle thromboembolische Ereignisse* wurden in den gepoolten INPULSIS Studien häufiger unter Nintedanib als unter Placebo (2,5 % vs. 0,7 %) beobachtet (Tabelle 4-46). *Myokardinfarkte* traten numerisch häufiger unter Nintedanib als unter Placebo (1,6 % vs. 0,5 %) auf.

Respiratorische Ereignisse

Eine *Progression der IPF* wurde in der gepoolten Analyse der INPULSIS-1 und INPULSIS-2 Studien in der Nintedanib-Gruppe von 10,0 % der Studienteilnehmer als *UE* angegeben, verglichen mit 14,4 % in der Placebo-Gruppe. Der Unterschied ist statistisch signifikant zugunsten von Nintedanib (OR [95 % KI]: 0,66 [0,45; 0,96]; p = 0,030). *Husten* als *UE* trat in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig auf (13,3 % vs. 13,5 %) (OR [95 % KI]: 0,99 [0,72; 1,35]; p = 0,943). *Dyspnoe* als *UE* wurde in den gepoolten INPULSIS Studien signifikant seltener bei Nintedanib-behandelten Studienteilnehmern (7,7 %) als bei Placebo-behandelten Studienteilnehmern (11,3 %) erfasst (OR [95 % KI]: 0,65; [0,43; 0,99]; p = 0,042). *Bronchitis* und *Nasopharyngitis* als *UE* fanden sich in den gepoolten INPULSIS Studien in beiden Behandlungsarmen vergleichbar häufig: *Bronchitis* in der Nintedanib-Gruppe bei 10,5 % der Studienteilnehmer vs. 10,6 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,99 [0,66; 1,47]; p = 0,943). *Nasopharyngitis* in der Nintedanib-Gruppe bei 13,6 % der Studienteilnehmer vs. 16,1 % der Studienteilnehmer in der Placebo-Gruppe (OR [95 % KI]: 0,82 [0,58; 1,16]; p = 0,271).

Blutungsereignisse

Blutungsereignisse fanden sich bei 10,3 % der Nintedanib- und 7,8 % der Placebo-behandelten Patienten. Der Gruppenunterschied war nicht statistisch signifikant.

Ein Hinweis auf einen Zusatznutzen gegenüber Placebo ergibt sich für die *unerwünschten Ereignisse von besonderem Interesse Progression der IPF* und *Dyspnoe*.

Bei den *unerwünschten Ereignissen von besonderem Interesse kardiale Ereignisse, schwerwiegende kardiale Ereignisse, tödliche kardiale Ereignisse, Husten, Bronchitis, Nasopharyngitis* und *Blutungsereignisse* liegen weder ein Zusatznutzen noch ein Nachteil von Nintedanib gegenüber Placebo vor.

Ein Nachteil von Nintedanib gegenüber Placebo ergibt sich für die *gastrointestinalen Ereignisse, arterielle thromboembolische Ereignisse* sowie *Leberenzymhöhung* und *Bilirubinerrhöhung*.

Fazit

Bei Ofev[®] handelt es sich um ein Arzneimittel zur Behandlung eines seltenen Leidens, der medizinische Zusatznutzen gilt gemäß § 35a Abs. 1 Satz 10 SGB V bereits durch die Zulassung als belegt. Lediglich das Ausmaß des Zusatznutzens muss für die Anzahl der Patienten, für die dieser Zusatznutzen besteht, auf Grundlage der Zulassungsstudien nachgewiesen werden.

Der therapeutische Bedarf in der Behandlung der IPF ist hoch (Behr et al., 2013). Ofev® ist das erste Arzneimittel für erwachsene Patienten in der Indikation IPF, das über alle Schweregrade hinweg eingesetzt werden kann. Ofev® wurde am 15.01.2015 im Rahmen eines „accelerated assessments“ als erstes Arzneimittel zur Behandlung aller Krankheitsstadien der IPF durch die Europäische Kommission zugelassen. Die US-amerikanische Gesundheitsbehörde FDA hatte Nintedanib bereits den Status eines Therapiedurchbruchs (Breakthrough Therapy Designation) erteilt und ebenfalls in einem beschleunigten Verfahren am 15. Oktober 2014 in den USA zugelassen (U. S. Food and Drug Administration, 2014). Gegenüber Placebo weist Nintedanib einen beträchtlichen Zusatznutzen in den Endpunktkategorien *Mortalität (FVC)* und *Morbidität (Reduktion adjudizierter akuter Exazerbationen)* auf. In der Endpunktkategorie *Lebensqualität* besteht ein nicht quantifizierbarer Zusatznutzen. Bezüglich der *unerwünschten Ereignisse* ist ein Nachteil anzunehmen.

Aus den dargestellten Daten geht hervor, dass bei der Behandlung mit Nintedanib für erwachsene Patienten mit IPF der Nutzen den potenziellen Nachteil überwiegt.

4.4.3 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht

Geben Sie auf Basis der in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse und unter Berücksichtigung des in Abschnitt 4.4.2 dargelegten Zusatznutzens sowie dessen Wahrscheinlichkeit und Ausmaß in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.

Tabelle 4-79: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Patienten mit IPF jeden Schweregrads	beträchtlich

4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte

4.5.1 Begründung für die Vorlage indirekter Vergleiche

Sofern mit dem Dossier indirekte Vergleiche (Abschnitt 4.3.2.1) eingereicht wurden, begründen Sie dies. Begründen Sie dabei auch, warum sich die ausgewählten Studien jeweils für einen indirekten Vergleich gegenüber dem zu bewertenden Arzneimittel und damit für den Nachweis eines Zusatznutzens durch indirekten Vergleich eignen.

Nicht zutreffend

4.5.2 Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nichtrandomisierte vergleichende Studien (Abschnitt 4.3.2.2) oder weitere Untersuchungen (Abschnitt 4.3.2.3) eingereicht wurden, nennen Sie die Gründe, nach denen es unmöglich oder unangemessen ist, zu den in diesen Studien bzw. Untersuchungen behandelten Fragestellungen Studien höchster Evidenzstufe (randomisierte klinische Studien) durchzuführen oder zu fordern.

Nicht zutreffend

4.5.3 Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen

Falls aus Ihrer Sicht valide Daten zu patientenrelevanten Endpunkten zum Zeitpunkt der Bewertung noch nicht vorliegen können, begründen Sie dies.

Nicht zutreffend

4.5.4 Verwendung von Surrogatendpunkten

Die Verwendung von Surrogatendpunkten bedarf einer Begründung (siehe Abschnitt 4.5.3). Zusätzlich soll dargelegt werden, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

Eine Validierung von Surrogatendpunkten bedarf in der Regel einer Meta-Analyse von Studien, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden patientenrelevanten Endpunkt untersucht wurden (Burzykowski 2005¹³, Molenberghs 2010¹⁴). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrunde liegende Anwendungsgebiet und das zu bewertende Arzneimittel sowie die Vergleichstherapie erlauben.

Eine Möglichkeit der Verwendung von Surrogatendpunkten ohne abschließende Validierung stellt die Anwendung des Konzepts eines sogenannten Surrogate-Threshold-Effekts (STE) (Burzykowski 2006¹⁵) dar. Daneben besteht die Möglichkeit einer Surrogatvalidierung in der quantitativen Betrachtung geeigneter Korrelationsmaße von Surrogatendpunkt und interessierendem patientenrelevanten Endpunkt („individuelle Ebene“) sowie von Effekten auf den Surrogatendpunkt und Effekten auf den interessierenden patientenrelevanten Endpunkt („Studienebene“). Dabei ist dann zu zeigen, dass die unteren Grenzen der entsprechenden 95 %-Konfidenzintervalle für solche Korrelationsmaße ausreichend hoch sind. Die Anwendung alternativer Methoden zur Surrogatvalidierung (siehe Weir 2006¹⁶) soll ausreichend begründet werden, insbesondere dann, wenn als Datengrundlage nur eine einzige Studie verwendet werden soll.

Berichten Sie zu den Studien zur Validierung oder zur Begründung für die Verwendung von Surrogatendpunkten mindestens folgende Informationen:

- Patientenpopulation
- Intervention
- Kontrolle
- Datenherkunft
- verwendete Methodik
- entsprechende Ergebnisse (abhängig von der Methode)
- Untersuchungen zur Robustheit
- ggf. Untersuchungen zur Übertragbarkeit

¹³ Burzykowski T (Ed.): *The evaluation of surrogate endpoints*. New York: Springer; 2005.

¹⁴ Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M: *A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials*. *Stat Methods Med Res* 2010; 19(3): 205-236.

¹⁵ Burzykowski T, Buyse M: *Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation*. *Pharm Stat* 2006; 5(3): 173-186.

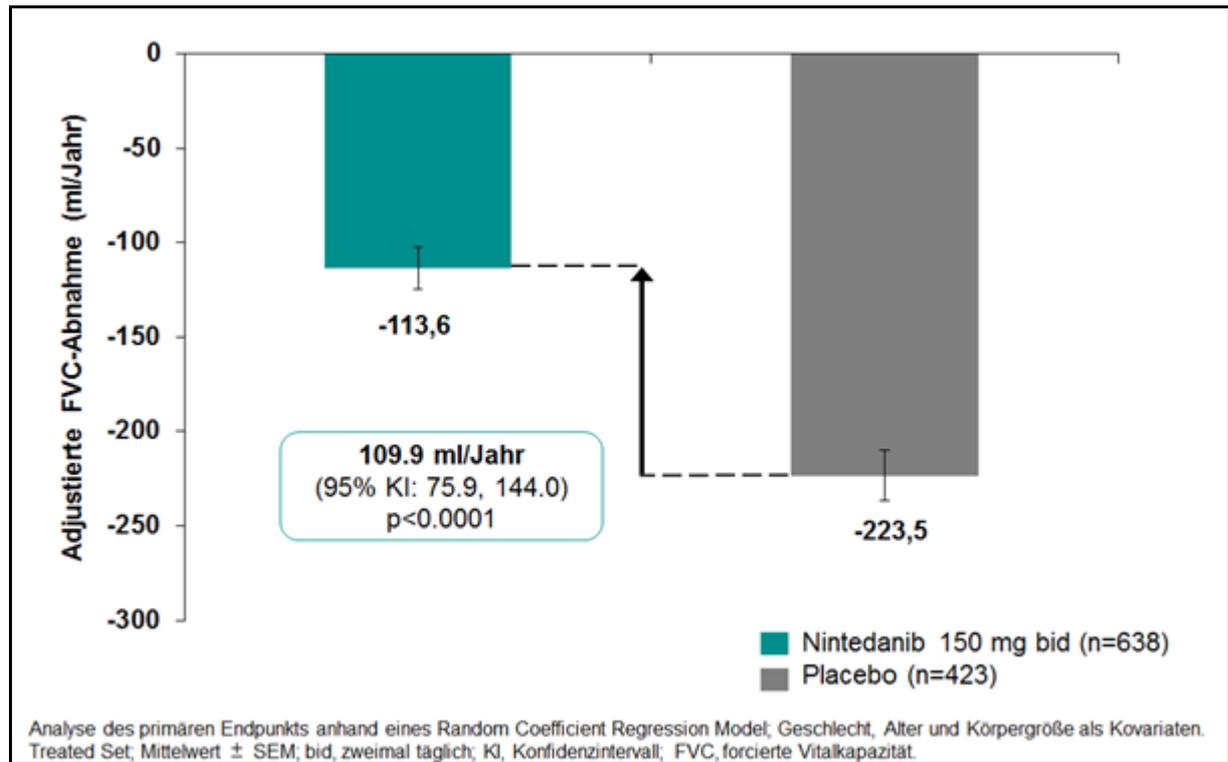
¹⁶ Weir CJ, Walley RJ: *Statistical evaluation of biomarkers as surrogate endpoints: a literature review*. *Stat Med* 2006; 25(2): 183-203.

Sofern Sie im Dossier Ergebnisse zu Surrogatendpunkten eingereicht haben, benennen Sie die Gründe für die Verwendung von Surrogatendpunkten. Beschreiben Sie, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

Zur vollständigen Beurteilung eines therapeutischen Effekts auf die mit einer IPF assoziierte Mortalität sind langjährige prospektive Studien notwendig, die eine Korrelation zwischen der eingesetzten Therapie und der Patientensterblichkeit herstellen. Um die Verfügbarkeit von Nintedanib für IPF-Patienten nicht zu verzögern, wurde keine Mortalitätsstudie durchgeführt, sondern der Surrogatparameter FVC genutzt, um indirekt Aussagen zu den zu erwartenden Mortalitätsunterschieden in einer größeren und länger dauernden Studie zu ermöglichen.

Bei der IPF handelt es um eine schwerwiegende und in der Regel zum Tod führende Erkrankung (Behr et al., 2013). Dabei ist die Mortalitätsrate höher als bei vielen Krebserkrankungen (Douglas et al., 2000). Wie in Abschnitt 4.2.5.2 ausgeführt, kommt es bei der IPF aufgrund der zunehmenden Vernarbung der Lunge zu einer mit der Zeit fortschreitenden Abnahme der FVC, sodass die Belüftung der Lunge zunehmend eingeschränkt wird, bzw. eine erhebliche Atemanstrengung des Patienten erfordert. Im Verlauf einer IPF verschlechtert sich die Lungenfunktion fortschreitend und unumkehrbar. Dies kann durch die Messung der FVC in regelmäßigen Intervallen im Krankheitsverlauf dokumentiert werden. FVC ist ein Lungenfunktionstest, der die Menge an Luft misst, die maximal aus der Lunge ausgeatmet werden kann, nachdem so tief wie möglich eingeatmet wurde. Für einen Patienten, der an IPF leidet, liegt die jährliche FVC-Abnahme im Bereich von 150 bis 200 ml (Ley et al., 2011). Dahingegen liegt der altersbedingte Verlust an FVC bei Gesunden bei etwa 30-65 ml pro Jahr (Griffith et al., 2001, Alexeeff et al., 2007, Braendli et al., 1996, Janssens, 2005).

Die *jährliche FVC-Abnahme* (ml) war in den gepoolten INPULSIS Studien unter Nintedanib mit 113,59 ml klinisch relevant und statistisch signifikant geringer als unter Placebo (223,53 ml).



Quelle: (Richeldi et al., 2014b)

Abbildung 4-11: Vergleich der FVC-Abnahme in ml – Nintedanib vs. Placebo

Sicht der Zulassungsbehörden

Die FVC wird weltweit von den Zulassungsbehörden als primärer Endpunkt für Studien im Anwendungsgebiet IPF anerkannt. Die allgemeine Akzeptanz der FVC als Prädiktor für die Mortalität spiegelt sich auch in den aktuellen Richtlinien der EMA wider. So wurde im Jahr 2011 im Rahmen eines Scientific Advice an die EMA der primäre Parameter FVC als adäquater Wirksamkeitsendpunkt bestätigt. Zudem hat die EMA den Parameter FVC als prognostisch für das Überleben der Patienten mit IPF akzeptiert (European Medicines Agency (EMA), 2011).

Die FVC ist ein valider Surrogatparameter, der in allen Phase II- und III-Studien sowohl bei Nintedanib als auch bei Pirfenidon sowie den meisten anderen Interventionsstudien im Anwendungsgebiet der IPF zum Einsatz kam (Antoniou et al., 2006, Raghu et al., 2004, Daniels et al., 2010). Gegenwärtig ist der Lungenfunktionsparameter FVC der am besten validierte und am breitesten eingesetzte Endpunkt in klinischen Studien im Indikationsgebiet. Du Bois et al. 2012 bekräftigen die Bedeutung der forcierten expiratorischen Vitalkapazität als robustes und valides Maß für die Krankheitsprogression bei IPF (du Bois et al., 2012). Der FVC-Wert erfülle alle Anforderungen an einen klinisch aussagekräftigen Endpunkt, mit dessen Hilfe qualitativ hochwertige RCTs durchgeführt werden können.

Sicht der Fachgesellschaften

Eine Abnahme der FVC von 5 % bis 10 % vom Sollwert innerhalb von sechs Monaten gilt bereits als prognostisch relevante Progression (Raghu et al., 2011, Jegal et al., 2005, Zappala et al., 2010). Diese Auffassung wird auch von der deutschen IPF-Konsensus-Konferenz vertreten. Nach Ansicht der Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin (DGP) ist der prognostische Stellenwert seit vielen Jahren durch zahlreiche klinisch kontrollierte Studien belegt. Daher kommt die DGP in ihrer Stellungnahme zur frühen Nutzenbewertung von Pirfenidon zum Schluss, dass die FVC den robustesten und sensitivsten Surrogatparameter des Überlebens von IPF-Patienten darstellt (Gemeinsamer Bundesausschuss, 2012).

Der Parameter FVC erfüllt darüber hinaus die im DIMDI-Bericht 91 (Deutsches Institut für Medizinische Dokumentation und Information) geforderten Validierungsanforderungen für einen patientenrelevanten Endpunkt (biologische Plausibilität, qualitativ und quantitativ gleichgerichtete Veränderung des klinischen Endpunktes, Wirkungsnachweis im Rahmen von RCT) (Mangiapane and Velasco Garrido, 2009). Insofern ist er ein valides Surrogat für die Mortalität dieser Patienten im Sinne der Anforderungen des DIMDI (Mangiapane and Velasco Garrido, 2009).

Der Zusammenhang zwischen FVC und Mortalität wurde im Rahmen verschiedener Studien untersucht, die im Folgenden beschrieben werden. Diese zeigten eine signifikante Korrelation von FVC mit der Patientensterblichkeit auf und bekräftigen die Verwendung des FVC als Surrogatparameter für die Mortalität bei IPF.

Studien zum Zusammenhang von FVC und Mortalität und Beurteilung der Schwellenwerte 10 % und 5 % vom Sollwert

Angesichts des heterogenen und nicht absehbaren Verlaufs einer IPF kommt prognostischen Faktoren eine besondere Bedeutung zu. Zahlreiche Studien zeigen, 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 ist. Tritt eine derartige Verschlechterung der FVC ein, steigt die Mortalität um bis zu einem Faktor von acht an, wobei die Zahl zwischen den einzelnen Studien etwas variiert (Collard et al., 2007a, Flaherty et al., 2003, Jegal et al., 2005, Latsi et al., 2003, Zappala et al., 2010, du Bois et al., 2011, Richeldi et al., 2012). Dabei erweist sich der Zusammenhang als robust über unterschiedliche Untersuchungsdauern und Patientenpopulationen.

Flaherty et al. konnten in einer retrospektiven Untersuchung an 80 Patienten mit UIP aufzeigen, dass eine Abnahme der FVC > 10 % über sechs Monate mit einem zweifachen Anstieg der Mortalität in dem anschließenden 6-Monats-Zeitraum in Verbindung gebracht werden kann (HR [95 %-KI]: 2,06 [1,09; 3,89]; p = 0,03), adjustiert nach Ausgangswert der FVC, Raucherstatus, Geschlecht und Zeit seit Auftreten der Symptome. Nach zwölf Monaten lag der Wert für das HR (Tod von Patienten mit > 10 % Rückgang in Prozent des FVC vom Sollwert) bei 1,70 (HR [95 %-KI]: 1,70 [0,76; 3,81]; p = 0,20) (Flaherty et al., 2003).

Collard et al. untersuchten das Überleben von 81 Patienten mit IPF (mittleres Alter: 62 Jahre, mittlere FVC : 67 % vom Sollwert) im Verhältnis zur absoluten Veränderung der FVC % vom Sollwert nach zwölf Monaten. Die Überlebenschancen nach vier Jahren betrug etwa 40 % bei einer Abnahme der FVC um mindestens 10 % vom Sollwert, 72 % bei stabiler FVC (± 10 % vom Sollwert) und 84 % bei einer Zunahme der FVC um mindestens 10 % vom Sollwert. Das Mortalitätsrisiko bei Patienten mit einer Abnahme der FVC um mindestens 10 % vom Sollwert stieg über einen Zeitraum von vier Jahren um das 2,8-fache an im Vergleich zu Patienten mit über 12 Monate stabiler FVC (± 10 % vom Sollwert) (Collard et al., 2003).

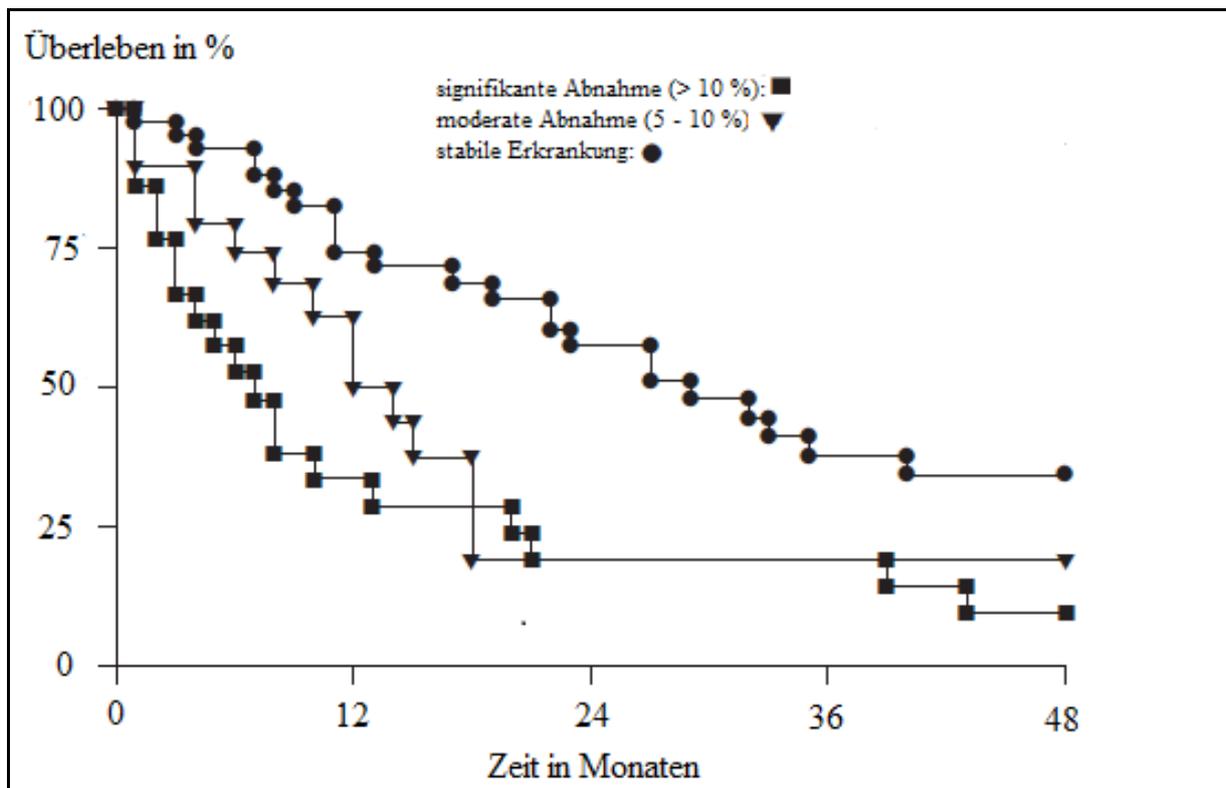
Richeldi et al. untersuchten den Zusammenhang zwischen der Veränderung der FVC % vom Sollwert und dem Transplantations-freien Überleben an 142 IPF-Patienten mit einem mittleren Alter von 67 Jahren und mittlerer Baseline-FVC von 2,7 l. In dieser Publikation wurden absolute und relative Änderungen unterschieden und die Mortalität über einen Zeitraum von zwei Jahren betrachtet. Eine relative bzw. absolute mindestens 10 %-ige Abnahme der FVC nach zwölf Monaten war dabei verbunden mit einem Anstieg des Risikos nach zwei Jahren zu versterben oder eine Transplantation erhalten zu müssen (OR [95%KI] 3,39 [1,14;10,07] bzw. 4,52 [1,27;16,12]) (Richeldi et al., 2012).

In einer von Du Bois et al. 2011 publizierten Studie mit 1.099 IPF-Patienten erwiesen sich neben dem Alter der Patienten auch eine vorangegangene Hospitalisierung aufgrund respiratorischer Beschwerden, der Ausgangswert der FVC und der DLco in Prozent des Sollwertes und der Abfall der FVC bzw. der DLco sowie der Parameter der Lebensqualität binnen 24 Wochen als unabhängige, prognostisch relevante Faktoren. Bereits ein absoluter Verlust der FVC von 5-10 % vom Sollwert innerhalb 24 Wochen (z. B. von 60% auf 55-50% vom Sollwert) war mit einem mehr als zweifach erhöhten Mortalitätsrisiko (HR [95 % KI]: 2,60 [1,75; 3,95]; $p < 0,001$), ein Abfall um mehr als 10 % vom Sollwert mit einem achtfach erhöhten Mortalitätsrisiko innerhalb des nächsten Jahres verbunden (HR [95 % KI]: 7,99 [5,26; 12,14]; $p < 0,001$) (du Bois et al., 2011).

In einer ebenfalls von Du Bois 2011 publizierten Studie mit 1.156 IPF-Patienten wurde die MCID der FVC mit verschiedenen Methoden ermittelt. Entsprechend dieser Studie liegt dieser Wert bei 2–6 % (je nach angewandter Methode; eher niedriger auf Ebene einer Population, eher höher im Individualfall), d. h. ein Abfall der FVC (in % vom Sollwert) um ca. fünf Prozentpunkte binnen 24 Wochen stellt bereits ein klinisch relevantes Ereignis mit potenziell therapeutischen Implikationen dar (du Bois et al., 2011).

In einer retrospektiven Untersuchung an 179 Patienten mit fibrotischer interstitieller Pneumonie (UIP/IPF 131; unspezifische interstitielle Pneumonie 48), war eine Veränderung der prozentualen erwarteten FVC (Anstieg von > 10 %, keine Veränderung, Abnahme von > 10 %) ein unabhängiger Prädiktor für das Überleben nach sechs Monaten ($p < 0,001$) und hatte einen stärkeren prognostischen Wert als jede andere Messung der Lungenfunktion (Jegal et al., 2005).

Zappala et al. untersuchten das Verhältnis zwischen der Veränderung der FVC % vom Sollwert und der 4-Jahres-Mortalität von 84 Patienten mit IPF. Die Auswertung erfolgte in den Kategorien „signifikante Abnahme“ (Abnahme der FVC > 10 % vom Sollwert), „moderate Abnahme“ (Abnahme der FVC um 5 % bis 10 % vom Sollwert) und „stabile Erkrankung“. Die Mortalität war bei Patienten mit signifikanter FVC-Abnahme nach sechs Monaten 2,8-fach höher als bei stabiler Erkrankung ($p < 0,001$). Patienten mit moderater FVC-Abnahme hatten im Vergleich zu Patienten mit stabiler Erkrankung ebenfalls ein 2,31-fach erhöhtes Risiko zu versterben ($p = 0,01$). Patienten mit signifikanter Abnahme und mit moderater Abnahme hatten somit eine miteinander vergleichbar stark erhöhte Mortalität (Zappala et al., 2010).



Quelle: (Zappala et al., 2010)

Abbildung 4-12: Vier-Jahres-Überleben in Abhängigkeit vom Ausmaß der FVC-Abnahme in % vom Sollwert über sechs Monate bei Patienten mit IPF (n = 84)

Fazit

Nach Beurteilung der in den vorausgehenden Abschnitten beschriebenen Studienlage, unter Berücksichtigung der relevanten Leitlinien und im Einklang mit den Vorgaben der Zulassungsbehörden wurde in den Studien zur Wirksamkeit und Verträglichkeit von Nintedanib die jährliche Abnahme der forcierten Vitalkapazität FVC als primärer Endpunkt für die Bestimmung des therapeutischen Effekts gewählt. Wie erläutert steht dieser Parameter in direkter Korrelation mit dem Krankheitsverlauf und stellt den **stärksten klinischen Prädiktor für die Mortalität bei IPF-Patienten dar und wird daher von BI als Surrogatparameter für die Mortalität angesehen und berücksichtigt.**

4.6 Liste der eingeschlossenen Studien

Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).

Tabelle 4-80: Liste der eingeschlossenen Studien

Studie	Quellen
INPULSIS-1	Studienprotokoll (Boehringer Ingelheim Pharma GmbH & Co. KG, 2012a) Studienbericht (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c) Registerbericht (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015c) Publikation: (Richeldi et al., 2014a, Richeldi et al., 2014b)
INPULSIS-2	Studienprotokoll (Boehringer Ingelheim Pharma GmbH & Co. KG, 2012b) Studienbericht (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d) Registerbericht (Boehringer Ingelheim Pharma GmbH & Co. KG, 2015d) Publikationen: (Richeldi et al., 2014a, Richeldi et al., 2014b)

4.7 Referenzliste

Listen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge), die Sie im vorliegenden Dokument angegeben haben (als fortlaufend nummerierte Liste). Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard). Geben Sie bei Fachinformationen immer den Stand des Dokuments an.

- [1] Aaronson, N. K. 1993. Assessment of quality of life and benefits from adjuvant therapies in breast cancer. *Recent Results Cancer Res*, 127, 201-10.
- [2] Alexeeff, S. E., Litonjua, A. A., Sparrow, D., Vokonas, P. S. & Schwartz, J. 2007. Statin use reduces decline in lung function: VA Normative Aging Study. *Am J Respir Crit Care Med*, 176, 742-7.
- [3] Antoniou, K. M., Nicholson, A. G., Dimadi, M., Malagari, K., Latsi, P., Rapti, A., Tzanakis, N., Trigidou, R., Polychronopoulos, V. & Bouros, D. 2006. Long-term clinical effects of interferon gamma-1b and colchicine in idiopathic pulmonary fibrosis. *Eur Respir J*, 28, 496-504.
- [4] Archibald, C. & Guidotti, T. 1987. Degree of objectively measured impairment and perceived shortness of breath with activities of daily living in patients with chronic obstructive pulmonary disease. *Can J Rehab*, 1, 45-54.
- [5] Azuma, A., Nukiwa, T., Tsuboi, E., Suga, M., Abe, S., Nakata, K., Taguchi, Y., Nagai, S., Itoh, H., Ohi, M., Sato, A. & Kudoh, S. 2005. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, 171, 1040-7.
- [6] Behr, J. 2013a. The diagnosis and treatment of idiopathic pulmonary fibrosis. *Dtsch Arztebl Int*, 110, 875-81.
- [7] 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. 2013b. S2K-Leitlinie zur Diagnostik und Therapie der idiopathischen Lungenfibrose. *Pneumologie*, 67, 81-111.
- [8] Boehringer Ingelheim Pharma GmbH & Co. KG 2012a. Studienprotokoll 1199.32 (Phase-III Studie).
- [9] Boehringer Ingelheim Pharma GmbH & Co. KG 2012b. Studienprotokoll 1199.34 (Phase-III Studie).
- [10] Boehringer Ingelheim Pharma GmbH & Co. KG 2014a. SCE-Supplement Phase-III Studien 1199.32 und 1199.34 (gepoolt).
- [11] Boehringer Ingelheim Pharma GmbH & Co. KG 2014b. SCS-Supplement Phase-III Studien 1199.32 und 1199.34
- [12] Boehringer Ingelheim Pharma GmbH & Co. KG 2014c. Studienbericht 1199.32 (Phase-III Studie).
- [13] Boehringer Ingelheim Pharma GmbH & Co. KG 2014d. Studienbericht 1199.34 (Phase-III Studie).
- [14] Boehringer Ingelheim Pharma GmbH & Co. KG 2014e. Übersicht Sensitivitätsanalysen Studien INPULSIS-1, INPULSIS-2.
- [15] Boehringer Ingelheim Pharma GmbH & Co. KG 2014f. Weitere Analysen INPULSIS-1, INPULSIS-2.
- [16] Boehringer Ingelheim Pharma GmbH & Co. KG 2015a. Baseline-Patientencharakteristika Subgruppen - Nachberechnung.
- [17] Boehringer Ingelheim Pharma GmbH & Co. KG 2015b. Fachinformation Ofev.
- [18] Boehringer Ingelheim Pharma GmbH & Co. KG. 2015c. *Safety and Efficacy of BIBF 1120 at High Dose in Idiopathic Pulmonary Fibrosis Patients - Registereintrag ClinicalTrials.gov - INPULSIS 1* [Online]. ClinicalTrials.gov. Available: <http://clinicaltrials.gov/ct2/show/NCT01335464?term=nintedanib+idiopathic+pulmonary+fibrosis&rank=6>.
- [19] Boehringer Ingelheim Pharma GmbH & Co. KG. 2015d. *Safety and Efficacy of BIBF 1120 at High Dose in Idiopathic Pulmonary Fibrosis Patients II - Registereintrag ClinicalTrials.gov - INPULSIS 2*, [Online]. ClinicalTrials.gov. Available: <http://clinicaltrials.gov/ct2/show/NCT01335477?term=nintedanib+idiopathic+pulmonary+fibrosis&rank=7>.
- [20] Braendli, O., Schindler, C., Kunzli, N., Keller, R. & Perruchoud, A. P. 1996. Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax*, 51, 277-83.
- [21] Brazier, J., Jones, N. & Kind, P. 1993. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Qual Life Res*, 2, 169-80.
- [22] Brown, A. W., Fischer, C. P., Shlobin, O. A., Buhr, R. G., Ahmad, S., Weir, N. A. & Nathan, S. D. 2015. Outcomes after hospitalization in idiopathic pulmonary fibrosis: a cohort study. *Chest*, 147, 173-9.
- [23] Bundesministerium der Justiz und für Verbraucherschutz 2010. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV). juris GmbH.

- [24] Bundesministerium der Justiz und für Verbraucherschutz 2014. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz - AMG). juris GmbH.
- [25] Collard, H. R., King, T. E., Jr., Bartelson, B. B., Vourlekis, J. S., Schwarz, M. I. & Brown, K. K. 2003. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, 168, 538-42.
- [26] Collard, H. R., Loyd, J. E., King, T. E., Jr. & Lancaster, L. H. 2007a. Current diagnosis and management of idiopathic pulmonary fibrosis: a survey of academic physicians. *Respir Med*, 101, 2011-6.
- [27] Collard, H. R., Moore, B. B., Flaherty, K. R., Brown, K. K., Kaner, R. J., King, T. E., Jr., Lasky, J. A., Loyd, J. E., Noth, I., Olman, M. A., Raghu, G., Roman, J., Ryu, J. H., Zisman, D. A., Hunninghake, G. W., Colby, T. V., Egan, J. J., Hansell, D. M., Johkoh, T., Kaminski, N., Kim, D. S., Kondoh, Y., Lynch, D. A., Muller-Quernheim, J., Myers, J. L., Nicholson, A. G., Selman, M., Toews, G. B., Wells, A. U., Martinez, F. J. & Idiopathic Pulmonary Fibrosis Clinical Research Network, I. 2007b. Acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, 176, 636-43.
- [28] Crawford, B., Monz, B., Hohlfeld, J., Roche, N., Rubin, B., Magnussen, H., Nivens, C., Ghafouri, M., McDonald, J. & Tetzlaff, K. 2008. Development and validation of a cough and sputum assessment questionnaire. *Respir Med*, 102, 1545-55.
- [29] Daniels, C. E., Lasky, J. A., Limper, A. H., Mieras, K., Gabor, E., Schroeder, D. R. & Imatinib, I. P. F. S. I. 2010. Imatinib treatment for idiopathic pulmonary fibrosis: Randomized placebo-controlled trial results. *Am J Respir Crit Care Med*, 181, 604-10.
- [30] de Boer, A., van Lanschott, J., Stalmeier, P., van Sandick, J., Hulscher, J., de Haes, J. & Sprangers, M. 2004. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual Life Res*, 2, 311-320.
- [31] Douglas, W. W., Ryu, J. H. & Schroeder, D. R. 2000. Idiopathic pulmonary fibrosis Impact of oxygen and colchicine, prednisone, or no therapy on survival. *161*, 1172-1178.
- [32] du Bois, R. M., Nathan, S. D., Richeldi, L., Schwarz, M. I. & Noble, P. W. 2012. Idiopathic pulmonary fibrosis: lung function is a clinically meaningful endpoint for phase III trials. *Am J Respir Crit Care Med*, 186, 712-5.
- [33] du Bois, R. M., Weycker, D., Albera, C., Bradford, W. Z., Costabel, U., Kartashov, A., King, T. E., Jr., Lancaster, L., Noble, P. W., Sahn, S. A., Thomeer, M., Valeyre, D. & Wells, A. U. 2011. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med*, 184, 1382-9.
- [34] Eakin, E. G., Resnikoff, P. M., Prewitt, L. M., Ries, A. L. & Kaplan, R. M. 1998. Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire. University of California, San Diego. *Chest*, 113, 619-24.
- [35] European Commission (EC). 2015a. *Community register of medicinal products for human use - Ofev* [Online]. Available: <http://ec.europa.eu/health/documents/community-register/html/h979.htm> [Accessed 27.01.2015].

- [36] European Commission (EC) 2015b. Note to the Members of the Standing Committee on Medicinal Products for Human Use/Standing Committee on Veterinary Medicinal Products - Adoption of Commission Implementing Decision granting marketing authorisation under Regulation (EC) No 726/2004 of the European Parliament and of the Council for "Ofev - Nintedanib", an orphan medicinal product for human use.
- [37] European Medicines Agency 2014. Rapporteur and Co-Rapporteur Day 120 Joint Response Assessment Report - Ofev.
- [38] European Medicines Agency (EMA) 1996. ICH Topic E 3 Structure and Content of Clinical Study Reports - Step 5 - Note for Guidance on Structure and Content of Clinical Study Reports (CPMP/ICH/137/95).
- [39] European Medicines Agency (EMA) 2011. Scientific Advice - BIBF 1120.
- [40] Flaherty, K. R., Mumford, J. A., Murray, S., Kazerooni, E. A., Gross, B. H., Colby, T. V., Travis, W. D., Flint, A., Toews, G. B., Lynch, J. P., 3rd & Martinez, F. J. 2003. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med*, 168, 543-8.
- [41] Gemeinsamer Bundesausschuss 2012. Zusammenfassende Dokumentation über die Änderung der Arzneimittel-Richtlinie (AM-AL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Pirfenidon.
- [42] Gemeinsamer Bundesausschuss 2013. Erstellung und Einreichung eines Dossiers zur Nutzenbewertung gemäß § 35a SGB V - Format und Gliederung des Dossiers, einzureichende Unterlagen, Vorgaben für technische Standards.
- [43] Gemeinsamer Bundesausschuss 2014. Verfahrensordnung des Gemeinsamen Bundesausschusses.
- [44] Gries, K. S., Esser, D. & Wiklund, I. 2013. Content validity of CASA-Q cough domains and UCSD-SOBQ for use in patients with Idiopathic Pulmonary Fibrosis. *Glob J Health Sci*, 5, 131-41.
- [45] Griffith, K. A., Sherrill, D. L., Siegel, E. M., Manolio, T. A., Bonekat, H. W. & Enright, P. L. 2001. Predictors of loss of lung function in the elderly: the Cardiovascular Health Study. *Am J Respir Crit Care Med*, 163, 61-8.
- [46] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2011. IQWiG-Berichte - Nr. 108 - Pirfenidon - Nutzenbewertung gemäß § 35a SGB V. Gemeinsamer Bundesausschuss.
- [47] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). 2013. *Allgemeine Methoden Version 4.1* [Online]. Available: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf [Accessed 22.01.2015].
- [48] Janssens, J. P. 2005. Aging of the respiratory system: impact on pulmonary function tests and adaptation to exertion. *Clin Chest Med*, 26, 469-84, vi-vii.
- [49] Jegal, Y., Kim, D. S., Shim, T. S., Lim, C. M., Do Lee, S., Koh, Y., Kim, W. S., Kim, W. D., Lee, J. S., Travis, W. D., Kitaichi, M. & Colby, T. V. 2005. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med*, 171, 639-44.

- [50] Johannson, K. & Collard, H. R. 2013. Acute Exacerbation of Idiopathic Pulmonary Fibrosis: A Proposal. *Curr Respir Care Rep*, 2.
- [51] Jones, P. W. 2014. Estimation and application of the minimum clinically important difference in COPD. *Lancet Respir Med*, 2, 167-9.
- [52] Jones, P. W., Donohue, J. F., Nedelman, J., Pascoe, S., Pinault, G. & Lassen, C. 2011. Correlating changes in lung function with patient outcomes in chronic obstructive pulmonary disease: a pooled analysis. *Respir Res*, 12, 161.
- [53] Jones, P. W., Quirk, F. H., Baveystock, C. M. & Littlejohns, P. 1992. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis*, 145, 1321-7.
- [54] Kim, D. S., Park, J. H., Park, B. K., Lee, J. S., Nicholson, A. G. & Colby, T. 2006. Acute exacerbation of idiopathic pulmonary fibrosis: frequency and clinical features. *Eur Respir J*, 27, 143-50.
- [55] King, T. E., Jr., Behr, J., Brown, K. K., du Bois, R. M., Lancaster, L., de Andrade, J. A., Stahler, G., Leconte, I., Roux, S. & Raghu, G. 2008. BUILD-1: a randomized placebo-controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, 177, 75-81.
- [56] Kupferberg, D. H., Kaplan, R. M., Slymen, D. J. & Ries, A. L. 2005. Minimal clinically important difference for the UCSD Shortness of Breath Questionnaire. *J Cardiopulm Rehabil*, 25, 370-7.
- [57] Latsi, P. I., du Bois, R. M., Nicholson, A. G., Colby, T. V., Bisirtzoglou, D., Nikolakopoulou, A., Veeraraghavan, S., Hansell, D. M. & Wells, A. U. 2003. Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med*, 168, 531-7.
- [58] Ley, B., Collard, H. R. & King, T. E., Jr. 2011. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, 183, 431-40.
- [59] Mangiapane, S. & Velasco Garrido, M. 2009. Use of Surrogate end points in HTA. *GMS Health Technol Assess*, 5, Doc12.
- [60] Miller, M. R., Crapo, R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Enright, P., van der Grinten, C. P., Gustafsson, P., Jensen, R., Johnson, D. C., MacIntyre, N., McKay, R., Navajas, D., Pedersen, O. F., Pellegrino, R., Viegi, G., Wanger, J. & Force, A. E. T. 2005. General considerations for lung function testing. *Eur Respir J*, 26, 153-61.
- [61] Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M. & Altman, D. G. 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c869.
- [62] Osoba, D., Rodrigues, G., Myles, J., Zee, B. & Pater, J. 1998. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol*, 16, 139-44.
- [63] Parambil, J. G., Myers, J. L. & Ryu, J. H. 2005. Histopathologic Features and Outcome of Patients With Acute Exacerbation of Idiopathic Pulmonary Fibrosis Undergoing Surgical Lung Biopsy. *Chest*, 125, 3310-3315.

- [64] Pickard, A. S., Wilke, C., Jung, E., Patel, S., Stavem, K. & Lee, T. A. 2008. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. *Respir Med*, 102, 519-36.
- [65] Rabin, R. & de Charro, F. 2001. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*, 33, 337-43.
- [66] Raghu, G., Brown, K. K., Bradford, W. Z., Starko, K., Noble, P. W., Schwartz, D. A., King, T. E., Jr. & Idiopathic Pulmonary Fibrosis Study, G. 2004. A placebo-controlled trial of interferon gamma-1b in patients with idiopathic pulmonary fibrosis. *N Engl J Med*, 350, 125-33.
- [67] Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., Hansell, D. M., Johkoh, T., Kim, D. S., King, T. E., Jr., Kondoh, Y., Myers, J., Muller, N. L., Nicholson, A. G., Richeldi, L., Selman, M., Dudden, R. F., Griss, B. S., Protzko, S. L., Schunemann, H. J. & Fibrosis, A. E. J. A. C. o. I. P. 2011. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*, 183, 788-824.
- [68] Richeldi, L., Cottin, V., Flaherty, K. R., Kolb, M., Inoue, Y., Raghu, G., Taniguchi, H., Hansell, D. M., Nicholson, A. G., Le Maulf, F., Stowasser, S. & Collard, H. R. 2014a. Design of the INPULSIS trials: two phase 3 trials of nintedanib in patients with idiopathic pulmonary fibrosis. *Respir Med*, 108, 1023-30.
- [69] Richeldi, L., du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., Cottin, V., Flaherty, K. R., Hansell, D. M., Inoue, Y., Kim, D. S., Kolb, M., Nicholson, A. G., Noble, P. W., Selman, M., Taniguchi, H., Brun, M., Le Maulf, F., Girard, M., Stowasser, S., Schlenker-Herceg, R., Disse, B., Collard, H. R. & Investigators, I. T. 2014b. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*, 370, 2071-82.
- [70] Richeldi, L., Ryerson, C. J., Lee, J. S., Wolters, P. J., Koth, L. L., Ley, B., Elicker, B. M., Jones, K. D., King, T. E., Jr., Ryu, J. H. & Collard, H. R. 2012. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax*, 67, 407-11.
- [71] Schulz, K. F., Altman, D. G., Moher, D. & Group, C. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340, c332.
- [72] Song, J. W., Hong, S. B., Lim, C. M., Koh, Y. & Kim, D. S. 2011. Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *Eur Respir J*, 37, 356-63.
- [73] Swigris, J. J., Brown, K. K., Behr, J., du Bois, R. M., King, T. E., Raghu, G. & Wamboldt, F. S. 2010. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. *Respir Med*, 104, 296-304.
- [74] Taniguchi, H., Ebina, M., Kondoh, Y., Ogura, T., Azuma, A., Suga, M., Taguchi, Y., Takahashi, H., Nakata, K., Sato, A., Takeuchi, M., Raghu, G., Kudoh, S., Nukiwa, T. & Pirfenidone Clinical Study Group in, J. 2010. Pirfenidone in idiopathic pulmonary fibrosis. *Eur Respir J*, 35, 821-9.

- [75] U. S. Food and Drug Administration. 2014. *FDA approves Esbriet to treat idiopathic pulmonray fibrosis* [Online]. Available: <http://www.fda.gov/newsevents/newsroom/pressannouncements/ucm418991.htm> [Accessed 04.02.2015].
- [76] Velikova, G., Keding, A., Harley, C., Cocks, K., Booth, L., Smith, A. B., Wright, P., Selby, P. J. & Brown, J. M. 2010. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. *Eur J Cancer*, 46, 2381-8.
- [77] Wilson, C. B., Jones, P. W., O'Leary, C. J., Cole, P. J. & Wilson, R. 1997. Validation of the St. George's Respiratory Questionnaire in bronchiectasis. *Am J Respir Crit Care Med*, 156, 536-41.
- [78] Yorke, J., Jones, P. W. & Swigris, J. J. 2010. Development and validity testing of an IPF-specific version of the St George's Respiratory Questionnaire. *Thorax*, 65, 921-6.
- [79] Zappala, C. J., Latsi, P. I., Nicholson, A. G., Colby, T. V., Cramer, D., Renzoni, E. A., Hansell, D. M., du Bois, R. M. & Wells, A. U. 2010. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J*, 35, 830-6.

Anhang 4-A: Suchstrategien – bibliografische Literaturrecherche

Geben Sie nachfolgend die Suchstrategien für die bibliografische(n) Literaturrecherche(n) an, und zwar getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.). Für jede durchsuchte Datenbank ist die verwendete Strategie separat darzustellen. Geben Sie dabei zunächst jeweils den Namen der durchsuchten Datenbank (z. B. EMBASE), die verwendete Suchoberfläche (z. B. DIMDI, Ovid etc.), das Datum der Suche, das Zeitsegment (z. B.: „1980 to 2010 week 50“) und die gegebenenfalls verwendeten Suchfilter (mit Angabe einer Quelle) an. Listen Sie danach die Suchstrategie einschließlich der resultierenden Trefferzahlen auf. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel (eine umfassende Suche soll Freitextbegriffe und Schlagwörter enthalten):

Datenbankname	EMBASE	
Suchoberfläche	Ovid	
Datum der Suche	08.12.2010	
Zeitsegment	1980 to 2010 week 50	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle ¹⁷] – Strategy minimizing difference between sensitivity and specificity	
#	Suchbegriffe	Ergebnis
1	Meglitinide/	848
2	Nateglinide/	1686
3	Repaglinide/	2118
4	(glinid* or meglitinid* or nateglinid* or repaglinid*).ab,ti.	1069
5	(starlix or novonorm or novo norm or prandin).ab,ti.	32
6	(105816-04-4 or 135062-02-1).rn.	2854
7	or/1-6	3467
8	Diabetes mellitus/	224164
9	Non Insulin dependent Diabetes mellitus/	91081
10	(diabet* or niddm or t2dm).ab,ti.	379777
11	or/8-10	454517
12	(random* or double-blind*).tw.	650136
13	placebo*.mp.	243550

¹⁷ Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006; 94(4): 451-455. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

14	or/12-13	773621
15	and/7,11,14	719

Anhang 4-A1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nicht zutreffend

Anhang 4-A2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend

Anhang 4-A3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend

Anhang 4-A4: Suche nach weiteren Untersuchungen

Nicht zutreffend

Anhang 4-B: Suchstrategien – Suche in Studienregistern

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern an. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Für jedes durchsuchte Studienregister ist eine separate Strategie darzustellen. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters (z. B. *clinicaltrials.gov*), die Internetadresse, unter der das Studienregister erreichbar ist (z. B. <http://www.clinicaltrials.gov>), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	08.12.2010
Suchstrategie	(Starlix OR Novonorm OR Prandin OR Nateglinid OR Repaglinid) [ALL-FIELDS] AND ("Phase II" OR "Phase III" OR "Phase IV") [PHASE]
Treffer	23

Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nicht zutreffend

Anhang 4-B2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend

Anhang 4-B3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend

Anhang 4-B4: Suche nach weiteren Untersuchungen

Nicht zutreffend

Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche)

Listen Sie nachfolgend die im Volltext gesichteten und ausgeschlossenen Dokumente aus der / den bibliografischen Literaturrecherche(n) auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Anhang 4-C1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nicht zutreffend

Anhang 4-C2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend

Anhang 4-C3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend

Anhang 4-C4: Suche nach weiteren Untersuchungen

Nicht zutreffend

Anhang 4-D: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern)

Listen Sie nachfolgend die durch die Studienregistersuche(n) identifizierten, aber ausgeschlossenen Studien auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Anhang 4-D1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nicht zutreffend

Anhang 4-D2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend

Anhang 4-D3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend

Anhang 4-D4: Suche nach weiteren Untersuchungen

Nicht zutreffend

Anhang 4-E: Methodik der eingeschlossenen Studien – RCT

Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.4 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle inklusive eines Flow-Charts für den Patientenfluss (Tabelle 4-81, Tabelle 4-82).

Sollten Sie im Dossier indirekte Vergleiche präsentieren, beschreiben Sie ebenfalls die Methodik jeder zusätzlich in den indirekten Vergleich eingeschlossenen Studie (Abschnitt 4.3.2.1). Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle inklusive eines Flow-Charts für den Patientenfluss.

Tabelle 4-81 (Anhang): Studiendesign und -methodik für Studie INPULSIS-1

Item ^a	Charakteristikum	Studieninformation
Studienziel		
2b	Genaue Ziele, Fragestellung und Hypothesen	<p>Zielsetzung</p> <p>Untersuchung der Wirksamkeit sowie des Nutzen-Risiko-Verhältnisses von oral verabreichtem Nintedanib 150 mg bid im Vergleich zu Placebo bei IPF. Als primärer Endpunkt wird die Verringerung der Lungenfunktionsabnahme, gemessen als jährliche Abnahme der FVC, ermittelt.</p> <p>Hypothesen</p> <p>Bewertung der Überlegenheit von Nintedanib 150 mg bid im Vergleich zu Placebo in Bezug auf die jährliche Abnahme der FVC bei Patienten mit IPF.</p> <p>FVC-Abnahme in 52 Wochen:</p> <ul style="list-style-type: none"> • Nullhypothese (H₀): Keine Unterschiede der jährlichen Abnahme der FVC von Baseline bis Woche 52 zwischen den Gruppen Nintedanib 150 mg bid und Placebo. • Alternativhypothese (H_a): Die jährliche Abnahme der FVC von Baseline bis Woche 52 ist bei Patienten unter Nintedanib 150 mg bid geringer als unter Placebo.
Methoden		
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	<p>Allokation: randomisiert</p> <p>Zuteilungsverhältnis: 3:2</p> <p>Randomisierungseinheit: Studienteilnehmer</p> <p>Verblindung: doppelblind</p> <p>Studienhorizont: 52 Wochen + 28 Tage Nachbeobachtung</p> <p>Design: parallel, zweiarmig, Placebo-kontrolliert</p> <p>Studienorganisation: multizentrisch</p> <p>Phase: III</p>

Item ^a	Charakteristikum	Studieninformation
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p>Amendment 1 (11/2011)</p> <ul style="list-style-type: none"> • Möglichkeit eines Re-Tests bei abnormalen Laborparametern zu Visit 1, bei Verdachtsfällen auf Messfehler oder fehlendem bzw. angezweifeltem Bezug auf den Patientenzustand. • Studienausschluss, wenn Patienten den Studienablauf nicht verstehen bzw. die Fragebögen nicht selbstständig ausfüllen können • Ergänzung der „schwerwiegenden unerwünschten Ereignisse“ entsprechend den neuen BI Standards, um deren ordnungsgemäße Dokumentation zu gewährleisten. • Einschlusskriterium 4 wurde abgeändert: „Thorax-HRCT innerhalb von 12 Monaten vor Visit 1“ anstelle von „Thorax HRCT innerhalb von 12 Monaten vor Visit 2“ <p>Amendment 2 (09/2012)</p> <ul style="list-style-type: none"> • Ergänzung explorativer Biomarker Analysen, um den Effekt von Nintedanib auf Biomarker in Verbindung mit IPF Pathologie und prognostischen Erkrankungsmarkern zu untersuchen, und Ergänzung pharmakogenomischer Analysen • Schlechte Compliance wurde als Protokollverletzung definiert
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • Schriftliche Einwilligungserklärung vor Studienbeginn • Alter ≥ 40 Jahre zu Visit 1 • Diagnose IPF nach ATS-, ERS-, JRS- und ALAT-Richtlinien (innerhalb von 5 Jahren vor Baseline) • Thorax HRCT innerhalb von 12 Monaten vor Visit 1) • HRCT-Muster und, wenn verfügbar, Lungenbiopsie-Muster übereinstimmend mit IPF Diagnose bewertet von zentralen Reviewern • DL_{CO}: 30 % bis 79 % (zu Visit 2; angepasst an Hämoglobin) • FVC % pred. ≥ 50 % (zu Visit 2) <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> • Laborparameter (zu Visit 1) <ul style="list-style-type: none"> • AST und ALT $> 1,5$-fach ULN • Bilirubin $> 1,5$-fach ULN • Lungenfunktion (zu Visit 2) <ul style="list-style-type: none"> • Atemwegsobstruktion (FEV₁/FVC $< 0,7$) • Lungentransplantation im Verlauf der Studie

Item ^a	Charakteristikum	Studieninformation
		<p>wahrscheinlich (nach Prüfarzt-Einschätzung)</p> <ul style="list-style-type: none"> • Andere Krankheiten <ul style="list-style-type: none"> • Herzerkrankung <ul style="list-style-type: none"> - Myokardinfarkt innerhalb 6 Monaten vor Visit 2 - Instabile Angina innerhalb 1 Monat vor Visit 2 • Blutungsrisiko <ul style="list-style-type: none"> - Bekannte Prädisposition für Blutungen - Fibrinolyse mit im Studienprotokoll festgelegten Interventionen - Hämorrhagisches ZNS-Ereignis innerhalb von 12 Monaten vor Visit 2 - weitere Kriterien innerhalb von 3 Monaten vor Visit 2 <ul style="list-style-type: none"> - Hämoptysis oder Hämaturie - Aktive gastrointestinale Blutung oder Geschwüre - Größere Verletzung oder Operation - Koagulations Parameter: INR > 2, Verlängerung der Prothrombinzeit und PTT (sek) um > 50 % ULN (Visit 1) • 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 Visit 2 • Bekannte Hypersensitivität gegenüber Studienmedikation oder deren Bestandteilen • Andere Krankheiten, die das Testprozedere oder die Studienteilnahme beeinflussen oder den Patienten bei Studienteilnahme in Gefahr bringen (Ermessen des Prüfarztes) • Lebenserwartung aufgrund anderer Krankheiten als IPF < 2,5 Jahre (Ermessen des Prüfarztes) <ul style="list-style-type: none"> • Allgemeine Ausschlusskriterien <ul style="list-style-type: none"> • Frühere Behandlung mit Nintedanib länger als 4 Wochen) • Andere Interventionen (Forschungsstudie) innerhalb von 8 Wochen vor Visit 1 • Einnahme von N-Acetylcystein (NAC), Prednison > 15 mg/Tag oder vergleichbare Arzneimittel innerhalb von 2 Wochen vor Visit 1 • Einnahme von Pirfenidon, Azathioprin, Cyclophosphamid, Cyclosporin A innerhalb von

Item ^a	Charakteristikum	Studieninformation
		<p>8 Wochen vor Visit 1</p> <ul style="list-style-type: none"> • 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 gebärfähigem Alter sind) • Aktiver Alkohol- oder Drogenmissbrauch • Patienten, die das Studienprozedere nicht verstehen oder einhalten (inkl. Ausfüllen von Fragebögen ohne Hilfe)
4b	Studienorganisation und Ort der Studiendurchführung	Multizentrische Studie, 98 Studienzentren in 13 Ländern: Australien, Belgien, China, Tschechische Republik, Frankreich, Deutschland, Indien, Irland, Israel, Italien, Japan, UK, USA
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<p>Alle Gruppen</p> <p>Visit 1: Screening</p> <p>Visit 2: Therapiestart und nachfolgende Einnahme des Arzneimittels über 52 Wochen</p> <p>Nintedanib (oral)</p> <ul style="list-style-type: none"> • 2-mal täglich 150 mg im Intervall von 12 Stunden über 52 Wochen • Bei unerwünschten Ereignissen Reduktion der Dosis auf 2-mal täglich 100 mg im Intervall von 12 Stunden über 52 Wochen bzw. in besonderen Fällen Unterbrechung der Therapie für maximal 4 Wochen bei therapieassoziierten unerwünschten Ereignissen und bis zu 8 Wochen bei Exerzabationen. <p>Placebo (oral)</p> <ul style="list-style-type: none"> • 2-mal täglich im Intervall von 12 Stunden über 52 Wochen • Bei unerwünschten Ereignissen Reduktion der Dosis auf 2-mal täglich Placebo zu 100 mg im Intervall von 12 Stunden über 52 Wochen bzw. in besonderen Fällen Unterbrechung der Therapie von maximal 4 Wochen bei therapieassoziierten unerwünschten Ereignissen und bis zu 8 Wochen bei Exerzabationen. <p>Alle Gruppen</p> <p>Follow-up: 28 Tage nach Therapieende</p>

Item ^a	Charakteristikum	Studieninformation
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Primäres Zielkriterium</p> <p>Untersuchung der Wirksamkeit sowie des Nutzen/Risiko-Verhältnisses von Nintedanib 150 mg bid im Vergleich zu Placebo bei IPF.</p> <p>Primäre Endpunkte</p> <ul style="list-style-type: none"> • Jährliche FVC- Abnahme über 52 Wochen (ml) <p>Sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Veränderung des <i>Saint George's Respiratory Questionnaire</i> Gesamtscores von Baseline bis Woche 52 (in Punkten) • Zeit bis zur ersten akuten Exazerbation über 52 Wochen (in Tagen) <p>Weitere sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Weitere Analysen der FVC <ul style="list-style-type: none"> • Absolute und relative Veränderung der FVC (ml) und FVC % vom Sollwert über 52 Wochen • Absolute kategorielle Veränderung der FVC % vom Sollwert bis zu 52 Wochen: Abfall > 5 %, Anstieg > 5 % und Veränderungen ≤ 5 % • Absolute kategorielle Veränderung der FVC % vom Sollwert von Baseline bis Woche 52: Abfall > 10 %, Anstieg > 10 % und Veränderung ≤ 10 % • Anteil FVC-Responder mit Abfall < 5 % vom FVC-Sollwert nach 52 Wochen • Anteil FVC-Responder mit Abfall < 10 % vom FVC-Sollwert nach 52 Wochen • Patient Reported Outcomes (PRO) gemessen von Baseline bis Woche 52: <ul style="list-style-type: none"> • 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[CD])-Score • Anteil PGI-C Responder • Veränderung EQ-5D-VAS • Akute Exazerbation: <ul style="list-style-type: none"> • Risiko einer akuten IPF Exazerbation • Überlebenszeitanalysen gemessen über 52 Wochen: <ul style="list-style-type: none"> • Zeit bis zum Tod

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • 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 DL_{CO} < 30 % vom Sollwert oder SpO₂ < 88 % in Ruhe) • Veränderung SpO₂ in Ruhe • Veränderung DL_{CO} in Ruhe <p>Sicherheit</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse • Vitalzeichen • Körperliche Untersuchung • Gewicht • Klinische Labortests • 12-Kanal-EKG (japanische Studienzentren) <p>Andere Endpunkte</p> <ul style="list-style-type: none"> • Prä-Dosis Plasmakonzentration (C_{pre,ss,N}) von Nintedanib, BIBF 1202 und BIBF 1202 Glukuronid Visit 4 (4 Wochen) und Visit 7 (24 Wochen) • Pharmakogenomische Auswertung • Biomarker-Auswertung
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Keine
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<ul style="list-style-type: none"> • Gepowert auf primären Endpunkt: Jährliche FVC-Abnahme über 52 Wochen (ml) • Erwarteter Effekt: Jährliche Differenz von 100 ml zwischen Nintedanib- und Placebo-Gruppe. • Erwartete Abweichung: 300 ml (auf Basis der Phase II-Studienergebnisse) • Erwarteter Anteil nicht auswertbarer Studienteilnehmer: 2 % • Zuteilungsverhältnis von 3:2 führt zu einer Fallzahl von: 291 zu 194 (n = 485) in Nintedanib zu Placebo • Verwendeter Test: t-Test • Teststärke: 90 % • (α): 0,025, einseitig • H₀-Hypothese: Es gibt keinen Unterschied zwischen den Behandlungsgruppen.

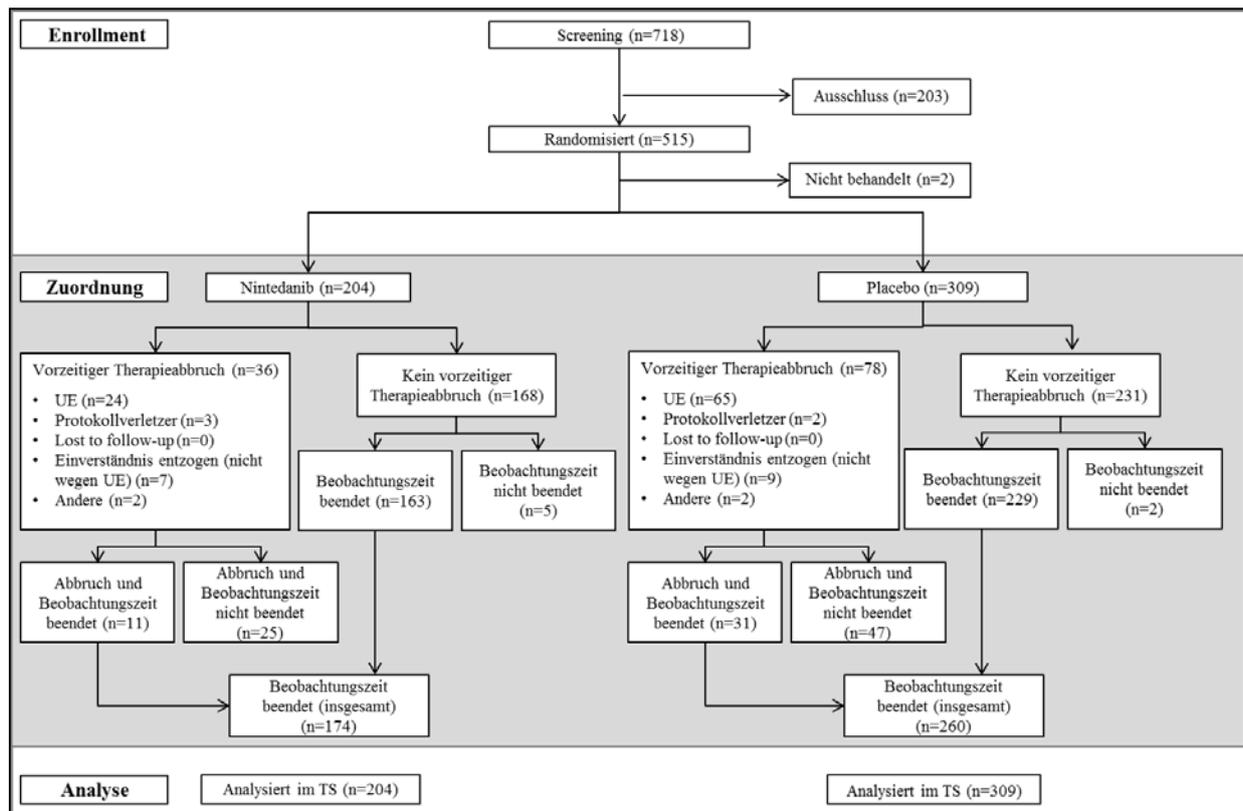
Item ^a	Charakteristikum	Studieninformation
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	<p>Zwischenanalyse</p> <ul style="list-style-type: none"> • keine <p>Kriterien für Patientenausschluss</p> <ul style="list-style-type: none"> • direkt nach Visit 2 (vor Arzneimittel Ersteinnahme) <ul style="list-style-type: none"> • AST, ALT > 1,5-fach ULN • Bilirubin > 1,5-fach ULN • INR > 2 • Prothrombinzeit oder PTT > 50 % ULN • Sonst keine weiteren Regeln für Patientenausschluss vorgesehen <p>Kriterien für patientenindividuellen Therapieausschluss</p> <ul style="list-style-type: none"> • Leberfunktion <ul style="list-style-type: none"> • Leberenzyme <ul style="list-style-type: none"> -ALT oder AST ≥ 8-fach ULN -ALT oder AST ≥ 3-fach ULN für mehr als 2 Wochen • Anzeichen akuter Leberschaden <ul style="list-style-type: none"> • ALT oder AST ≥ 3-fach ULN und Gesamt-Bilirubin ≥ 1,5-fach ULN oder INR > 1,5 oder Anzeichen von Fatigue, Übelkeit, Erbrechen, Schmerz oder Druckempfindlichkeit im rechten oberen abdominalen Quadranten, Fieber, Ausschlag, und/oder Eosinophile > 5 % • Andere Kriterien <ul style="list-style-type: none"> • Wiederauftreten einer Diarrhöe (> 4 Stuhlgänge/Tag) für 8 oder mehr aufeinander folgende Tage trotz Dosisreduktion und Begleitbehandlung • Gleichzeitige Einnahme von gemäß der Ausschlusskriterien verbotenen Arzneimitteln • Schwangerschaft • Widerruf der Einwilligung <p>Kriterien für dauerhaften Abbruch der Studienmedikation</p> <ul style="list-style-type: none"> • Größere Operation (abdominal und intestinal) • Patienten mit vollständiger Antikoagulation • Größere thromboembolische Ereignisse (Schlaganfall, tiefe Venenthrombose, Lungenembolie, Myokardinfarkt) • Erhöhtes Blutungsrisiko (hämorrhagisches ZNS-Ereignis, Hämoptysis oder Hämaturie, aktive gastrointestinale Blutung oder Geschwüre) <p>Kriterien für Studienausschluss</p>

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Widerruf der Einwilligung <p>Kriterien für Studienabbruch durch Sponsor</p> <ul style="list-style-type: none"> • Nichteinhaltung der Aufnahmekriterien • Auftreten von Wirkung/Sicherheit Informationen, die möglicherweise die Fortsetzung des Tests beeinflusst haben • Verstoß GCP, CTP oder Vertragsbruch seitens Prüfzentrum oder Prüfer
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Randomisierung über Interaktives Telefon/Internet-Response-System (IXRS)
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	<ul style="list-style-type: none"> • Blockrandomisierung: Blockgröße 5 • Nicht stratifizierte Randomisierung
9	<p>Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment)</p> <p>Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war</p>	<ul style="list-style-type: none"> • Zuteilung eines nummerierten Behandlungskits erfolgte mittels IVXS gemäß Randomisierungsliste des Sponsors. • Zugewiesene Arzneimittelnummer wurde im eCRF dokumentiert. • Zuteilung der Arzneimittel durch ausgewiesenes Personal. • Das Studienpersonal erhielt keine Information über die Zuteilung.
10	<p>Randomisierung, Durchführung</p> <p>Wer hat die Randomisierungsliste erstellt, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?</p>	<ul style="list-style-type: none"> • Erstellung der Randomisierungsliste durch den Sponsor und Abruf der Zuteilung durch den Prüfarzt über IVXS • Aufnahme der Probanden in die Studie durch Prüfarzt und weiteres Studienpersonal
11	Verblindung	
11a	<p>Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?</p>	<p>a) verblindet b) verblindet c) verblindet</p> <p>Verblindung wurde gewährleistet durch</p> <ul style="list-style-type: none"> • Zuteilung der Behandlung durch IVXS • Identifikation der Studienmedikation über Identifikations-Codes • Identisches Aussehen der Verpackung und Beschriftung sowie der Kapseln von Verum und Placebo

Item ^a	Charakteristikum	Studieninformation
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	<ul style="list-style-type: none"> • Identische Verpackung und Beschriftung der Studienmedikation • Verschiedene Farben der 150 mg- und 100 mg-Kapseln, jedoch für Nintedanib und Placebo identisch • Identische Einnahmezeitpunkte von Verum und Placebo
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p>Primäre Zielkriterien</p> <ul style="list-style-type: none"> • Jährliche Abnahme FVC in ml über 52 Wochen mittels randomisiertem Koeffizienz-Regressionsmodell. <p>Sekundäre Zielkriterien</p> <ul style="list-style-type: none"> • Veränderung des SGRQ-Gesamtscore bis Woche 52 mittels MMRM. • Zeit der ersten akuten Exazerbation über 52 Wochen (proportionales Cox-Hazardmodell, log-rank Test, Kaplan-Meier). <p>Für alle weiteren sekundäre Zielkriterien</p> <ul style="list-style-type: none"> • Veränderung FVC bis Woche 52 mittels MMRM • Absolute und relative Veränderung FVC bis Woche 52 mittels MMRM • Alle weiteren kontinuierlichen Endpunkte (PROs) mittels MMRM. • Alle Responder-Analysen (95 % KI nach Wilson, logistische Regression). • Alle weiteren Überlebensanalysen (proportionales Cox-Hazardmodell, log-rank Test, Kaplan-Meier)
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<ul style="list-style-type: none"> • Japanisch oder Nicht-Japanisch <p>Zusätzlich für die gepoolte Auswertung der INPULSIS Studien:</p> <ul style="list-style-type: none"> • Geschlecht (männlich/weiblich) • Alter (< 65, ≥ 65 Jahre) • Ethnie (kaukasisch, asiatisch) • Raucherstatus (Nichtraucher, Ex-Raucher/Raucher) • FVC % pred. (≤ 70 %, > 70 %) • SGRQ-Gesamtscore (≤ 40, > 40 Punkte) • systemische Kortikosteroid-Therapie (ja, nein) • Verwendung Bronchodilator (ja, nein) • Dosisintensität (≤ 90 %, > 90 %) (nur für den primären Endpunkt)
Resultate		
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	s. Flow-Chart

Item ^a	Charakteristikum	Studieninformation
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	a) n = 515 Nintedanib 150 mg bid: n = 309 Placebo: n = 206 b) n = 513 Nintedanib 150 mg bid: n = 309 Placebo: n = 204 c) n = 513 Nintedanib 150 mg bid: n = 309 Placebo: n = 204
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Vorzeitiger Abbruch der Studienmedikation <u>Nintedanib: n = 78 (25,2 %)</u> <ul style="list-style-type: none"> • n = 65 (21,0 %) unerwünschte Ereignisse <ul style="list-style-type: none"> • n = 9 (2,9 %) unerwartete Verschlechterung der IPF • n = 2 (0,6 %) unerwartete Verschlechterung einer anderen bereits vorhandenen Krankheit • n = 54 (17,5 %) andere unerwünschte Ereignisse • n = 2 (0,6 %) Protokollverletzer • n = 0 Lost to follow-up • n = 9 (2,9 %) Verweigerung der Medikation • n = 2 (0,6 %) andere <u>Placebo: n = 36 (17,6 %)</u> <ul style="list-style-type: none"> • n = 24 (11,8 %) unerwünschte Ereignisse <ul style="list-style-type: none"> • n = 9 (4,4 %) unerwartete Verschlechterung der IPF • n = 2 (1,0 %) unerwartete Verschlechterung einer anderen bereits vorhandenen Krankheit • n = 13 (6,4 %) andere unerwünschte Ereignisse • n = 3 (1,5 %) Protokollverletzer • n = 0 Lost to follow-up • n = 7 (3,4 %) Verweigerung der Medikation • n = 2 (1,0 %) andere
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	<ul style="list-style-type: none"> • Studienaufnahme: 05/2011 bis 10/2013 • Follow-up: 28 Tage
14b	Informationen, warum die Studie endete oder beendet wurde	Beendigung der Studie gemäß Studienprotokoll
a: nach CONSORT 2010		

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014c)

Abbildung 4-13: Patientenfluss INPULSIS-1 Studie

Tabelle 4-82 (Anhang): Studiendesign und -methodik für Studie INPULSIS-2

Item ^a	Charakteristikum	Studieninformation
Studienziel		
2b	Genauere Ziele, Fragestellung und Hypothesen	<p>Zielsetzung</p> <p>Untersuchung der Wirksamkeit sowie des Nutzen-Risiko-Verhältnisses von oral verabreichtem Nintedanib 150 mg bid im Vergleich zu Placebo bei IPF. Als primärer Endpunkt wird die Verringerung der Lungenfunktionsabnahme, gemessen als jährliche Abnahme der FVC, ermittelt.</p> <p>Hypothesen</p> <p>Bewertung der Überlegenheit von Nintedanib 150 mg bid im Vergleich zu Placebo in Bezug auf die jährliche Abnahme der FVC bei Patienten mit IPF.</p> <p>FVC-Abnahme in 52 Wochen:</p> <ul style="list-style-type: none"> • Nullhypothese (H₀): Keine Unterschiede der jährlichen Abnahme der FVC von Baseline bis Woche 52 zwischen den Gruppen Nintedanib 150 mg bid und Placebo. • Alternativhypothese (H_a): Die jährliche Abnahme der FVC von Baseline bis Woche 52 ist bei Patienten unter Nintedanib 150 mg bid geringer als unter Placebo.
Methoden		
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	<p>Allokation: randomisiert</p> <p>Zuteilungsverhältnis: 3:2</p> <p>Randomisierungseinheit: Studienteilnehmer</p> <p>Verblindung: doppelblind</p> <p>Studienhorizont: 52 Wochen + 28 Tage Nachbeobachtung</p> <p>Design: parallel, zweiarmig, Placebo-kontrolliert</p> <p>Studienorganisation: multizentrisch</p> <p>Phase: III</p>
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p>Amendment 1 (11/2011)</p> <ul style="list-style-type: none"> • Möglichkeit eines Re-Tests bei abnormalen Laborparametern zu Visit 1, bei Verdachtsfällen auf Messfehler oder fehlendem bzw. angezweifeltem Bezug auf den Patientenzustand. • Studienausschluss, wenn Patienten den Studienablauf nicht verstehen bzw. die Fragebögen nicht selbstständig ausfüllen können • Ergänzung der „schwerwiegenden unerwünschten Ereignisse“ entsprechend den neuen BI Standards, um deren ordnungsgemäße Dokumentation zu gewährleisten. • Einschlusskriterium 4 wurde abgeändert: „Thorax-HRCT innerhalb von 12 Monaten vor Visit 1“ anstelle

Item ^a	Charakteristikum	Studieninformation
		<p>von „Thorax HRCT innerhalb von 12 Monaten vor Visit 2“</p> <p>Amendment 2 (09/2012)</p> <ul style="list-style-type: none"> • Ergänzung explorativer Biomarker Analysen, um den Effekt von Nintedanib auf Biomarker in Verbindung mit IPF Pathologie und prognostischen Erkrankungsmarkern zu untersuchen, und Ergänzung pharmakogenomischer Analysen • Schlechte Compliance wurde als Protokollverletzung definiert
4	Probanden / Patienten	.
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • Schriftliche Einwilligungserklärung vor Studienbeginn • Alter \geq 40 Jahre zu Visit 1 • Diagnose IPF nach ATS-, ERS-, JRS- und ALAT-Richtlinien (innerhalb von 5 Jahren vor Baseline) • Thorax HRCT innerhalb von 12 Monaten vor Visit 1) • HRCT-Muster und, wenn verfügbar, Lungenbiopsie-Muster übereinstimmend mit IPF Diagnose bewertet von zentralen Reviewern • DL_{CO}: 30 % bis 79 % (zu Visit 2; angepasst an Hämoglobin) • FVC \geq 50 % vom Sollwert (zu Visit 2) <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> • Laborparameter (zu Visit 1) <ul style="list-style-type: none"> • AST und ALT > 1,5-fach ULN • Bilirubin > 1,5-fach ULN • Lungenfunktion (zu Visit 2) <ul style="list-style-type: none"> • Atemwegsobstruktion (FEV₁/FVC < 0,7) • Lungentransplantation im Verlauf der Studie wahrscheinlich (nach Prüfarzt-Einschätzung) • Andere Krankheiten <ul style="list-style-type: none"> • Herzerkrankung <ul style="list-style-type: none"> - Myokardinfarkt innerhalb 6 Monaten vor Visit 2 - Instabile Angina innerhalb 1 Monat vor Visit 2 • Blutungsrisiko <ul style="list-style-type: none"> - Bekannte Prädisposition für Blutungen - Fibrinolyse mit im Studienprotokoll festgelegten Interventionen - Hämorrhagisches ZNS-Ereignis innerhalb von 12 Monaten vor Visit 2 - weitere Kriterien innerhalb von 3 Monaten vor Visit 2

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> - Hämoptysis oder Hämaturie - Aktive gastrointestinale Blutung oder Geschwüre - Größere Verletzung oder Operation - Koagulations Parameter: INR > 2, Verlängerung der Prothrombinzeit und PTT (sek) um > 50 % ULN (Visit 1) • 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 Visit 2 • Bekannte Hypersensitivität gegenüber Studienmedikation oder deren Bestandteilen • Andere Krankheiten, die das Testprozedere oder die Studienteilnahme beeinflussen oder den Patienten bei Studienteilnahme in Gefahr bringen (Ermessen des Prüfarztes) • Lebenserwartung aufgrund anderer Krankheiten außer IPF < 2,5 Jahre (Ermessen des Prüfarztes) • Allgemeine Ausschlusskriterien <ul style="list-style-type: none"> • Frühere Behandlung mit Nintedanib länger als 4 Wochen) • Andere Interventionen (Forschungsstudie) innerhalb von 8 Wochen vor Visit 1 • Einnahme von N-Acetylcystein (NAC), Prednison > 15 mg/Tag oder vergleichbare Arzneimittel innerhalb von 2 Wochen vor Visit 1 • Einnahme von Pirfenidon, Azathioprin, Cyclophosphamid, Cyclosporin A innerhalb von 8 Wochen vor Visit 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 gebärfähigem Alter sind) • Aktiver Alkohol- oder Drogenmissbrauch

Item ^a	Charakteristikum	Studieninformation
		Fragebögen ohne Hilfe)
4b	Studienorganisation und Ort der Studiendurchführung	Multizentrische Studie, 107 Studienzentren in 17 Ländern: Kanada, Chile, China, Finnland, Frankreich, Deutschland, Griechenland, Indien, Japan, Korea, Mexiko, Niederlande, Portugal, Russland, Spanien, Türkei, USA
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<p>Alle Gruppen</p> <p>Visit 1: Screening</p> <p>Visit 2: Therapiestart und nachfolgende Einnahme des Arzneimittels über 52 Wochen</p> <p>Nintedanib (oral)</p> <ul style="list-style-type: none"> • 2-mal täglich 150 mg im Intervall von 12 Stunden über 52 Wochen • Bei unerwünschten Ereignissen Reduktion der Dosis auf 2-mal täglich 100 mg im Intervall von 12 Stunden über 52 Wochen bzw. in besonderen Fällen Unterbrechung der Therapie von maximal 4 Wochen bei therapieassoziierten unerwünschten Ereignissen und bis zu 8 Wochen bei Exazerbationen <p>Placebo (oral)</p> <ul style="list-style-type: none"> • 2-mal täglich im Intervall von 12 Stunden über 52 Wochen • Bei unerwünschten Ereignissen Reduktion der Dosis auf 2-mal täglich Placebo zu 100 mg im Intervall von 12 Stunden über 52 Wochen bzw. in besonderen Fällen Unterbrechung der Therapie von maximal 4 Wochen bei therapieassoziierten unerwünschten Ereignissen und bis zu 8 Wochen bei Exazerbationen <p>Alle Gruppen</p> <p>Follow-up: 28 Tage nach Therapieende</p>
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Primäres Zielkriterium</p> <p>Untersuchung der Wirksamkeit sowie des Nutzen/Risiko-Verhältnisses von Nintedanib 150 mg bid im Vergleich zu Placebo bei IPF.</p> <p>Primäre Endpunkte</p> <ul style="list-style-type: none"> • Jährliche FVC- Abnahme über 52 Wochen (ml) <p>Sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Veränderung des <i>Saint George's Respiratory Questionnaire</i> Gesamtscore von Baseline bis Woche 52 (in Punkten) • Zeit bis zur ersten akuten Exazerbation über 52 Wochen (in Tagen) <p>Weitere sekundäre Endpunkte</p> <ul style="list-style-type: none"> • Weitere Analysen der FVC

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Absolute und relative Veränderung der FVC (ml) und FVC (% vom Sollwert) über 52 Wochen • Absolute kategorielle Veränderung der FVC (% vom Sollwert) bis zu 52 Wochen: Abfall > 5 %, Anstieg > 5 % und Veränderungen ≤ 5 % • Absolute kategorielle Veränderung der FVC (% vom Sollwert) von Baseline bis Woche 52: Abfall > 10 %, Anstieg > 10 % und Veränderung ≤ 10 % • Anteil FVC-Responder mit Abfall < 5 % vom FVC-Sollwert nach 52 Wochen • Anteil FVC-Responder mit Abfall < 10 % vom FVC-Sollwert nach 52 Wochen • Patient Reported Outcomes (PRO) gemessen von Baseline bis Woche 52: <ul style="list-style-type: none"> • 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[CD])-Score • Anteil PGI-C Responder • Veränderung EQ-5D-VAS • Akute Exazerbation: <ul style="list-style-type: none"> • Risiko einer akuten IPF Exazerbation • Überlebenszeitanalysen gemessen über 52 Wochen: <ul style="list-style-type: none"> • 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 DL_{CO} < 30 % vom Sollwert oder SpO₂ < 88 % in Ruhe) • Veränderung SpO₂ in Ruhe • Veränderung DL_{CO} in Ruhe <p>Sicherheit</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse • Vitalzeichen • Körperliche Untersuchung • Gewicht • Klinische Labortests

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • 12-Kanal-EKG (japanische Studienzentren) <p>Andere Endpunkte</p> <ul style="list-style-type: none"> • Prä-Dosis Plasmakonzentration ($C_{pre,ss,N}$) von Nintedanib, BIBF 1202 und BIBF 1202 Glukuronid Visit 4 (4 Wochen) und Visit 7 (24 Wochen) • Pharmakogenomische Auswertung • Biomarker-Auswertung
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	keine
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<ul style="list-style-type: none"> • Gepowert auf primären Endpunkt: Jährliche FVC-Abnahme über 52 Wochen (ml) • Erwarteter Effekt: Jährliche Differenz von 100 ml zwischen Nintedanib- und Placebo-Gruppe. • Erwartete Abweichung: 300 ml (auf Basis der Phase II-Studienergebnisse) • Erwarteter Anteil nicht auswertbarer Studienteilnehmer: 2 % • Zuteilungsverhältnis von 3:2 führt zu einer Fallzahl von: 291 zu 194 (n = 485) in Nintedanib zu Placebo • Verwendeter Test: t-Test • Teststärke: 90 % • (α): 0,025, einseitig • H0-Hypothese: Es gibt keinen Unterschied zwischen den Behandlungsgruppen.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	<p>Zwischenanalyse</p> <ul style="list-style-type: none"> • keine <p>Kriterien für Patientenausschluss</p> <ul style="list-style-type: none"> • direkt nach Visit 2 (vor Arzneimittel Ersteinnahme) <ul style="list-style-type: none"> • AST, ALT > 1,5-fach ULN • Bilirubin > 1,5-fach ULN • INR > 2 • Prothrombinzeit oder PTT > 50 % ULN • Sonst keine weiteren Regeln für Patientenausschluss vorgesehen <p>Kriterien für Therapieausschluss</p> <ul style="list-style-type: none"> • Leberfunktion <ul style="list-style-type: none"> • Leberenzyme <ul style="list-style-type: none"> -ALT oder AST \geq 8-fach ULN -ALT oder AST \geq 3-fach ULN für mehr als 2 Wochen

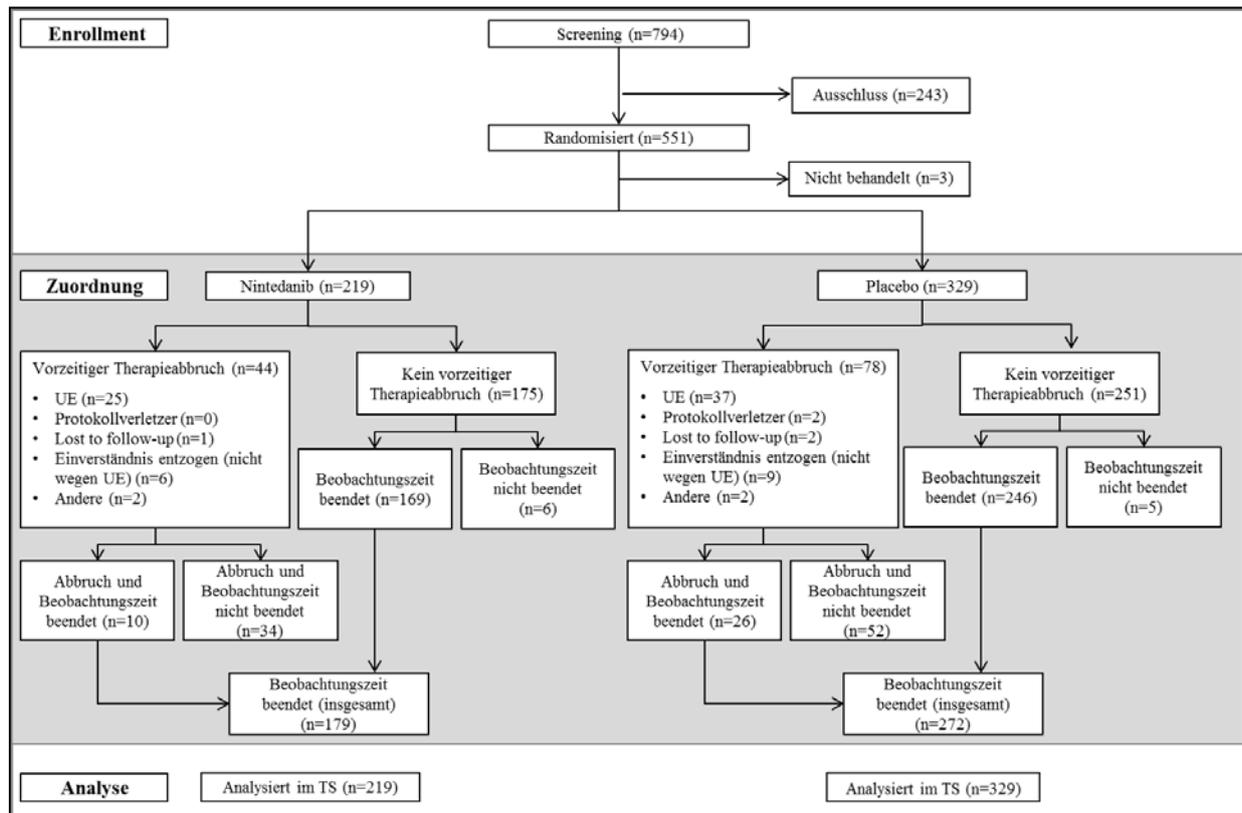
Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Anzeichen akuter Leberschaden <ul style="list-style-type: none"> • ALT oder AST \geq 3-fach ULN und Gesamt-Bilirubin \geq 1,5-fach ULN oder INR $>$ 1,5 oder Anzeichen von Fatigue, Übelkeit, Erbrechen, Schmerz oder Druckempfindlichkeit im rechten oberen abdominalen Quadranten, Fieber, Ausschlag, und/oder Eosinophile $>$ 5 % • Andere Kriterien <ul style="list-style-type: none"> • Wiederauftreten einer Diarrhöe ($>$ 4 Stuhlgänge/Tag) für 8 oder mehr aufeinander folgende Tage trotz Dosisreduktion und Begleitbehandlung • Gleichzeitige Einnahme von gemäß der Ausschlusskriterien verbotenen Arzneimitteln • Schwangerschaft • Widerruf der Einwilligung <p>Kriterien für dauerhaften Abbruch der Studienmedikation</p> <ul style="list-style-type: none"> • Größere Operation (abdominal und intestinal) • Patienten mit vollständiger Antikoagulation • Größere thromboembolische Ereignisse (Schlaganfall, tiefe Venenthrombose, Lungenembolie, Myokardinfarkt) • Erhöhtes Blutungsrisiko (hämorrhagisches ZNS-Ereignis, Hämoptysis oder Hämaturie, aktive gastrointestinale Blutung oder Geschwüre) <p>Kriterien für Studienausschluss</p> <ul style="list-style-type: none"> • Widerruf der Einwilligung <p>Kriterien für Studienabbruch durch Sponsor</p> <ul style="list-style-type: none"> • Nichteinhaltung der Aufnahmekriterien • Auftreten von Wirkung/Sicherheit Informationen, die möglicherweise die Fortsetzung des Tests beeinflusst haben • Verstoß GCP, CTP oder Vertragsbruch seitens Prüfzentrum oder Prüfer
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Randomisierung über Interaktives Telefon/Internet-Response-System (IXRS)
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	<ul style="list-style-type: none"> • Blockrandomisierung: Blockgröße 5 • Nicht stratifizierte Randomisierung

Item ^a	Charakteristikum	Studieninformation
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	<ul style="list-style-type: none"> • Zuteilung eines nummerierten Behandlungskits erfolgte mittels IVXS gemäß Randomisierungsliste des Sponsors. • Zugewiesene Arzneimittelnummer wurde im eCRF dokumentiert. • Zuteilung der Arzneimittel durch ausgewiesenes Personal. • Das Studienpersonal erhielt keine Information über die Zuteilung.
10	Randomisierung, Durchführung Wer hat die Randomisierungsliste erstellt, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	<ul style="list-style-type: none"> • Erstellung der Randomisierungsliste durch den Sponsor und Abruf der Zuteilung durch den Prüfarzt über IVXS • Aufnahme der Probanden in die Studie durch Prüfarzt und weiteres Studienpersonal
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	a) verblindet b) verblindet c) verblindet Verblindung wurde gewährleistet durch <ul style="list-style-type: none"> • Zuteilung der Behandlung durch IVXS • Identifikation der Studienmedikation über Identifikations-Codes • Identisches Aussehen der Verpackung und Beschriftung sowie der Kapseln von Verum und Placebo
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	<ul style="list-style-type: none"> • Identische Verpackung und Beschriftung der Studienmedikation • Verschiedene Farben der 150 mg- und 100 mg-Kapseln, jedoch für Nintedanib und Placebo identisch • Identische Einnahmezeitpunkte von Verum und Placebo
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p>Primäre Zielkriterien</p> <ul style="list-style-type: none"> • Jährliche Abnahme FVC in ml über 52 Wochen mittels randomisiertem Koeffizienz-Regressionsmodell. <p>Sekundäre Zielkriterien</p> <ul style="list-style-type: none"> • Veränderung des SGRQ-Gesamtscores bis Woche 52 mittels MMRM. • Zeit der ersten akuten Exazerbation über 52 Wochen (proportionales Cox-Hazardmodell, log-rank Test, Kaplan-Meier). <p>Für alle weiteren sekundäre Zielkriterien</p> <ul style="list-style-type: none"> • Veränderung FVC bis Woche 52 mittels MMRM

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Absolute und relative Veränderung FVC bis Woche 52 mittels MMRM • Alle weiteren kontinuierlichen Endpunkte (PROs) mittels MMRM. • Alle Responder-Analysen (95 % KI nach Wilson, logistische Regression). • Alle weiteren Überlebensanalysen (proportionales Cox-Hazardmodell, log-rank Test, Kaplan-Meier)
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<ul style="list-style-type: none"> • Japanisch oder Nicht-Japanisch Zusätzlich für die gepoolte Auswertung der INPULSIS Studien: <ul style="list-style-type: none"> • Geschlecht (männlich/weiblich) • Alter (< 65, ≥ 65 Jahre) • Ethnie (kaukasisch, asiatisch) • Raucherstatus (Nichtraucher, Ex-Raucher/Raucher) • FVC % vom Sollwert (≤ 70 %, > 70 %) • SGRQ-Gesamtscore (≤ 40, > 40 Punkte) • systemische Kortikosteroid-Therapie (ja, nein) • Verwendung Bronchodilator (ja, nein) • Dosisintensität (≤ 90 %, > 90 %) (nur für den primären Endpunkt)
Resultate		
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	s. Flow-Chart
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	a) n = 551 Nintedanib 150 mg bid: n = 331 Placebo: n = 220 b) n = 548 Nintedanib 150 mg bid: n = 329 Placebo: n = 219 c) n = 548 Nintedanib 150 mg bid: n = 329 Placebo: n = 219
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Vorzeitiger Abbruch der Studienmedikation <u>Nintedanib: n = 78 (23,7 %)</u> <ul style="list-style-type: none"> • n = 62 (18,8 %) unerwünschte Ereignisse <ul style="list-style-type: none"> • n = 10 (3,0 %) unerwartete Verschlechterung der Krankheit • n = 3 (0,9 %) unerwartete Verschlechterung einer bereits vorhandenen Krankheit • n = 49 (14,9 %) andere unerwünschte Ereignisse

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • n = 2 (0,6 %) Protokollverletzer • n = 0 Lost to follow-up • n = 11 (3,3 %) Verweigerung der Medikation • n = 3 (0,9 %) andere <p><u>Placebo: n = 44 (20,1 %)</u></p> <ul style="list-style-type: none"> • n = 35 (16,0) unerwünschte Ereignisse <ul style="list-style-type: none"> • n = 18 (8,2 %) unerwartete Verschlechterung der Krankheit • n = 2 (0,9 %) unerwartete Verschlechterung einer bereits vorhandenen Krankheit • n = 15 (6,8 %) andere unerwünschte Ereignisse • n = 1 (0,5 %) Protokollverletzer • n = 1 (0,5 %) Lost to follow-up • n = 6 (2,7 %) Verweigerung der Medikation • n = 1 (0,5 %) andere
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	<ul style="list-style-type: none"> • Studienaufnahme: 05/2011 bis 10/2013 • Follow-up: 28 Tage
14b	Informationen, warum die Studie endete oder beendet wurde	Beendigung der Studie gemäß Studienprotokoll
a: nach CONSORT 2010		

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.



Quelle: (Boehringer Ingelheim Pharma GmbH & Co. KG, 2014d)

Abbildung 4-14: Patientenfluss INPULSIS-2 Studie

Anhang 4-F: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten

Der nachfolgend dargestellte Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- Verzerrungsaspekte auf Studienebene. In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- Verzerrungsaspekte auf Endpunktebene. In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich sollen die Bögen studienbezogen ausgefüllt werden. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen ist die Blankoversion des Bogens. Dieser Blankobogen ist für jede Studie heranzuziehen. Im Anschluss daran ist ein Bewertungsbogen inklusive Ausfüllhinweisen abgebildet, der als Ausfüllhilfe dient, aber nicht als Vorlage verwendet werden soll.

Beschreiben Sie nachfolgend die Verzerrungsaspekte jeder eingeschlossenen Studie (einschließlich der Beschreibung für jeden berücksichtigten Endpunkt). Erstellen Sie hierfür je Studie eine separate Version des nachfolgend dargestellten Bewertungsbogens.

Tabelle 4-83 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie INPULSIS-1

Studie: INPULSIS-1**Tabelle: Liste der für die Bewertung herangezogenen Quellen**

Genauere Benennung der Quelle	Kürzel
Boehringer Ingelheim. Clinical Trial Report 1199-0032	INPULSIS-1 CTR
Boehringer Ingelheim. Clinical Trial Protocol 1199-0032	INPULSIS-1 CTP
Boehringer Ingelheim Trial Statistical Analysis Plan 1199-0032	INPULSIS-1 TSAP
Richeldi et al. 2014b Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis	Richeldi 2014

A: Verzerrungsaspekte auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Computergenerierte Zuteilung zu einer Behandlungsgruppe.

1.

für randomisierte Studien: Adäquate Erzeugung der Randomisierungssequenz

ja **unklar** **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen

ja **unklar** **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“) ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. **Verblindung von Patienten und behandelnden Personen****Patient:** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

Die Gruppenzuteilung blieb während der gesamten Studie verdeckt, so dass der Patient keine Rückschlüsse auf die Behandlung ziehen konnte.

behandelnde bzw. weiterbehandelnde Personen: ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

Die Gruppenzuteilung blieb während der gesamten Studie verdeckt, so dass die behandelnde bzw. weiterbehandelnde Person keine Rückschlüsse auf die Behandlung ziehen konnte.

4. **Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

niedrig hoch

Begründung für die Einstufung:

Es handelt sich um eine doppelblinde, randomisierte, Placebo-kontrollierte, multizentrische Studie. Randomisierung und Verblindung wurden adäquat durchgeführt und über den gesamten Studienzeitraum gewährleistet. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Studienebene ist somit als niedrig einzustufen.

B: Verzerrungsaspekte auf Endpunktebene pro Endpunkt**Endpunkt: Zeit bis Tod unabhängig von der Ursache****1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die Feststellung eines Todesfalles ist zudem objektiv möglich.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis respiratorisch bedingtem Tod**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die Feststellung eines Todesfalles ist zudem objektiv möglich. Die Beurteilung der Todesursache wurde sowohl durch den Prüfarzt als auch durch ein unabhängiges Bewertungskomitee anhand der Patientenakten beurteilt. Dadurch kann von einer objektiven Einschätzung der Todesursache ausgegangen werden.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Jährliche FVC-Abnahme (ml)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die FVC ist zudem objektiv bestimmbar.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: FVC-Responder (Schwellenwert <5 % und <10 % vom Sollwert)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die FVC ist zudem objektiv bestimmbar.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Schwellenwerte 5 % und 10 % vom Sollwert sind anerkannt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis zur ersten akuten Exazerbation**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt, einem Facharzt mit Erfahrung in der Behandlung der IPF.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis zur ersten adjudizierten akuten Exazerbation**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Zusätzlich wurden die Exazerbationen durch ein unabhängiges Bewertungskomitee anhand der Patientenakten beurteilt. Dadurch kann von einer objektiven Einschätzung des Ereignisses Exazerbation ausgegangen werden.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SGRQ-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: SGRQ-(\leq -4 Punkte)-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Als Cut off-Wert für die Response wurde die für COPD validierte und für IPF ebenfalls akzeptierte MCID von vier Punkten zugrunde gelegt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SGRQ-I-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: SGRQ-I (≤ -4 Punkte)-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Als Cut off-Wert für die Response wurde die für COPD validierte und für IPF ebenfalls akzeptierte MCID von vier Punkten zugrunde gelegt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SOBQ-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung CASA-Q Husten-Symptome und Husten-Impact**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: PGI-C-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung EQ-5D-VAS**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Unerwünschte Ereignisse (Gesamtzahl UE, SUE, UE zum Therapieabbruch oder Tod führend, UE von besonderen Interesse)

1. Verblindung der Endpunkterheber

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Auswertungskriterien waren im Prüfplan a priori definiert. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Tabelle 4-84 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie INPULSIS-2

Studie: INPULSIS-2

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Boehringer Ingelheim. Clinical Trial Report 1199-0034	INPULSIS-2 CTR
Boehringer Ingelheim. Clinical Trial Protocol 1199-0034	INPULSIS-2 CTP
Boehringer Ingelheim Trial Statistical Analysis Plan 1199-0034	INPULSIS-2 TSAP
Richeldi 2014. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis	Richeldi 2014

A: Verzerrungsaspekte auf Studienebene:**Einstufung als randomisierte Studie** **ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien **nein:** → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

 Computergenerierte Zuteilung zu einer Behandlungsgruppe.

1.

für randomisierte Studien: Adäquate Erzeugung der Randomisierungssequenz **ja** **unklar** **nein**Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen** **ja** **unklar** **nein**Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“) ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Verblindung von Patienten und behandelnden Personen****Patient:** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

Die Gruppenzuteilung blieb während der gesamten Studie verdeckt, so dass der Patient keine Rückschlüsse auf die Behandlung ziehen konnte.

behandelnde bzw. weiterbehandelnde Personen: ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

Die Gruppenzuteilung blieb während der gesamten Studie verdeckt, so dass die behandelnde bzw. weiterbehandelnde Person keine Rückschlüsse auf die Behandlung ziehen konnte.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

niedrig **hoch**

Begründung für die Einstufung:

Es handelt sich um eine doppelblinde, randomisierte, Placebo-kontrollierte, multizentrische Studie. Randomisierung und Verblindung wurden adäquat durchgeführt und über den gesamten Studienzeitraum gewährleistet. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Studienebene ist somit als niedrig einzustufen.

B: Verzerrungsaspekte auf Endpunktebene pro Endpunkt**Endpunkt: Zeit bis Tod, unabhängig von der Ursache****1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die Feststellung eines Todesfalles ist zudem objektiv möglich.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis respiratorisch bedingtem Tod**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die Feststellung eines Todesfalles ist zudem objektiv möglich. Die Beurteilung der Todesursache wurde sowohl durch den Prüfarzt als auch durch ein unabhängiges Bewertungskomitee anhand der Patientenakten beurteilt. Dadurch kann von einer objektiven Einschätzung der Todesursache ausgegangen werden.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Jährliche FVC-Abnahme (ml)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die FVC ist zudem objektiv bestimmbar.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: FVC-Responder (Schwellenwert 5 % und 10 % vom Sollwert)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Die FVC ist zudem objektiv bestimmbar.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Schwellenwerte 5 % und 10 % vom Sollwert sind anerkannt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis zur ersten akuten Exazerbation**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt, einem Facharzt mit Erfahrung in der Behandlung der IPF.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Zeit bis zur ersten adjudizierten akuten Exazerbation**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind durch den Prüfarzt. Zusätzlich wurden die Exazerbationen durch ein unabhängiges Bewertungskomitee anhand der Patientenakten beurteilt. Dadurch kann von einer objektiven Einschätzung des Ereignisses Exazerbation ausgegangen werden.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SGRQ-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: SGRQ-(\leq -4 Punkte)-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Als Cut off-Wert für die Response wurde die für COPD validierte und für IPF ebenfalls akzeptierte MCID von vier Punkten zugrunde gelegt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SGRQ-I-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: SGRQ-I (≤ -4 Punkte)-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Als Cut off-Wert für die Response wurde die für COPD validierte und für IPF ebenfalls akzeptierte MCID von vier Punkten zugrunde gelegt. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung SOBQ-Gesamtscore**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung CASA-Q Husten-Symptome und Husten-Impact**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: PGI-C-Responder**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Veränderung EQ-5D-VAS**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine** ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können** ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:**Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):** niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Rücklaufquoten waren in den beiden Studienarmen vergleichbar. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Endpunkt: Unerwünschte Ereignisse (Gesamtzahl UE, SUE, UE zum Therapieabbruch oder Tod führend, UE von besonderem Interesse)

1. Verblindung der Endpunkterheber

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Datenerhebung erfolgte doppelblind.

2. Adäquate Umsetzung des ITT-Prinzips

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig hoch

Begründung für die Einstufung:

Die Verblindung der Endpunkterheber war über den gesamten Studienzeitraum gewährleistet. Das ITT-Prinzip wurde konsequent eingehalten. Weder eine ergebnisgesteuerte Berichterstattung noch sonstige Aspekte, die das Verzerrungspotenzial beeinflussen könnten, lagen vor. Die Auswertungskriterien waren im Prüfplan a priori definiert. Das Verzerrungspotenzial auf Endpunktebene ist somit als niedrig einzustufen.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunkübergreifend; B: endpunktspezifisch).

A: Verzerrungsaspekte auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anzeichen vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:**Verdeckung der Gruppenzuteilung („allocation concealment“)** **ja:** Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

 unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau. **nein:** Die Gruppenzuteilung erfolgte nicht verdeckt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**für nicht randomisierte Studien:****Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** **ja:** Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anzeichen dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

 unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau. **nein:** Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:**3. Verblindung von Patienten und behandelnden Personen****Patient** **ja:** Die Patienten waren verblindet. **unklar:** Es finden sich keine diesbezüglichen Angaben. **nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

behandelnde bzw. weiterbehandelnde Personen

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die behandelnden Personen nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*

- geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben
- Prüfen, ob nicht statistisch signifikante Ergebnisse weniger ausführlich dargestellt sind.
- Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.

Anzumerken ist, dass Anzeichen für eine ergebnisgesteuerte Darstellung eines Endpunkts zu Verzerrungen der Ergebnisse der übrigen Endpunkte führen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anzeichen dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, sind Verzerrungen für die anderen Endpunkte möglich. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Verzerrung der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und dies nicht zur Verzerrung anderer Endpunkte führt.

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrung führen können

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien
- intransparenter Patientenfluss
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:

B: Verzerrungsaspekte auf Endpunktebene pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: _____

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

2. Adäquate Umsetzung des ITT-Prinzips

Kommen in einer Studie Patienten vor, die die Studie entweder vorzeitig abgebrochen haben oder wegen Protokollverletzung ganz oder teilweise aus der Analyse ausgeschlossen wurden, so sind diese ausreichend genau zu beschreiben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen zu berücksichtigen (i. d. R. ITT-Analyse, siehe Äquivalenzstudien). Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass

solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:
