

Dokumentvorlage, Version vom 16.12.2021

Dossier zur Nutzenbewertung gemäß § 35a SGB V

Sotatercept (WINREVAIR®)

MSD Sharp & Dohme GmbH

Modul 4 A

Pulmonale Arterielle Hypertonie

Medizinischer Nutzen und
medizinischer Zusatznutzen,
Patientengruppen mit therapeutisch
bedeutsamem Zusatznutzen

Stand: 13.09.2024

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	4
Abbildungsverzeichnis	10
Abkürzungsverzeichnis	11
4 Modul 4 – allgemeine Informationen	14
4.1 Zusammenfassung der Inhalte von Modul 4.....	15
4.2 Methodik.....	24
4.2.1 Fragestellung.....	24
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung.....	26
4.2.3 Informationsbeschaffung.....	28
4.2.3.1 Studien des pharmazeutischen Unternehmers.....	28
4.2.3.2 Bibliografische Literaturrecherche.....	28
4.2.3.3 Suche in Studienregistern/Studienergebnisdatenbanken.....	30
4.2.3.4 Suche auf der Internetseite des G-BA.....	31
4.2.3.5 Selektion relevanter Studien.....	32
4.2.4 Bewertung der Aussagekraft der Nachweise.....	33
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.....	35
4.2.5.2.1 Patientencharakteristika.....	36
4.2.5.2.2 Patientenrelevanz der Endpunkte.....	36
4.2.5.2.3 Verwendete statistische Methoden.....	42
4.2.5.3 Meta-Analysen.....	45
4.2.5.4 Sensitivitätsanalysen.....	46
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren.....	49
4.2.5.6 Indirekte Vergleiche.....	54
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen.....	57
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	57
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel.....	57
4.3.1.1.1 Studien des pharmazeutischen Unternehmers.....	57
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche.....	60
4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken.....	61
4.3.1.1.4 Studien aus der Suche auf der Internetseite des G-BA.....	63
4.3.1.1.5 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	64
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	65
4.3.1.2.1 Studiendesign und Studienpopulationen.....	65
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	74

4.3.1.3	Ergebnisse aus randomisierten kontrollierten Studien.....	75
4.3.1.3.1	Endpunkte – RCT.....	77
4.3.1.3.1.1	Mortalität – RCT	80
4.3.1.3.1.1.1	Mortalität zu Woche 24	80
4.3.1.3.1.2	Morbidität – RCT	85
4.3.1.3.1.2.1	Verbesserung der 6MWD – RCT	85
4.3.1.3.1.2.2	Zeit bis zur klinischen Verschlechterung oder Tod – RCT	90
4.3.1.3.1.2.3	Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test – RCT.....	99
4.3.1.3.1.2.4	Mehrkomponentenverbesserung.....	102
4.3.1.3.1.2.5	Krankheitssymptomatik und Gesundheitszustand – RCT	106
4.3.1.3.1.3	Gesundheitsbezogene Lebensqualität – RCT	113
4.3.1.3.1.3.1	Gesundheitsbezogene Lebensqualität – RCT	113
4.3.1.3.1.4	Nebenwirkungen – RCT	117
4.3.1.3.1.4.1	Unerwünschte Ereignisse Gesamtraten – RCT	117
4.3.1.3.1.4.2	Unerwünschte Ereignisse (gegliedert nach SOC und PT) – RCT 120	
4.3.1.3.1.4.3	Unerwünschte Ereignisse von Interesse – RCT	127
4.3.1.3.2	Subgruppenanalysen – RCT.....	129
4.3.1.3.2.1	Überblick über Ergebnisse der Interaktionstests aus Subgruppenanalysen	133
4.3.1.3.2.2	Ergebnisse für Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$).....	143
4.3.1.4	Liste der eingeschlossenen Studien – RCT.....	149
4.3.2	Weitere Unterlagen.....	150
4.3.2.1	Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien	150
4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche	150
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	150
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen	151
4.3.2.1.3.1	<Endpunkt xxx> – indirekte Vergleiche aus RCT	151
4.3.2.1.3.2	Subgruppenanalysen – indirekte Vergleiche aus RCT	154
4.3.2.1.4	Liste der eingeschlossenen Studien – indirekte Vergleiche aus RCT...	154
4.3.2.2	Nicht randomisierte vergleichende Studien.....	154
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien	155
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	155
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien	156
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien.....	156
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien	158
4.3.2.2.4	Liste der eingeschlossenen Studien – nicht randomisierte vergleichende Studien	158
4.3.2.3	Weitere Untersuchungen.....	158
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen	158
4.3.2.3.2	Charakteristika der weiteren Untersuchungen	159
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen	159
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen	159

4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen.....	160
4.3.2.3.4	Liste der eingeschlossenen Studien – weitere Untersuchungen.....	160
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	161
4.4.1	Beurteilung der Aussagekraft der Nachweise	161
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	161
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht	167
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte	168
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	168
4.5.2	Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen.....	168
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen	168
4.5.4	Verwendung von Surrogatendpunkten.....	168
4.6	Referenzliste.....	170
Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche		173
Anhang 4-B : Suchstrategien – Suche in Studienregistern/ Studienergebnisdatenbanken.....		176
Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....		178
Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/ Studienergebnisdatenbanken).....		179
Anhang 4-E : Methodik der eingeschlossenen Studien – RCT		190
Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten		198
Anhang 4-G : Weitere Ergebnisse		217

Tabellenverzeichnis

	Seite
Tabelle 4-1: Übersicht der Ein- und Ausschlusskriterien für die Suche nach RCT mit dem zu bewertenden Arzneimittel zum medizinischen Nutzen und Zusatznutzen.....	16
Tabelle 4-2: Übersicht über die zur Ableitung des Zusatznutzens herangezogenen Endpunkte der Studie STELLAR.....	18
Tabelle 4-3: Übersicht der Ein- und Ausschlusskriterien für die Suche nach RCT mit dem zu bewertenden Arzneimittel zum medizinischen Nutzen und Zusatznutzen.....	27
Tabelle 4-4: Übersicht zu den Endpunktkategorien und Endpunkten.....	36
Tabelle 4-5: Systematik zur Darstellung des Endpunkts Unerwünschte Ereignisse (gegliedert nach SOC und PT)	44
Tabelle 4-6: Sensitivitätsanalyse für den Endpunkt „Zeit bis zur klinischen Verschlechterung oder Tod“	49
Tabelle 4-7: Kriterien für die Berechnung der Interaktionstests in Anlehnung an das IQWiG-Methodenpapier 7.0 (3).....	52
Tabelle 4-8: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel	58
Tabelle 4-9: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	60
Tabelle 4-10: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern/ Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel.....	63
Tabelle 4-11: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel.....	64
Tabelle 4-12: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	65
Tabelle 4-13: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel	66
Tabelle 4-14: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	68
Tabelle 4-15: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel.....	68
Tabelle 4-16: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	75
Tabelle 4-17: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel	75
Tabelle 4-18: Operationalisierung des Endpunkts Mortalität zu Woche 24	80
Tabelle 4-19: Bewertung des Verzerrungspotenzials für Mortalität zu Woche 24 in RCT mit dem zu bewertenden Arzneimittel	81
Tabelle 4-20: Ergebnisse für Mortalität zu Woche 24 aus RCT mit dem zu bewertenden Arzneimittel.....	82

Tabelle 4-21: Ergebnisse für Gesamtüberleben mit dem zu bewertenden Arzneimittel (Studie STELLAR und weiterführende Daten aus SOTERIA)	83
Tabelle 4-22: Ergebnisse für Gesamtüberleben mit dem zu bewertenden Arzneimittel (RPSFT-Methode, Studie STELLAR und weiterführende Daten aus der LTE-Studie SOTERIA).....	84
Tabelle 4-23: Operationalisierung des Endpunkts Verbesserung der 6MWD.....	86
Tabelle 4-24: Bewertung des Verzerrungspotenzials für Verbesserung der 6MWD in RCT mit dem zu bewertenden Arzneimittel	87
Tabelle 4-25: Ergebnisse für Verbesserung der 6MWD aus RCT mit dem zu bewertenden Arzneimittel – Responderanalyse.....	87
Tabelle 4-26: Ergebnisse für Veränderung der 6MWD aus RCT mit dem zu bewertenden Arzneimittel – Veränderung gegenüber Baseline	88
Tabelle 4-27: Operationalisierung von Zeit bis zur klinischen Verschlechterung oder Tod ...	90
Tabelle 4-28: Bewertung des Verzerrungspotenzials für Zeit bis zur klinischen Verschlechterung oder Tod in RCT mit dem zu bewertenden Arzneimittel.....	91
Tabelle 4-29: Ergebnisse für Zeit bis zur klinischen Verschlechterung oder Tod aus RCT mit dem zu bewertenden Arzneimittel – Hauptanalyse	92
Tabelle 4-30: Ergebnisse Zeit bis zur klinischen Verschlechterung oder Tod getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Hauptanalyse	93
Tabelle 4-31: Ergebnisse für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod aus RCT mit dem zu bewertenden Arzneimittel – Sensitivitätsanalyse	96
Tabelle 4-32: Ergebnisse für Zeit bis zur klinischen Verschlechterung oder Tod getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Sensitivitätsanalyse.....	97
Tabelle 4-33: Operationalisierung des Endpunkts Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test.....	99
Tabelle 4-34: Bewertung des Verzerrungspotenzials für Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test in RCT mit dem zu bewertenden Arzneimittel..	100
Tabelle 4-35: Ergebnisse für Verbesserung der Dyspnoe um 15 % der Skalenspannweite gemäß Borg CR10-Skala pre 6MWD-Test aus RCT mit dem zu bewertenden Arzneimittel – Responderanalysen.....	101
Tabelle 4-36: Ergebnisse für Veränderung der Dyspnoe gemäß Borg CR10-Skala aus RCT mit dem zu bewertenden Arzneimittel – Veränderung gegenüber Baseline	101
Tabelle 4-37: Operationalisierung des Endpunkts Mehrkomponentenverbesserung.....	102
Tabelle 4-38: Bewertung des Verzerrungspotenzials für Mehrkomponentenverbesserung in RCT mit dem zu bewertenden Arzneimittel	103
Tabelle 4-39: Ergebnisse für Mehrkomponentenverbesserung aus RCT mit dem zu bewertenden Arzneimittel	103

Tabelle 4-40: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung der 6MWD (≥ 30 m)	104
Tabelle 4-41: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung in NT-proBNP (Verringerung des NT-proBNP Wertes um ≥ 30 %) oder Erhaltung bzw. Erreichung eines NT-proBNP Wertes von < 300 ng/L)	104
Tabelle 4-42: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung oder Erhaltung der WHO-FK	104
Tabelle 4-43: Operationalisierung des Endpunkts Krankheitssymptomatik und Gesundheitszustand	106
Tabelle 4-44: Bewertung des Verzerrungspotenzials für Krankheitssymptomatik und Gesundheitszustand in RCT mit dem zu bewertenden Arzneimittel	108
Tabelle 4-45: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiopulmonale Symptome) – Responderanalyse	109
Tabelle 4-46: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiovaskuläre Symptome) – Responderanalyse	109
Tabelle 4-47: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (EQ-5D VAS) – Responderanalyse	110
Tabelle 4-48: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiopulmonale Symptome) – Veränderung gegenüber Baseline.....	111
Tabelle 4-49: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiovaskuläre Symptome) – Veränderung gegenüber Baseline.....	111
Tabelle 4-50: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (EQ-5D VAS) – Veränderung gegenüber Baseline..	111
Tabelle 4-51: Operationalisierung des Endpunkts Gesundheitsbezogene Lebensqualität.....	113
Tabelle 4-52: Bewertung des Verzerrungspotenzials für Krankheitssymptomatik und Gesundheitszustand in RCT mit dem zu bewertenden Arzneimittel	114
Tabelle 4-53: Ergebnisse für Verbesserung der Gesundheitsbezogenen Lebensqualität um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Körperliche Auswirkungen) – Responderanalyse	115
Tabelle 4-54: Ergebnisse für Verbesserung der Gesundheitsbezogenen Lebensqualität um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kognitive/Emotionale Auswirkungen) – Responderanalyse.....	115
Tabelle 4-55: Ergebnisse für Gesundheitsbezogene Lebensqualität aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Körperliche Auswirkungen) – Veränderung gegenüber Baseline	116

Tabelle 4-56: Ergebnisse für Gesundheitsbezogene Lebensqualität aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kognitive/Emotionale Auswirkungen) – Veränderung gegenüber Baseline.....	117
Tabelle 4-57: Operationalisierung des Endpunkts Unerwünschte Ereignisse Gesamtraten zu Woche 24.....	118
Tabelle 4-58: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse Gesamtraten in RCT mit dem zu bewertenden Arzneimittel	119
Tabelle 4-59: Ergebnisse für Unerwünschte Ereignisse Gesamtraten aus RCT mit dem zu bewertenden Arzneimittel	120
Tabelle 4-60: Operationalisierung des Endpunkts Unerwünschte Ereignisse (gegliedert nach SOC und PT).....	120
Tabelle 4-61: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse (gegliedert nach SOC und PT) in RCT mit dem zu bewertenden Arzneimittel.....	122
Tabelle 4-62: Ergebnisse für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel	122
Tabelle 4-63: Ergebnisse für den Endpunkt Schwerwiegende Unerwünschte Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel	125
Tabelle 4-64: Ergebnisse für den Endpunkt Schwere Unerwünschte Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel	125
Tabelle 4-65: Ergebnisse für den Endpunkt Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel	126
Tabelle 4-66: Operationalisierung des Endpunkts Unerwünschte Ereignisse von besonderem Interesse	127
Tabelle 4-67: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse von Interesse in RCT mit dem zu bewertenden Arzneimittel	128
Tabelle 4-68: Ergebnisse für Unerwünschte Ereignisse von Interesse aus RCT mit dem zu bewertenden Arzneimittel	129
Tabelle 4-69 Matrix der durchgeführten Subgruppenanalysen.....	131
Tabelle 4-70: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Verbesserung der 6MWD – Responderanalyse.....	135
Tabelle 4-71: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt 6MWD – Veränderung gegenüber Baseline (Teil 1) ..	135
Tabelle 4-72: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt 6MWD – Veränderung gegenüber Baseline (Teil 2) ..	136
Tabelle 4-73: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod.....	137
Tabelle 4-74: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala.	137
Tabelle 4-75: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (PAH-SYMPACT – Kardiopulmonale Symptome).....	138

Tabelle 4-76: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (PAH-SYMPACT – Kardiovaskuläre Symptome)	138
Tabelle 4-77: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (EQ-5D VAS).....	139
Tabelle 4-78: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Gesundheitsbezogene Lebensqualität (PAH-SYMPACT – Körperliche Auswirkungen)	140
Tabelle 4-79: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Gesundheitsbezogene Lebensqualität (PAH-SYMPACT – Kognitive/Emotionale Auswirkungen).....	140
Tabelle 4-80: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Unerwünschte Ereignisse Gesamtraten	141
Tabelle 4-81: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT).....	142
Tabelle 4-82: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt 6MWD Veränderung gegenüber Baseline zu Woche 24 aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)	144
Tabelle 4-83: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Krankheitssymptomatik und Gesundheitszustand (EQ-5D VAS) aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)	145
Tabelle 4-84: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) für SOC aus RCT mit dem zu bewertenden Arzneimittel (STELLAR).....	146
Tabelle 4-85: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) für PT aus RCT mit dem zu bewertenden Arzneimittel (STELLAR).....	148
Tabelle 4-86: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche	151
Tabelle 4-87: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	152
Tabelle 4-88: Operationalisierung von <Endpunkt xxx>.....	152
Tabelle 4-89: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche	153
Tabelle 4-90: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	153
Tabelle 4-91: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien	156
Tabelle 4-92: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien	156
Tabelle 4-93: Operationalisierung von <Endpunkt xxx>.....	157

Tabelle 4-94: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien	157
Tabelle 4-95: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen	159
Tabelle 4-96: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen.....	160
Tabelle 4-97: Übersicht über die zur Ableitung des Zusatznutzens herangezogenen Endpunkte der Studie STELLAR.....	163
Tabelle 4-98: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	168
Tabelle 4-99 (Anhang): Studiendesign und -methodik für Studie STELLAR.....	191
Tabelle 4-100 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie STELLAR	199

Abbildungsverzeichnis

	Seite
Abbildung 1: Domänen des PAH-SYMPACT-Fragebogens (21)	41
Abbildung 2: Mehrstufiges Vorgehen zur Durchführung von Subgruppenanalysen und methodisches Vorgehen zur Bewertung möglicher Effektmodifikationen (Eigene Darstellung)	52
Abbildung 3: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel	61
Abbildung 4: Schematische Darstellung des Studiendesigns - Studie STELLAR (Quelle: CSP)	71
Abbildung 5: Kaplan-Meier-Kurve für die Zeit bis Studienende (STELLAR)	72
Abbildung 6: Überlebenszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Gesamtüberleben (Studie STELLAR und weiterführende Daten aus der LTE-Studie SOTERIA).....	83
Abbildung 7: Überlebenszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Gesamtüberleben (RPSFT-Methode, Studie STELLAR und weiterführende Daten aus SOTERIA).....	85
Abbildung 8: Ereigniszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod der Studie STELLAR – Hauptanalyse	93
Abbildung 9: Ereigniszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod der Studie STELLAR – Sensitivitätsanalyse	97
Abbildung 10: Flow-Chart der Studie STELLAR	197

Abkürzungsverzeichnis

Abkürzung	Bedeutung
6MWD	6-minutes-walking-distance (6-Minuten-Gehstrecke)
AEOI	Adverse Events of Interest (Unerwünschte Ereignisse von Interesse)
AEOSI	Adverse Events of Special Interest (Unerwünschte Ereignisse von besonderem Interesse)
AESI	Adverse Event of Special Interest (Unerwünschtes Ereignis von besonderem Interesse)
AMIce	Arzneimittel-Informationssystem
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ANCOVA	Analysis of Covariance (Kovarianzanalyse)
ASE	Asymptotic Standard Error (asymptotischer Standardfehler)
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
CONSORT	Consolidated Standards of Reporting Trials
COVID-19	Coronavirus Disease 2019
CR10-Skala	10 Point Category Ratio Scale
CTCAE	Common Terminology Criteria for Adverse Events
CTD	Common Technical Document (Zulassungsdossier)
DBPC	Double-Blind Placebo-Controlled (doppelblind Placebo-kontrolliert)
dFDR	Double FDR
DIMDI	Deutsches Institut für Medizinische Dokumentation
EMA	European Medicines Agency (Europäische Arzneimittel-Agentur)
EMBASE	Excerpta Medica Database
EQ-5D	EuroQoL-5 Dimensions
ESC	European Society of Cardiology
EU-CTR	EU Clinical Trials Register
EudraCT	European Union Drug Regulating Authorities Clinical Trials
EuroQoL	European Quality of Life 5 Dimensions
FAS	Full Analysis Set
FDR	False Discovery Rate
FK	Funktionsklasse
FSH	Follikel-stimulierendes Hormon
G-BA	Gemeinsamer Bundesausschuss

Hb	Hämoglobin
HIV	Humanes Immundefizienz-Virus
hPAH	Heritable pulmonale arterielle Hypertonie
HR	Hazard Ratio
ICTRP	International Clinical Trials Registry Platform
IPAH	Idiopathische pulmonale arterielle Hypertonie
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IRT	Interactive Response Technology
ITT	Intention-To-Treat
kg	Kilogramm
KI	Konfidenzintervall
L	Liter
LTDB	Long-Term Double-Blind (langfristige doppelblinde Behandlungsphase)
LTE	Long-Term Extension
Max	Maximum
MedDRA	Medical Dictionary for Regulatory Activities
MEDLINE	MEDical Literature Analysis and Retrieval System onLINE
mg	Milligramm
MID	Minimal Important Difference
Min	Minimum
MMRM	Mixed effect Model Repeat Measurement
MTC	Mixed Treatment Comparison
N	Anzahl der Patient:innen
n	Patient:innen mit Ereignis
n.a.	Not Applicable
n.c.	not calculated (nicht berechnet)
n.s.	nicht signifikant (adjustierter p-Wert $\geq 0,05$)
NCT	National Clinical Trial
ng	Nanogramm
NT-proBNP	N-Terminal Pro-B-Type Natriuretic Peptide
OR	Odds Ratio
PAH	Pulmonale arterielle Hypertonie

PAH-SYMPACT	Pulmonary Arterial Hypertension – Symptoms and Impact
PT	Preferred Terms
PVR	Pulmonary Vascular Resistance
Q1	Erstes Quartil
Q3	Drittes Quartil
RCT	Randomized Controlled Trial (randomisierte kontrollierte Studie)
RHK	Rechtsherzkatheterisierung
RoW	Rest of the World (Rest der Welt)
RR	Relatives Risiko
s.c.	Subkutan
SD	Standard Deviation
SE	Standard Error (Standardfehler)
SGB	Sozialgesetzbuch
SMDs	Standardisierte Mittelwertdifferenzen
SMQs	Standardised MedDRA Queries
SOC	System Organ Class (Systemorganklasse)
STE	Surrogate Threshold Effect
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SUE	Schwerwiegendes UE
TREND	Transparent Reporting of Evaluations with Non-Randomized Design
TTCW	Time to Clinical Worsening (Zeit bis zur klinischen Verschlechterung oder Tod)
TTE	Time-To-Event
UE	Unerwünschtes Ereignis
ULN	Upper Limit of Normal (obere Normgrenze)
VAS	Visual Analog Scale (visuelle Analogskala)
VerfO	Verfahrensordnung
WHO	World Health Organization (Weltgesundheitsorganisation)
WHO-FK	WHO-Funktionsklasse
WHO-FC	WHO Functional Class (WHO-Funktionsklasse)
WU	Wood Unit (Wood-Einheit)
zVT	Zweckmäßige Vergleichstherapie

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 4.3.2)
- 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.

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

Das Ziel der vorliegenden Untersuchung ist die Ermittlung des Ausmaßes des medizinischen Nutzens und medizinischen Zusatznutzens von Sotatercept in Kombination mit anderen Therapien gegen pulmonale arterielle Hypertonie (PAH) im Vergleich zur zweckmäßigen Vergleichstherapie (zVT) für die Behandlung von PAH bei erwachsenen Patienten mit der Weltgesundheitsorganisation (World Health Organization, WHO)-Funktionsklasse (FK) II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Patientenpopulation

Im vorliegenden Anwendungsgebiet A handelt es sich um die Erstzulassung von Sotatercept in Kombination mit anderen Therapien gegen PAH für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Intervention

Für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit, beträgt die empfohlene Dosis von Sotatercept einmalig zu Beginn 0,3 mg/kg und danach 0,7 mg/kg alle drei Wochen als subkutane Injektion.

Vergleichstherapie

Der Gemeinsame Bundesausschuss (G-BA) legte im Rahmen der Beratungsanforderung 2022-B-349 die zVT zur Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit wie folgt fest (1):

„Patientenindividuell optimierte medikamentöse Therapie unter Berücksichtigung folgender Therapien:

- Endothelin-Rezeptor-Antagonisten (Ambrisentan, Bosentan, Macitentan)
- Phosphodiesterase-Typ-5-Inhibitoren (Sildenafil, Tadalafil)
- Prostazyklin-Analoga (Iloprost)
- Selektive Prostazyklin-Rezeptor-Agonisten (Selexipag)
- Stimulator der löslichen Guanylatzyklase (Riociguat)

MSD folgt grundsätzlich der Entscheidung des G-BA (1) sieht aber aufgrund des Patientenkollektivs im Anwendungsgebiet auch die parenteral zu verabreichenden Prostazykline als relevanten Bestandteil der zVT an. Die Patient:innen sollen abhängig von ihren Vortherapien und dem jeweiligen Gesundheitszustand patientenindividuell optimiert behandelt werden (1). Somit ist es plausibel, dass Patient:innen sich im Verlauf der Erkrankung

in ihrer WHO-FK verbessern können – dass für sie das individuell wirksame Therapieschema jedoch beibehalten wird.

Endpunkte

Die Bewertung erfolgt hinsichtlich der patientenrelevanten Endpunkte der Endpunktkategorien Mortalität, Morbidität, Gesundheitsbezogene Lebensqualität und Nebenwirkungen (siehe Abschnitt 4.3.1.3). Der Nutzen eines Arzneimittels ist nach § 3 Abs. 1 des 5. Kapitels der Verfahrensordnung (VerfO) des G-BA „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ (2). Der Nutzen einer Maßnahme soll sich dabei auf die Patient:innen beziehen und patientenrelevante Endpunkte beeinflussen, wobei als patientenrelevant in diesem Zusammenhang verstanden werden soll „wie eine Patientin oder ein Patient fühlt, ihre oder seine Funktionen und Aktivitäten wahrnehmen kann oder ob sie oder er überlebt“ (3).

Studientypen

Die Bewertung wird nach Möglichkeit auf Grundlage von randomisierten kontrollierten Studien (Randomized Controlled Trials, RCT) zur oben genannten Fragestellung vorgenommen.

Datenquellen

Neben der Suche in internen Informationssystemen und Datenbanken wurde eine systematische Literaturrecherche innerhalb externer Datenbanken durchgeführt. Die systematische bibliografische Literaturrecherche erfolgte in den Datenbanken MEDical Literature Analysis and Retrieval System onLINE (MEDLINE), Excerpta Medica Database (EMBASE) und Cochrane Central Register of Controlled Trials. Zudem wurde eine systematische Suche in den Studienregistern ClinicalTrials.gov, EU Clinical Trials Register (EU-CTR) und im International Clinical Trials Registry Platform (ICTRP)-Suchportal durchgeführt. Des Weiteren wurden die Internetseite des G-BA, das Suchportal der Europäischen Arzneimittel-Agentur (European Medicines Agency, EMA) sowie das Arzneimittel-Informationssystem (AMIce) nach bewertungsrelevanten Dokumenten durchsucht.

Mit der Suche in den herangezogenen Datenquellen konnte die Zulassungsstudie STELLAR identifiziert werden. Diese dient als Grundlage für den Nachweis des medizinischen Nutzens und medizinischen Zusatznutzens im vorliegenden Anwendungsgebiet.

Ein-/Ausschlusskriterien für Studien

Tabelle 4-1: Übersicht der Ein- und Ausschlusskriterien für die Suche nach RCT mit dem zu bewertenden Arzneimittel zum medizinischen Nutzen und Zusatznutzen

Einschlusskriterien		Ausschlusskriterien
E1 Patientenpopulation	Erwachsene Patient:innen mit pulmonaler arterieller Hypertonie (PAH) der WHO-FK II bis III	Nicht E1

Einschlusskriterien		Ausschlusskriterien
E2 Prüfintervention	Sotatercept, in Kombination mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, Dosierung gemäß Fachinformation, Dosierung gemäß Fachinformation	Nicht E2
E3 Vergleichsintervention	Eine patientenindividuell optimierte medikamentöse Therapie unter Berücksichtigung der Vortherapien und des Gesundheitszustandes, unter Berücksichtigung folgender Therapien: <ul style="list-style-type: none"> • Endothelin-Rezeptor-Antagonisten (Ambrisentan, Bosentan, Macitentan) • Phosphodiesterase-Typ-5-Inhibitoren (Sildenafil, Tadalafil) • Prostazyklin-Analoga (Iloprost) • Selektive Prostazyklin-Rezeptor-Agonisten (Selexipag) • Stimulator der löslichen Guanylatzyklase (Riociguat) 	Nicht E3
E4 Patientenrelevante Endpunkte	Erhebung von mindestens einem patientenrelevanten Endpunkt aus den folgenden Endpunktkategorien: <ul style="list-style-type: none"> • Mortalität • Morbidität • Gesundheitsbezogene Lebensqualität • Nebenwirkungen 	Nicht E4
E5 Studiendauer	Studien mit einer Dauer von mindestens 24 Wochen	Nicht E5
E6 Studientyp	RCT	Nicht E6
E7 Publikationstyp	Vollpublikation, Ergebnisbericht aus einem Studienregister bzw. Studienbericht verfügbar	Nicht E7 ^a
<p>a: Nicht E7 sowie jeglicher Publikationstyp, in welchem keine zusätzlichen Informationen (zu bereits identifizierten Informationsquellen) dargestellt werden. Dies beinhaltet auch in der bibliografischen Literaturrecherche identifizierte Studienregistereinträge, welche ebenfalls in der Studienregistersuche identifiziert werden</p> <p>PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation); WHO-FK: WHO-Funktionsklasse; RCT: Randomisierte kontrollierte Studie</p>		

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

Zur Bewertung der Aussagekraft der eingeschlossenen Studien werden die Verzerrungsaspekte der Ergebnisse endpunktübergreifend und endpunktspezifisch eingeschätzt. Die Einschätzung erfolgt anhand der zur Verfügung stehenden Quellen wie Studienberichte/-protokolle, aus denen alle relevanten Angaben in den Bewertungsbogen zur Einschätzung von Verzerrungsaspekten im Anhang 4-F extrahiert werden. Das Verzerrungspotenzial wird

entweder als „niedrig“ oder „hoch“ eingestuft. Von einem niedrigen Verzerrungspotenzial kann ausgegangen werden, wenn eine relevante Verzerrung der Ergebnisse mit großer Wahrscheinlichkeit ausgeschlossen werden kann. Eine relevante Verzerrung liegt dann vor, wenn sich die Ergebnisse in ihrer Grundaussage verändern würden.

Die Verzerrungsaspekte auf Studienebene und Endpunktebene werden in den Bewertungsbögen in Anhang 4-F ausführlich dargestellt.

Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Tabelle 4-2: Übersicht über die zur Ableitung des Zusatznutzens herangezogenen Endpunkte der Studie STELLAR

Studie: STELLAR ^a	Sotatercept			Placebo			Sotatercept vs. Placebo		Zusatznutzen
	N	Mediane Patienten mit Ereignis n (%)	Mediane Ereigniszeit [95 %-KI]	N	Mediane Patienten mit Ereignis n (%)	Mediane Ereigniszeit [95 %-KI]	Effekt-schätzer [95 %-KI]	p-Wert	
Mortalität									
Mortalität zu Woche 24	163 ^b	0 (0,0)	n.a.	160 ^b	6 (3,8)	n.a.	Peto-OR ^c : 0,13 [0,03; 0,64]	0,013 ^d	Hinweis auf einen erheblichen Zusatznutzen
Morbidität									
Verbesserung der 6MWD	163 ^e	72 (44,2)	n.a.	159 ^e	29 (18,2)	n.a.	RR ^c : 2,40 [1,66; 3,48]	< 0,001 ^d	Hinweis auf einen erheblichen Zusatznutzen
Zeit bis zur klinischen Verschlechterung	163 ^b	11 (6,7)	Nicht erreicht [-; -] ^g	160 ^b	42 (26,3)	78,00 [-; -] ^g	HR ^g : 0,23 [0,12; 0,44]	< 0,001 ^{g,h}	Hinweis auf einen erheblichen Zusatznutzen
Verbesserung der Dyspnoe gemäß Borg CR10-Skala	160 ^f	38 (23,8)	n.a.	159 ^f	37 (23,3)	n.a.	RR ^c : 1,02 [0,69; 1,51]	0,918 ^d	Zusatznutzen nicht belegt
Krankheitssymptomatik und Gesundheitszustand									
Verbesserung der PAH-SYMPACT Domäne: kardiopulmonale Symptome	115 ^f	47 (40,9)	n.a.	117 ^f	35 (29,9)	n.a.	RR ^c : 1,35 [0,95; 1,93]	0,095 ^d	Zusatznutzen nicht belegt
Verbesserung der PAH-SYMPACT Domäne: kardiovaskuläre Symptome	115 ^f	49 (42,6)	n.a.	117 ^f	34 (29,1)	n.a.	RR ^c : 1,48 [1,04; 2,11]	0,030 ^d	Hinweis auf einen geringen Zusatznutzen
Verbesserung des EQ-5D VAS	124 ^f	29 (23,4)	n.a.	126 ^f	20 (15,9)	n.a.	RR ^c : 1,49 [0,89; 2,49]	0,131 ^d	Zusatznutzen nicht belegt
Gesundheitsbezogene Lebensqualität									
Verbesserung der PAH-SYMPACT Domäne: Auswirkungen auf die körperliche Funktion	117 ^f	39 (33,3)	n.a.	123 ^f	31 (25,2)	n.a.	RR ^c : 1,31 [0,87; 1,96]	0,193 ^d	Zusatznutzen nicht belegt

Studie: STELLAR ^a	Sotatercept		Placebo		Sotatercept vs. Placebo	Zusatznutzen
Verbesserung der PAH-SYMPACT Domäne: kognitive/emotionale Auswirkungen	117 ^f	30 (25,6) n.a.	123 ^f	30 (24,4) n.a.	RR ^c : 0,866 ^d 1,04 [0,67; 1,60]	
Nebenwirkungen						
Unerwünschte Ereignisse Gesamtraten						
Schwerwiegende unerwünschte Ereignisse	163 ⁱ	40 (24,54) n.a.	160 ⁱ	47 (29,38) n.a.	RR ^j : 0,3281 ^k 0,84 [0,58; 1,20]	Zusatznutzen nicht belegt
Schwere unerwünschte Ereignisse	163 ⁱ	24 (14,72) n.a.	160 ⁱ	33 (20,63) n.a.	RR ^j : 0,1649 ^k 0,71 [0,44; 1,15]	
Therapieabbruch wegen unerwünschter Ereignisse	163 ⁱ	6 (3,68) n.a.	160 ⁱ	11 (6,88) n.a.	RR ^j : 0,1994 ^k 0,54 [0,20; 1,41]	
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)</p> <p>d: Zweiseitiger p-Wert (Wald-Test)</p> <p>e: Anzahl der Patient:innen: Full Analysis Set Population. Patient:innen mit fehlender Erhebung zu Woche 24 und nicht aufgrund von COVID-19 werden als Non-Responder gewertet. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden ausgeschlossen</p> <p>f: Anzahl Patient:innen:</p> <ul style="list-style-type: none"> mit Erhebung sowohl zu Baseline als auch zu Woche 24 sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt. <p>Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.</p> <p>g: Basierend auf einem Cox Regressionsmodell mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)</p> <p>h: Zweiseitiger p-Wert (Wald-Test; Score-Test im Fall von null Ereignissen in einem der Studienarme)</p> <p>i: Anzahl der Patient:innen: Safety Set Population</p> <p>j: 2x2 Kontingenztabelle mit Konfidenzintervall nach Wald. Bei keinem:keiner Patient:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztabelle, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppe kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet</p> <p>k: Basierend auf einen Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsmarmen aufweisen, wird n.a. berichtet</p> <p>KI: Konfidenzintervall; n.a.: nicht anwendbar; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization; COVID-19: Coronavirus Disease of 2019; 6MWD: 6-Minuten Gehstrecke; G-BA: Gemeinsamer Bundesausschuss; PAH-SYMPACT: pulmonale arterielle Hypertonie-Symptome und Auswirkungen; EQ-5D: European Quality of Life 5 Dimensions; VAS: Visuelle Analogskala</p>						

Mortalität

Gesamtüberleben

Für den Endpunkt Mortalität zu Woche 24 zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (Peto-Odds Ratio [OR] [95 %- Konfidenzintervall (KI)]: 0,13 [0,03; 0,64]; $p = 0,013$), dies ist eine erhebliche Reduktion des Mortalitätsrisikos.

Für den Endpunkt Mortalität zu Woche 24 ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept

gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Morbidität

Verbesserung der 6MWD

Für den Endpunkt Verbesserung der 6-Minuten-Gehstrecke (6-minutes-walking-distance, 6MWD) zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (Relatives Risiko (RR) [95 %-KI]: 2,40 [1,66; 3,48]; $p = < 0,001$). 44,2 % der Patient:innen im Sotatercept-Arm konnten ihre innerhalb von 6 Minuten zurückgelegte Gehdistanz um mindestens 40 Meter gegenüber ihrem Baseline-Wert verbessern, im Placebo-Arm lediglich 18,2 %. Es ist anzumerken, dass diese bisher nicht erreichte, erhebliche Verbesserung der körperlichen Belastbarkeit unter Verwendung einer Minimal Important Difference (MID) erreicht wird, welche die in der Fachliteratur als klinisch relevant erachtete MID von 33 Metern weit überschreitet. Darüber hinaus ist festzuhalten, dass es sich in der Studie STELLAR durch das Add-on Design um bereits therapierte Patient:innen handelt.

Für den Endpunkt Verbesserung der 6MWD ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Zeit bis zur klinischen Verschlechterung oder Tod

Für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (Hazard Ratio HR [95 %-KI]: 0,23 [0,12; 0,44]; $p = < 0,001$). Das Risiko zu versterben oder eine klinische Verschlechterung zu erfahren, beträgt etwa ein Viertel für die Patient:innen im Sotatercept-Arm im Vergleich zum Placebo-Arm.

Für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Verbesserung der Dyspnoe gemäß Borg CR10-Skala

Für den Endpunkt Verbesserung der Dyspnoe gemäß Borg 10 Point Category Ratio Scale (CR10-Skala) (Verbesserung der Dyspnoe gemäß Borg CR10-Skala um 15 % der Skalenspannweite) zeigt sich kein Unterschied zwischen Sotatercept und Placebo.

Für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala ist der **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Krankheitssymptomatik und Gesundheitszustand

Für die Krankheitssymptomatik anhand des Symptom-Items kardiopulmonale Symptome des Pulmonary Arterial Hypertension – Symptoms and Impact (PAH-SYMPACT) und Gesundheitszustand (anhand der EuroQoL-5 Dimensions [EQ-5D] VAS) sind positive Effekte der Behandlung mit Sotatercept zu beobachten, die in den Responderanalysen jedoch keine statistische Signifikanz erreichen. Für die Krankheitssymptomatik anhand des Symptom-Items kardiovaskuläre Symptome des PAH-SYMPACT zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (RR [95 %-KI]: 1,48 [1,04; 2,11]; $p = 0,030$).

Für den Endpunkt Krankheitssymptomatik und Gesundheitszustand ist das Verzerrungspotenzial als niedrig zu bewerten. In der Gesamtschau liegen positive Effekte und ein Zusatznutzen hinsichtlich der kardiovaskulären Symptome vor.

Somit ist für die patientenberichtete Krankheitssymptomatik und den Gesundheitszustand anhand des Symptom-Items kardiopulmonale Symptome des PAH-SYMPACT und Gesundheitszustand (anhand der EQ-5D Visuelle Analogskala (Visual Analog Scale, VAS) ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Für den Endpunkt Krankheitssymptomatik anhand des Symptom-Items kardiovaskuläre Symptome des PAH-SYMPACT ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen geringen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Gesundheitsbezogene Lebensqualität

Gesundheitsbezogene Lebensqualität

In Bezug auf die Auswirkungen auf die körperliche Funktion in Alltagssituationen berichten Patient:innen unter Sotatercept häufiger eine Verbesserung als im Placebo-Arm. Diese erreicht jedoch keine statistische Signifikanz. Für das Impact-Item „Kognitive/Emotionale Auswirkungen“ lässt sich kein Unterschied zwischen der Behandlung mit Sotatercept und Placebo feststellen.

Für den Endpunkt Gesundheitsbezogene Lebensqualität (anhand der Impact-Items des PAH-SYMPACT) ist das Verzerrungspotenzial als niedrig zu bewerten und in der Gesamtschau liegt kein fazitrelevanter Zusatznutzen vor. Somit ist für die Gesundheitsbezogene Lebensqualität ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Nebenwirkungen

Unerwünschte Ereignisse Gesamtraten

In der Studie STELLAR zeigt sich bei den Unerwünschten Ereignissen Gesamtraten kein Unterschied zwischen Sotatercept und Placebo.

Für den Endpunkt Unerwünschte Ereignisse Gesamtraten ist das Verzerrungspotenzial als niedrig zu bewerten. Somit ist für den Endpunkt Unerwünschte Ereignisse Gesamtraten ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Die Ableitung des Zusatznutzens für Sotatercept, in Kombination mit anderen Therapien gegen PAH, für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit, basiert auf den Ergebnissen der randomisierten, doppelblinden und Placebo-kontrollierten Studie STELLAR. Diese Studie umfasst Patient:innen der WHO-FK II und III, behandelt mit stabiler, patientenindividuell optimierten PAH-Hintergrundtherapie als Mono- (4,0 %), Zweifachkombinations- (34,7 %) oder Dreifachkombinationstherapie (61,3 %). Etwa 40 % der Studienteilnehmer:innen erhielten aufgrund des bereits fortgeschrittenen Krankheitsstadiums eine parenterale Prostazyklin-Therapie. Bei der Behandlung dieser Patient:innen wurden bereits alle weiteren verfügbaren Wirkansätze ausgeschöpft.

Die PAH ist trotz der aktuell verfügbaren Behandlungsoptionen nach wie vor eine rasch fortschreitende, lebenszeitverkürzende und letztlich tödlich verlaufende Erkrankung, die für betroffene Patient:innen mit einem hohen Leidensdruck einhergeht. Im Unterschied zu den aktuell verfügbaren Substanzklassen steht mit Sotatercept nun erstmals ein Wirkstoff zur Verfügung, der über die Behandlung der Symptome der PAH hinaus auch kausal in die Pathophysiologie der PAH eingreift. Dabei steuert Sotatercept einen neuen, bisher nicht verfolgten, 4. Signalweg an und ermöglicht sogar bei Patient:innen, die bisher verfügbare Therapieoptionen ausgeschöpft haben, umfangreiche klinisch relevante und für die Patient:innen unmittelbar spürbare Verbesserungen.

Durch eine Behandlung mit Sotatercept wird eine nachhaltige und bisher nicht erreichte große Verbesserung des therapielevanten Nutzens gegenüber der zVT erzielt. Hervorzuheben ist hier einerseits die erhebliche Verbesserung der körperlichen Belastbarkeit der Patient:innen, welche sich unmittelbar in alltäglichen Funktionen und Aktivitäten widerspiegelt und andererseits die nachhaltige, erhebliche Reduktion des Mortalitätsrisikos. Mit Sotatercept kann nicht nur eine Verbesserung, sondern auch eine Verzögerung dieser progredienten chronischen Erkrankung erreicht werden. Die Patient:innen erleiden erheblich seltener eine klinische Verschlechterung.

Zusammenfassend handelt es sich bei Sotatercept, in Kombination mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, für die

Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit, um eine hochwirksame und gut verträgliche Therapieoption. In der Gesamtschau ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept im Vergleich zur zVT.

4.2 Methodik

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).

Die Benennung der Vergleichstherapie in Modul 4 muss zur Auswahl der zweckmäßigen Vergleichstherapie im zugehörigen Modul 3 konsistent sein.

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.

Das Ziel der vorliegenden Untersuchung ist die Ermittlung des Ausmaßes des medizinischen Nutzens und medizinischen Zusatznutzens von Sotatercept in Kombination mit anderen Therapien gegen PAH im Vergleich zur zweckmäßigen Vergleichstherapie (zVT) für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Patientenpopulation

Im vorliegenden Anwendungsgebiet A handelt es sich um die Erstzulassung von Sotatercept in Kombination mit anderen Therapien gegen PAH für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Intervention

Für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit beträgt die empfohlene Dosis von Sotatercept einmalig zu Beginn 0,3 mg/kg und danach 0,7 mg/kg alle drei Wochen als subkutane Injektion.

Vergleichstherapie

Der Gemeinsame Bundesausschuss (G-BA) legte im Rahmen der Beratungsanforderung 2022-B-349 die zVT zur Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit wie folgt fest (1):

„Patientenindividuell optimierte medikamentöse Therapie unter Berücksichtigung folgender Therapien:

- Endothelin-Rezeptor-Antagonisten (Ambrisentan, Bosentan, Macitentan)
- Phosphodiesterase-Typ-5-Inhibitoren (Sildenafil, Tadalafil)
- Prostazyklin-Analoga (Iloprost)
- Selektive Prostazyklin-Rezeptor-Agonisten (Selexipag)
- Stimulator der löslichen Guanylatzyklase (Riociguat)

MSD folgt grundsätzlich der Entscheidung des G-BA (1) sieht aber aufgrund des Patientenkollektivs im Anwendungsgebiet auch die parenteral zu verabreichenden Prostazykline als relevanten Bestandteil der zVT an. Die Patient:innen sollen abhängig von ihren Vortherapien und dem jeweiligen Gesundheitszustand patientenindividuell optimiert behandelt werden (1). Somit ist es plausibel, dass Patient:innen sich im Verlauf der Erkrankung in ihrer WHO-FK verbessern können – dass für sie das individuell wirksame Therapieschema jedoch beibehalten wird.

Endpunkte

Die Bewertung erfolgt hinsichtlich der patientenrelevanten Endpunkte der Endpunktkategorien Mortalität, Morbidität, Gesundheitsbezogene Lebensqualität und Nebenwirkungen (siehe Abschnitt 4.3.1.3). Der Nutzen eines Arzneimittels ist nach § 3 Abs. 1 des 5. Kapitels der VerfO des G-BA „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ (2). Der Nutzen einer Maßnahme soll sich dabei auf die Patient:innen beziehen und patientenrelevante Endpunkte beeinflussen, wobei als patientenrelevant in diesem Zusammenhang verstanden werden soll „wie eine Patientin oder ein Patient fühlt, ihre oder seine Funktionen und Aktivitäten wahrnehmen kann oder ob sie oder er überlebt“ (3).

Studientypen

Die Bewertung wird nach Möglichkeit auf Grundlage von RCT zur oben genannten Fragestellung vorgenommen.

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/ einer Studienergebnisdatenbank erfolgen, während ein Kongressabstrakt 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. Erstellen Sie dabei für unterschiedliche Themen der Recherche (z. B. unterschiedliche Fragestellungen) jeweils eine separate Übersicht.

Begründung der Wahl der Selektionskriterien

Das untersuchte Anwendungsgebiet A umfasst die bewertungsrelevante Patientenpopulation erwachsene Patienten mit PAH der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Die Prüflintervention ist Sotatercept (WINREVAIR®), die gemäß den Angaben der Fachinformation zulassungskonform in einer Dosierung von einmalig zu Beginn 0,3 mg/kg und danach 0,7 mg/kg alle drei Wochen als subkutane Injektion angewendet wird (4).

Der G-BA legte im Rahmen der Beratungsanforderung 2022-B-349 die zVT zur Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit wie folgt fest (1):

Patientenindividuell optimierte medikamentöse Therapie unter Berücksichtigung folgender Therapien:

- Endothelin-Rezeptor-Antagonisten (Ambrisentan, Bosentan, Macitentan)
- Phosphodiesterase-Typ-5-Inhibitoren (Sildenafil, Tadalafil)
- Prostazyklin-Analoga (Iloprost)
- Selektive Prostazyklin-Rezeptor-Agonisten (Selexipag)
- Stimulator der löslichen Guanylatzyklase (Riociguat)

Die Bewertung des Zusatznutzens von Sotatercept gegenüber der zVT erfolgt anhand der patientenrelevanten Endpunkte aus den Endpunktkategorien Mortalität, Morbidität,

Gesundheitsbezogene Lebensqualität und Nebenwirkungen gemäß § 2 Abs. 3 der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) (siehe auch Abschnitt 4.2.5.2.2).

Bei der PAH handelt es sich um eine chronische Erkrankung. Der G-BA hält hierzu im Rahmen seiner Beratungsanforderung 2022-B-349 fest, dass bei chronischen Erkrankungen grundsätzlich eine Studiendauer von mindestens 24 Wochen als notwendig erachtet wird (1). Dies ist auch im Einklang mit der Leitlinie der EMA zur klinischen Erprobung medizinischer Produkte in der PAH (5).

Um Studien hochgradiger Evidenz zu identifizieren, wird eine Suche nach dem Studientyp RCT mit dem zu bewertenden Arzneimittel durchgeführt.

Bezüglich des Publikationstyps werden Studien berücksichtigt, deren Bewertung auf Grundlage eines Studienberichts, einer Vollpublikation oder eines ausführlichen Ergebnisberichts aus einem Studienregister erfolgen kann (Tabelle 4-3).

Tabelle 4-3: Übersicht der Ein- und Ausschlusskriterien für die Suche nach RCT mit dem zu bewertenden Arzneimittel zum medizinischen Nutzen und Zusatznutzen

Einschlusskriterien		Ausschlusskriterien
E1 Patientenpopulation	Erwachsene Patient:innen mit pulmonaler arterieller Hypertonie (PAH) der WHO-FK II bis III	Nicht E1
E2 Prüflintervention	Sotatercept, in Kombination mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, Dosierung gemäß Fachinformation	Nicht E2
E3 Vergleichsintervention	Eine patientenindividuell optimierte medikamentöse Therapie unter Berücksichtigung der Vortherapien und des Gesundheitszustandes, unter Berücksichtigung folgender Therapien: <ul style="list-style-type: none"> • Endothelin-Rezeptor-Antagonisten (Ambrisentan, Bosentan, Macitentan) • Phosphodiesterase-Typ-5-Inhibitoren (Sildenafil, Tadalafil) • Prostazyklin-Analoga (Iloprost) • Selektive Prostazyklin-Rezeptor-Agonisten (Selexipag) • Stimulator der löslichen Guanylatzyklase (Riociguat) 	Nicht E3

Einschlusskriterien		Ausschlusskriterien
E4 Patientenrelevante Endpunkte	Erhebung von mindestens einem patientenrelevanten Endpunkt aus den folgenden Endpunktkategorien: <ul style="list-style-type: none"> • Mortalität • Morbidität • Gesundheitsbezogene Lebensqualität • Nebenwirkungen 	Nicht E4
E5 Studiendauer	Studien mit einer Dauer von mindestens 24 Wochen	Nicht E5
E6 Studientyp	RCT	Nicht E6
E7 Publikationstyp	Vollpublikation, Ergebnisbericht aus einem Studienregister bzw. Studienbericht verfügbar	Nicht E7 ^a

a: Nicht E7 sowie jeglicher Publikationstyp, in welchem keine zusätzlichen Informationen (zu bereits identifizierten Informationsquellen) dargestellt werden. Dies beinhaltet auch in der bibliografischen Literaturrecherche identifizierte Studienregistereinträge, welche ebenfalls in der Studienregistersuche identifiziert werden
PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation);
WHO-FK: WHO-Funktionsklasse; RCT: RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)

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 (inklusive „in-process & other non-indexed citations“) 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.

Die systematische bibliografische Literaturrecherche nach relevanten Studien erfolgt in den Datenbanken

- EMBASE
- MEDLINE
- Cochrane Central Register of Controlled Trials.

Die für die jeweilige Datenbank adaptierten Suchstrategien sind in Blöcken getrennt nach Indikation, Intervention und Studientyp aufgebaut. Für die Einschränkung nach Studientyp (RCT) wird für die Datenbanken EMBASE und MEDLINE der Filter nach Wong zur Optimierung von Genauigkeit und Sensitivität verwendet (6). Die Suchstrategie in der Cochrane-Datenbank wird hinsichtlich des Studientyps nicht eingeschränkt. Es werden keine Jahreseinschränkungen und Einschränkungen hinsichtlich der Sprache vorgenommen.

Für die Suchen in den Datenbanken wird die Plattform Ovid verwendet.

Das Ergebnis der systematischen bibliografischen Literaturrecherche ist in Abschnitt 4.3.1.1.2 beschrieben.

Die Suchstrategien und die Anzahl der Treffer sind im Anhang 4-A1 dokumentiert.

4.2.3.3 Suche in Studienregistern/Studienergebnisdatenbanken

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

Eine Suche in Studienregistern/ Studienergebnisdatenbanken 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 sowie Studienergebnisdatenbanken 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/ Studienergebnisdatenbanken [clinicaltrials.gov](http://www.clinicaltrials.gov) (www.clinicaltrials.gov), EU Clinical Trials Register (EU-CTR, www.clinicaltrialsregister.eu), International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal), Suchportal der WHO, Clinical Data Suchportal der European Medicines Agency (<https://clinicaldata.ema.europa.eu>) sowie dem Arzneimittel-Informationssystem (AMIS, <https://www.pharmnet-bund.de/dynamic/de/arsneimittel-informationssystem/index.html>) durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Studienregistern / Studienergebnisdatenbanken (z. B. krankheitsspezifische Studienregister oder Studienregister einzelner pharmazeutischer Unternehmen) durchgeführt werden. Die Suche in Studienregistern/ Studienergebnisdatenbanken anderer pharmazeutischer Unternehmer ist insbesondere bei indirekten Vergleichen sinnvoll, wenn Studien zu anderen Arzneimitteln identifiziert werden müssen.

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

Für Clinical Data (Suchportal der European Medicines Agency) und das Arzneimittel-Informationssystem (AMIS) genügt hingegen die Suche nach Einträgen mit Ergebnisberichten zu Studien, die bereits anderweitig (z.B. über die bibliografische Literaturrecherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Studienregistern/ Studienergebnisdatenbanken 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.

Die Suche nach abgeschlossenen, abgebrochenen und laufenden Studien erfolgt in den folgenden Studienregistern mit einer für das jeweilige Studienregister adaptierten Suchstrategie:

- ClinicalTrials.gov
- EU-CTR
- ICTRP

Es werden keine Jahreseinschränkungen und Einschränkungen hinsichtlich der Sprache vorgenommen.

Die Ergebnisse der Suche in den Studienregistern sind in Abschnitt 4.3.1.1.2 beschrieben.

Die Suchstrategie und die Anzahl der Treffer sind im Anhang 4-B1 dokumentiert.

Für die in der bibliografischen Literaturrecherche oder Studienregistersuche identifizierten, im vorliegenden Anwendungsgebiet relevanten Studien erfolgt zusätzlich eine Suche nach Ergebnisberichten über das Suchportal der EMA und in der Datenbank des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM), dem AMIce.

4.2.3.4 Suche auf der Internetseite des G-BA

Die Internetseite des G-BA ist grundsätzlich zu durchsuchen, um sicherzustellen, dass alle vorliegenden Daten zu Studienmethodik und –ergebnissen von relevanten Studien in die Bewertung einfließen.

Auf der Internetseite des G-BA werden Dokumente zur frühen Nutzenbewertung nach §35a SGB V veröffentlicht. Diese enthalten teilweise anderweitig nicht veröffentlichte Daten zu Studienmethodik und –ergebnissen¹. Solche Daten sind dabei insbesondere in den Modulen 4 der Dossiers pharmazeutischer Unternehmer, in IQWiG-Nutzenbewertungen sowie dem Beschluss des G-BA einschließlich der Tragenden Gründe und der Zusammenfassenden Dokumentation zu erwarten.

Die Suche auf der Internetseite des G-BA 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

¹ Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörn H, Moritz G, Staack K, Wieseler B. Information on new drugs at market entry: retrospective analysis of health technology assessment reports, journal publications, and registry reports. *BMJ* 2015;350:h796

weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche auf der G-BA Internetseite immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird. Die Suche ist dann sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie durchzuführen. Es genügt die Suche nach Einträgen zu Studien, die bereits anderweitig (z.B. über die bibliografische Literaturrecherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

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

Beschreiben Sie nachfolgend das Vorgehen für die Suche. Benennen Sie die Wirkstoffe und die auf der Internetseite des G-BA genannten zugehörigen Vorgangsnummern, zu denen Sie eine Suche durchgeführt haben.

Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Für die in der bibliografischen Literaturrecherche oder Studienregistersuche identifizierten, im vorliegenden Anwendungsgebiet relevanten Studien erfolgt zusätzlich eine Suche nach Studienergebnissen oder Angaben zur Studienmethodik auf der Internetseite des G-BA. Als Suchbegriffe werden, sofern verfügbar, die European Union Drug Regulating Authorities Clinical Trials (EudraCT)-Nummer, National Clinical Trial (NCT)-Nummer, die interne Nummer des Prüfplans sowie weitere bekannte bzw. in den entsprechenden Registern/Publicationen genannten Studienbezeichnungen verwendet.

Die identifizierten Bewertungsverfahren wurden auf relevante Informationen in den Modulen 4, in den Nutzenbewertungen des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und in den tragenden Gründen zum Beschluss des G-BA durchsucht. Relevante Studien werden im Abschnitt 4.3.1.1.4 aufgeführt.

4.2.3.5 Selektion relevanter Studien

Beschreiben Sie das Vorgehen bei der Selektion relevanter Studien aus dem Ergebnis der in den Abschnitten 4.2.3.2, 4.2.3.3 und 4.2.3.4 beschriebenen Rechenschritte. Begründen Sie das Vorgehen, falls die Selektion nicht von zwei Personen unabhängig voneinander durchgeführt wurde.

Selektion relevanter Publikationen aus den Ergebnissen der bibliografischen Literaturrecherche

Zwei Reviewer bewerten unabhängig voneinander, ob die identifizierten Publikationen die in Abschnitt 4.2.2 gelisteten Ein- und Ausschlusskriterien erfüllen und somit für die vorliegende Nutzenbewertung relevant sind. Die Bewertung erfolgt in zwei Schritten: [1] Sichtung des Titels und – sofern vorhanden – des Abstracts aller identifizierten Publikationen, [2] Sichtung des Volltextes der nach Selektionsschritt 1 verbleibenden Publikationen. Diskrepanzen zwischen den Reviewern werden durch Diskussion aufgelöst, und wenn nötig wird ein dritter Reviewer in die Diskussion miteinbezogen. Der Selektionsprozess wird elektronisch

dokumentiert. Die im Volltext gesichteten und ausgeschlossenen Publikationen aus der bibliografischen Literaturrecherche sind mit Ausschlussgrund im Anhang 4-C1 dokumentiert.

Selektion relevanter Studien aus den Ergebnissen der Recherche in Studienregistern/Studienergebnisdatenbanken

Die aus den Recherchen in den Studienregistern/Studienergebnisdatenbanken identifizierten Studien werden ebenfalls von zwei unabhängigen Reviewern anhand der zuvor festgelegten Kriterien (siehe Abschnitt 4.2.2) auf ihre Relevanz hin bewertet. Diskrepanzen zwischen den beiden Reviewern werden durch Diskussion aufgelöst, und wenn nötig wird ein dritter Reviewer in die Diskussion miteinbezogen. Die als relevant erachteten Studien werden dem Studienpool zugeführt. Der Selektionsprozess wird elektronisch dokumentiert. Ausgeschlossene Studien aus der Studienregistersuche werden mit Ausschlussgrund im Anhang 4-D1 dokumentiert.

Selektion relevanter Dokumente aus den Ergebnissen der Suche auf der Internetseite des G-BA

Die aus den Recherchen auf der Internetseite des G-BA identifizierten Dokumente werden ebenfalls von zwei unabhängigen Reviewern auf ihre Relevanz hin bewertet. Diskrepanzen zwischen den beiden Reviewern werden durch Diskussion gelöst, und wenn nötig wird ein dritter Reviewer in die Diskussion miteinbezogen. Die als relevant erachteten Studiendokumente werden zusätzlich als Informationsquelle hinsichtlich Studienergebnissen oder Angaben zur Studienmethodik berücksichtigt.

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.

Zur Bewertung der Aussagekraft der eingeschlossenen Studien werden die Verzerrungsaspekte der Ergebnisse endpunktübergreifend und endpunktspezifisch eingeschätzt. Die Einschätzung erfolgt anhand der zur Verfügung stehenden Quellen wie Studienberichte/-protokolle, aus denen alle relevanten Angaben in den Bewertungsbogen zur Einschätzung von Verzerrungsaspekten im Anhang 4-F extrahiert werden. Das Verzerrungspotenzial wird entweder als „niedrig“ oder „hoch“ eingestuft. Von einem niedrigen Verzerrungspotenzial kann ausgegangen werden, wenn eine relevante Verzerrung der Ergebnisse mit großer Wahrscheinlichkeit ausgeschlossen werden kann. Eine relevante Verzerrung liegt dann vor, wenn sich die Ergebnisse in ihrer Grundaussage verändern würden.

Die Verzerrungsaspekte auf Studienebene und Endpunktebene werden in den Bewertungsbögen in Anhang 4-F ausführlich dargestellt.

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)². Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-³ bzw. STROBE-Statements⁴ 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.

Das Design und die Methodik der für die Nutzenbewertung relevanten Studien werden mithilfe des Consolidated Standards of Reporting Trials (CONSORT)-Statements (Item 2b bis 14) beschrieben und im Anhang 4-E dargestellt. Für die Erstellung des CONSORT-Statements wird der entsprechende Studienbericht herangezogen. Der Patientenfluss wird gemäß CONSORT-Flow-Chart 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

² Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c332.

³ 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.

⁴ 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.

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.

Sofern zur Berechnung von Ergebnissen von Standardverfahren und –software abgewichen wird (insbesondere beim Einsatz spezieller Software oder individueller Programmierung), sind die Berechnungsschritte und ggf. verwendete Software explizit abzubilden. Insbesondere der Programmcode ist in lesbarer Form anzugeben.

4.2.5.2.1 Patientencharakteristika

Die Patientencharakteristika der eingeschlossenen Studie werden in Tabelle 4-15 dargestellt.

4.2.5.2.2 Patientenrelevanz der Endpunkte

Die Bewertung erfolgt in der vorliegenden Nutzenbewertung hinsichtlich der untersuchten Endpunkte aus den Endpunktkategorien Mortalität, Morbidität, Gesundheitsbezogene Lebensqualität und Nebenwirkungen (siehe Abschnitt 4.3.1).

Der Nutzen eines Arzneimittels ist nach § 3 Abs. 1 des 5. Kapitels der Verfo des G-BA „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ (2). Der Nutzen einer Maßnahme soll sich dabei auf die Patient:innen beziehen und patientenrelevante Endpunkte beeinflussen, wobei als patientenrelevant in diesem Zusammenhang verstanden werden soll „wie eine Patientin oder ein Patient fühlt, ihre oder seine Funktionen und Aktivitäten wahrnehmen kann oder ob sie oder er überlebt“ (3).

Die folgende Tabelle 4-4 bietet eine Übersicht sämtlicher im Dossier berichteter Endpunkte sowie deren Zuordnung zu den Endpunktkategorien. Die Patientenrelevanz der einzelnen Endpunkte ist nachfolgend zur Tabelle näher erläutert.

Tabelle 4-4: Übersicht zu den Endpunktkategorien und Endpunkten

Endpunktkategorie	Endpunkte
Mortalität	<ul style="list-style-type: none"> • <u>Mortalität zu Woche 24</u>
Morbidität	<ul style="list-style-type: none"> • <u>Verbesserung der 6MWD zu Woche 24</u> • <u>Zeit bis zur klinischen Verschlechterung oder Tod</u> • <u>Verbesserung der Dyspnoe gemäß Borg-Skala zu Woche 24</u> • <u>Mehrkomponentenverbesserung^a</u> • <u>Krankheitssymptomatik und Gesundheitszustand zu Woche 24</u> <ul style="list-style-type: none"> – Symptom-Items des PAH-SYMPACT – EQ-5D VAS
Gesundheitsbezogene Lebensqualität	<ul style="list-style-type: none"> • <u>Gesundheitsbezogene Lebensqualität zu Woche 24</u> <ul style="list-style-type: none"> – Impact-Items des PAH-SYMPACT

Endpunktkategorie	Endpunkte
Nebenwirkungen	<ul style="list-style-type: none"> • <u>Unerwünschte Ereignisse Gesamtraten zu Woche 24</u> <ul style="list-style-type: none"> – Unerwünschte Ereignisse gesamt^a – Schwerwiegende unerwünschte Ereignisse – Schwere unerwünschte Ereignisse – Therapieabbruch wegen unerwünschter Ereignisse • <u>Unerwünschte Ereignisse (gegliedert nach SOC und PT)^a zu Woche 24</u> <ul style="list-style-type: none"> – Unerwünschte Ereignisse gesamt (SOC und PT) – Schwerwiegende unerwünschte Ereignisse (SOC und PT) – Schwere unerwünschte Ereignisse (SOC und PT) – Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) • <u>Unerwünschte Ereignisse (AEOI und AEOSI)^a zu Woche 24</u> <ul style="list-style-type: none"> – Schwerwiegende AEO(S)I – Schwere AEO(S)I
<p>a: Diese Endpunkte werden in der vorliegenden Nutzenbewertung ergänzend dargestellt. 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); AEOI: Adverse Events of Interest (Unerwünschte Ereignisse von Interesse); AEOSI: Adverse Events of Special Interest (Unerwünschte Ereignisse von besonderem Interesse); PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact; EQ-5D: EuroQoL-5 Dimensions; PT: Preferred Term; SOC: System Organ Class (Systemorganklasse); VAS: Visual Analog Scale (visuelle Analogskala)</p>	

Mortalität

Mortalität zu Woche 24

Das Überleben gilt indikationsübergreifend unbestritten als patientenrelevanter Endpunkt. Entsprechend wird dieser Endpunkt auch im Sozialgesetzbuch (SGB) V als eine der vorrangig zu berücksichtigenden Zielgrößen genannt (2).

Das Überleben bzw. die Mortalität ist auch in chronisch-progredienten Erkrankungen mit erhöhtem Mortalitätsrisiko von großer Bedeutung. Obwohl hier Langzeitdaten von besonderer Bedeutung sind, ist bei Vorliegen verzerrender Aspekte nach einem bestimmten Zeitpunkt eine Analyse zu einem früheren – aber dafür unverzerrt messbaren Zeitpunkt - sinnvoll. Daher wird für die primäre Analyse zur Mortalität die Auswertung zu Woche 24 herangezogen, zur Begründung siehe Abschnitt 4.2.5.4.

Gesamtüberleben

Analysen zum Gesamtüberleben werden ergänzend vorgelegt, siehe Abschnitt 4.2.5.4.

Morbidität

6-Minuten-Gehstrecke (6 Minute Walk Distance, 6MWD)

Die 6-Minuten-Gehstrecke dient zur Beurteilung der körperlichen Belastbarkeit. Dies ist für Patient:innen direkt erfahrbar und hat unmittelbar Auswirkungen auf Funktion und darauf, wie Aktivitäten wahrgenommen werden können.

Demzufolge ist die 6MWD ein patientenrelevanter Endpunkt. Dies konstatiert der G-BA auch in seinen Beschlüssen zu Macitentan und Riociguat (7, 8) sowie in der Niederschrift zur Beratungsanforderung 2022-B-349 (1).

Für die vorliegende Nutzenbewertung wird zur Beurteilung des Zusatznutzens eine klinische Relevanzschwelle von 40 Metern herangezogen. Dies ist im Einklang mit Publikationen, welche eine klinische Relevanzschwelle zwischen 33 und 42 Metern bei erwachsenen PAH-Patient:innen beschreiben (9-12). In einer neuen systematischen Übersichtsarbeit von Moutchia et al. wird eine MID von 33 Metern in der PAH über alle im Rahmen der Literaturrecherche festgelegten Subgruppenmerkmale (Alter, Geschlecht, Ethnie, PAH-Ätiologie, Body-Mass-Index, WHO-FK und PAH-Hintergrundtherapie sowie Baseline 6MWD-Wert) hinweg als klinisch signifikantes Responsekriterium beschrieben (13). In vergangenen Verfahren hat der G-BA eine klinische Relevanzschwelle von 40 Metern akzeptiert (8). MSD folgt der vom G-BA bisher akzeptierten MID von 40 Metern, jedoch ist anzumerken, dass die in der Fachliteratur als klinisch relevant erachtete MID damit weit überschritten wird.

Zeit bis zur klinischen Verschlechterung oder Tod (Time to Clinical Worsening, TTCW)

Der kombinierte Endpunkt „Zeit bis zur klinischen Verschlechterung oder Tod“ setzt sich aus Komponenten aus verschiedenen Endpunktkategorien (Mortalität und Morbidität) zusammen:

Mortalität:

- Tod

Morbidität:

- Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression
- Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Dosierung der Infusion mit Prostazyklin um 10 % oder mehr
- Notwendigkeit einer Atriostomie
- Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden)
- Verschlechterung der PAH definiert durch das Auftreten beider folgender Ereignisse, im Vergleich zu Baseline:
 - Verschlechterung der WHO-FK
 - Verschlechterung der 6MWD um ≥ 15 % bestätigt durch zwei Tests mit einem Zeitabstand von mindestens vier Stunden, aber nicht mehr als eine Woche

Im Rahmen der Beratungsanforderung 2022-B-349 stuft der G-BA neben der Gesamtmortalität in der Endpunktkategorie Morbidität die Dyspnoe, die körperliche Belastbarkeit und Funktionalität sowie die Hospitalisierung als patientenrelevant ein (1).

Darüber hinaus bewertet der G-BA die Leistung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression sowie die Notwendigkeit einer Atrioseptostomie als grundsätzlich patientenrelevant (1). Auch die Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Dosiserhöhung einer Prostazyklin-Gabe um 10 % oder mehr wird als patientenrelevant eingestuft, sofern diese symptomatisch bedingt ist (1).

Im Rahmen des kombinierten Endpunktes „Zeit bis zur klinischen Verschlechterung oder Tod“ ist die 6MWD als „Verschlechterung der 6MWD um ≥ 15 %, bestätigt durch 2 Tests mit einem Zeitabstand von mindestens 4 Stunden, aber nicht mehr als 1 Woche“ operationalisiert. Der G-BA merkt hierzu an, dass das methodische Vorgehen des IQWiG Methodenpapier 6.1 (14) zur Beurteilung klinischer Relevanzschwellen bei komplexen Skalen bei „natürlichen“ Skalen wie beispielsweise der Gehstrecke in Metern nicht zur Anwendung kommt. Vielmehr muss eine klinische Relevanzschwelle für eine natürliche Skala im Dossier dargestellt und begründet werden (1).

Für die Ableitung des Zusatznutzens wird die präspezifizierte Definition des kombinierten Endpunktes „Zeit bis zur klinischen Verschlechterung oder Tod“ herangezogen. Unter Würdigung der im Rahmen der Beratungsanforderung durch den G-BA festgehaltenen Anmerkungen wird zusätzlich eine Sensitivitätsanalyse präsentiert, die lediglich die vom G-BA als direkt patientenrelevant erachteten Komponenten beinhaltet (diese wird in Abschnitt 4.2.5.4 beschrieben).

Verbesserung der Dyspnoe gemäß Borg-Skala pre 6MWD-Test:

Im Rahmen der Beratungsanforderung 2022-B-349 stuft der G-BA die Dyspnoe als patientenrelevanten Endpunkt ein (1). Die Borg-Skala wird unmittelbar vor und nach der Durchführung des 6MWD erhoben. In der vorliegenden Nutzenbewertung werden die Ergebnisse zur Verbesserung der Dyspnoe gemäß Borg-Skala pre 6MWD-Test präsentiert. Dies erfolgt vor dem Hintergrund, dass die Patient:innen während des 6MWD dazu angehalten werden, bis an ihr körperliches Limit zu gehen. Daher sind auch bei einer unterschiedlichen körperlichen Leistungsfähigkeit vergleichbare Werte hinsichtlich der Dyspnoe gemäß Borg-Skala post 6MWD-Test zu erwarten.

Mehrkomponentenverbesserung:

Der kombinierte Endpunkt „Mehrkomponentenverbesserung“ setzt sich aus verschiedenen Einzelkomponenten zusammen:

- Verbesserung der 6MWD (≥ 30 Meter)

- Verbesserung in N-Terminal Pro-B-Type Natriuretic Peptide (NT-proBNP) (Verringerung des NT-proBNP Wertes um $\geq 30\%$) oder Erhaltung bzw. Erreichung eines NT-proBNP Wertes von < 300 ng/L)
- Verbesserung der WHO-FK oder Erhaltung der WHO-FK II

NT-proBNP:

In den Leitlinien der European Society of Cardiology (ESC) wird die Bestimmung von NT-proBNP bei Verdacht auf Herzinsuffizienz empfohlen und stellt einen wichtigen Marker für die behandelnden Ärztinnen und Ärzte dar (15).

WHO-FK:

Mithilfe der WHO-FK wird die Belastbarkeit der Patient:innen in vier Klassen (I-IV) eingeteilt. Die WHO-FK wird in der EMA-Guideline für klinische Studien in der PAH (5) als Endpunkt für klinische Studien im Bereich der PH empfohlen. In der Analyse eines PAH-Registers konnte die WHO-FK als eine Variable identifiziert werden, die eng mit dem Mortalitätsrisiko verknüpft ist (16). Des Weiteren stellt die WHO-FK im deutschen Versorgungsalltag einen gebräuchlichen Parameter dar und ist eine Grundlage für die Therapiezieldefinition (17). Auch dient sie in Übereinstimmung mit der europäischen Leitlinie (15) als Verlaufsparemeter, der die Morbidität der Patient:innen direkt widerspiegelt. Demzufolge ist die Verschlechterung der WHO-FK ein patientenrelevanter Endpunkt.

Zur Patientenrelevanz der 6MWD siehe oben. In der vorliegenden Nutzenbewertung wird die Mehrkomponentenverbesserung als ergänzender Morbiditätsendpunkt dargestellt.

Krankheitssymptomatik und Gesundheitszustand:

Die Erfassung der Krankheitssymptomatik und des Gesundheitszustands erfolgt über die Fragebögen PAH-SYMPACT (18) und die EQ-5D VAS (19).

PAH-SYMPACT

Der Fragebogen PAH-SYMPACT ist ein PAH-spezifischer Fragebogen zur Erfassung der Symptome und Auswirkungen der Erkrankung auf die körperliche und emotional/kognitive Funktionalität der Patient:innen und besteht aus insgesamt 23 Fragen. Zur Erfassung der Krankheitssymptomatik und Gesundheitszustand werden die 11 Fragen zu kardiopulmonalen und kardiovaskulären Symptomen sowie eine Frage zum Sauerstoffgebrauch zusammengefasst (siehe Tabelle 4-43). Der Fragebogen ist entsprechend den Vorgaben der U.S.-Gesundheitsbehörde Food and Drug Administration (FDA) validiert (18, 20).

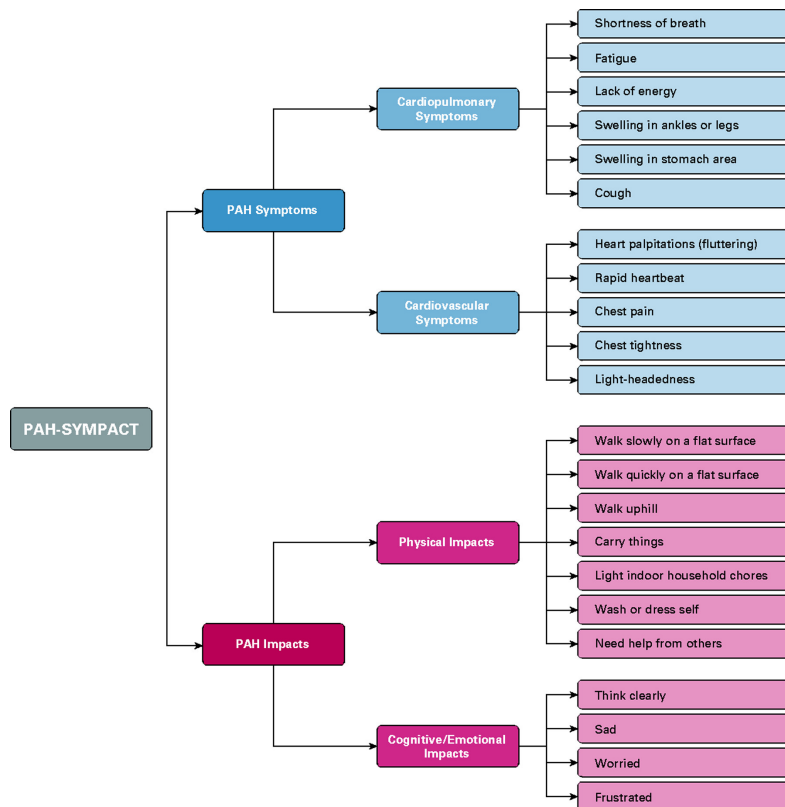


Abbildung 1: Domänen des PAH-SYMPACT-Fragebogens (21)

PAH: Pulmonale arterielle Hypertonie; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact

EQ-5D VAS

Der EQ-5D ist ein validierter und vielfach eingesetzter, krankheitsübergreifender Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität (22). Er besteht aus zwei Teilen, dem deskriptiven System und der VAS, wobei für die vorliegende Nutzenbewertung lediglich die VAS berücksichtigt wird. Anhand der VAS schätzen die Patient:innen ihren Gesundheitszustand auf einer Skala von 0 (schlechtest möglicher aktueller Gesundheitszustand) bis 100 (bestmöglicher aktueller Gesundheitszustand) ein. Entsprechend der Definition der Entwickler des Erhebungsbogens stellt die EQ-5D VAS eine Bewertung der gesamten individuellen gesundheitsbezogenen Lebensqualität dar (22). In der Nutzenbewertung gemäß § 35a SGB V wird die EQ-5D VAS nicht der Endpunktkategorie Gesundheitsbezogene Lebensqualität, sondern der Endpunktkategorie Morbidität zugeordnet. Daher werden die Ergebnisse der EQ-5D VAS im vorliegenden Dossier unter der Endpunktkategorie Morbidität dargestellt.

Gesundheitsbezogene Lebensqualität

Gesundheitsbezogene Lebensqualität

Krankheitsbedingte und therapiebedingte Symptome, mentale Gesundheitsprobleme sowie soziale Aspekte stellen bei Patient:innen mit einer PAH eine große Belastung dar, die sich negativ auf die Lebensqualität und den Krankheitsverlauf auswirken (23). Die gesundheitsbezogene Lebensqualität wird von den Patient:innen selbst berichtet und mithilfe

standardisierter Fragebögen erfasst. Sie spiegelt somit die Lebensrealität der Patient:innen wider und ist per Definition ein patientenrelevanter Endpunkt (2, 3). Zur Erfassung der gesundheitsbezogenen Lebensqualität werden die Impact-Items (insgesamt 11 Fragen) des validierten Fragebogens PAH-SYMPACT herangezogen (siehe voriger Abschnitt und Tabelle 4-51).

Nebenwirkungen

Unerwünschte Ereignisse Gesamtraten

Die Unerwünschte Ereignisse Gesamtraten umfassen

- Unerwünschte Ereignisse gesamt
- Schwerwiegende unerwünschte Ereignisse
- Schwere unerwünschte Ereignisse
- Therapieabbruch wegen unerwünschter Ereignisse.

Der Endpunkt Unerwünschte Ereignisse gesamt wird in der vorliegenden Nutzenbewertung ergänzend dargestellt.

Unerwünschte Ereignisse (gegliedert nach SOC und PT)

Für Unerwünschte Ereignisse gesamt (System Organ Class [SOC] und Preferred Terms [PT]), Schwerwiegende unerwünschte Ereignisse (SOC und PT), Schwere unerwünschte Ereignisse (SOC und PT) und Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) werden einzelne SOC - kodiert nach dem Medical Dictionary for Regulatory Activities (MedDRA) Version 25.0 oder höher - betrachtet. Die Darstellung von ausgewählten SOC und PT erfolgt ergänzend nach den in Tabelle 4-5 genannten Kriterien.

Unerwünschte Ereignisse von Interesse

In der vorliegenden Nutzenbewertung werden die Endpunkte Schwerwiegende Unerwünschte Ereignisse von Interesse (Adverse Events of Interest, AEOI) und Schwere Unerwünschte Ereignisse von Interesse (AEOI) sowie Schwerwiegende Unerwünschte Ereignisse von Besonderem Interesse (Adverse Events of Special Interest, AEOSI) und Schwere Unerwünschte Ereignisse von Besonderem Interesse (AEOSI) als ergänzende Endpunkte dargestellt.

4.2.5.2.3 Verwendete statistische Methoden

Generell werden diejenigen statistischen Methoden verwendet, die in der jeweiligen Studie präspezifiziert wurden. Sollten sachliche Gründe dagegensprechen (z. B. Anforderungen der Formatvorlage an Datenschnitte, verletzte Verteilungsannahmen) wird eine andere geeignete Methode verwendet. Für chronische Erkrankungen sind dies in der Regel Auswertungen zu einem festen Analysezeitpunkt mittels dichotomer Analysen oder Analysen für kontinuierliche Endpunkte zu einem festen Zeitpunkt.

Auswertungspopulationen

Full-Analysis-Set (FAS)

Die FAS-Population ist definiert als alle randomisierten Patient:innen, unabhängig davon, ob eine Studienmedikation verabreicht wurde oder nicht.

Die FAS-Population wird für die Analysen sämtlicher Wirksamkeitsendpunkte sowie patientenberichteter Endpunkte herangezogen.

Safety-Analysis-Set

Das Safety-Analysis-Set ist definiert als alle randomisierten Patient:innen, die mindestens eine Dosis der Studienmedikation verabreicht bekommen haben. Patient:innen werden bei den Auswertungen der Behandlung hinzugerechnet, welche sie anfänglich tatsächlich bekommen haben.

Das Safety-Analysis-Set wird für die Analysen sämtlicher Endpunkte der Kategorie Nebenwirkungen herangezogen.

Datenschnitte

Falls keine sachlichen Gründe dagegensprechen, wird der Datenschnitt mit der längsten Nachbeobachtungsdauer herangezogen. Sollte ein früherer Datenschnitt einen Informationsgewinn mit sich bringen, wird dieser zusätzlich dargestellt. Eine detaillierte Übersicht der geplanten bzw. durchgeführten Datenschnitte aus der Studie STELLAR bietet Abschnitt 4.3.1.2.1.

Auswertungsmethoden

Auswertungen werden durch adäquate Effektschätzer mit zugehörigen 95 %-KI und p-Werten geeigneter zweiseitiger statistischer Tests zu einem Signifikanzniveau von 5 % dargestellt. Die Analysemethode unterscheidet sich abhängig von der gemessenen Skala des jeweiligen Endpunkts und wird im Folgenden beschrieben. Für Subgruppenanalysen wird eine Analysemethode analog zur Hauptanalyse verwendet. Die Analysemethode zur Untersuchung der Interaktionsterme zwischen den Subgruppenkategorien ist in Abschnitt 4.2.5.5 beschrieben.

Dichotome Analysen

Dichotome Endpunkte werden mithilfe der Stratum-adjustierten Mantel-Haenszel-Methode ausgewertet. Zum Vergleich zwischen den Behandlungsarmen wird das RR mit zugehörigen 95 %-KI- und zweiseitigem p-Wert (Wald-Test) dargestellt. Bei Ereignisraten $\leq 1\%$ oder $\geq 99\%$ in mindestens einem Arm wird das Peto-OR anstatt des RR angegeben. Zusätzlich wird die adjustierte Differenz mit zugehörigem 95 %-KI gezeigt, die mithilfe der Miettinen und Nurminen-Methode ausgewertet wird (24). Alle Analysen werden stratifiziert nach WHO-FK (Klasse II vs. Klasse III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) durchgeführt.

Für die Analyse der Endpunkte der Kategorie Nebenwirkungen wird das RR basierend auf einer 2 x 2-Kontingenztafel mit einem zweiseitigen 95 %-Wald-KI (basierend auf einer Log-

Transformation des RR) berechnet. Falls in einem Arm keine Ereignisse auftreten, wird eine Nullzellenkorrektur mit einem Wald-modifizierten KI und dazugehörigen nicht adjustierten p-Werten basierend auf einem Cochran-Mantel-Haenszel-Modell durchgeführt. Falls in beiden Armen keine Patient:innen ein Ereignis oder in beiden Armen alle Patient:innen ein Ereignis aufweisen, so wird n.a. (not applicable) berichtet. Unerwünschte Ereignisse (gegliedert nach SOC und PT) werden dargestellt, sofern die unter Tabelle 4-5 genannten Inzidenzkriterien erfüllt werden.

Tabelle 4-5: Systematik zur Darstellung des Endpunkts Unerwünschte Ereignisse (gegliedert nach SOC und PT)

Endpunkte	Darstellung der SOC bzw. PT im Dossier (Kriterium)
Unerwünschte Ereignisse gesamt (SOC und PT)	Ereignisse bei mindestens 10 % der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen UND bei mindestens 1 % der Patient:innen in einem Studienarm
Schwerwiegende unerwünschte Ereignisse (SOC und PT) und Schwere unerwünschte Ereignisse (SOC und PT)	Ereignisse bei mindestens 5 % der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen UND bei mindestens 1 % der Patient:innen in einem Studienarm
Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT)	Ereignisse bei > 0 % der Patient:innen in einem Studienarm werden deskriptiv dargestellt
PT: Preferred Term; SOC: System Organ Class (Systemorganklasse)	

Für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) wird unterstützend zu dem nominalen p-Wert ein adjustierter p-Wert dargestellt. Dieser p-Wert wird auf SOC-Ebene anhand der False Discovery Rate (FDR)-Methodik (25) bzw. auf Ebene der PT anhand der Double FDR (dFDR)-Methodik (26) berechnet. Durch Berücksichtigung des adjustierten p-Werts können falsch-positive (signifikante) Ergebnisse des nominalen p-Wertes aufgedeckt und somit der Einfluss des multiplen hierarchischen Testens auf die Ergebnisinterpretation der unerwünschten Ereignisse nach SOC bzw. PT kontrolliert werden.

Deskriptive Analysen

Folgende Parameter werden im Rahmen der deskriptiven Analyse nach Behandlungsarm und Visite dargestellt: Anzahl der Patient:innen, Mittelwert (inklusive Standardabweichung), Median (inklusive erstes Quartil und drittes Quartil) sowie Minimum und Maximum.

Für kombinierte Endpunkte werden zusätzlich zur Gesamtauswertung die Einzelkomponenten anhand der Anzahl der Patient:innen (inklusive Prozentangabe) nach Behandlungsarm deskriptiv dargestellt.

Time-to-Event Analysen

Die Berechnung des HR inklusive zugehörigem 95 %-KI und zweiseitigem p-Wert (Wald-Test; Score-Test im Fall von 0 Ereignissen in einem der Studienarme) basiert auf der Anwendung eines Cox-Regressionsmodells mit Behandlung als Kovariate, stratifiziert nach WHO-FK (Klasse II vs. Klasse III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie).

Für die Schätzung der medianen Ereigniszeit wird die Kaplan-Meier-Methode verwendet. Zusätzlich werden Kaplan-Meier-Kurven dargestellt.

Analysen kontinuierlicher Endpunkte

Für den primären Endpunkt wurde bereits bei der Studienplanung der Studie STELLAR eine nicht parametrische Methode vorgesehen, da angenommen wurde, dass der Endpunkt keiner Normalverteilung folgt. In der Studie STELLAR sind die meisten kontinuierlich erhobenen Endpunkte nicht normalverteilt (Shapiro-Wilk-Test p-Wert > 0,05), daher wird für alle kontinuierlichen Endpunkte die folgende, für den primären Endpunkt präspezifizierte, nichtparametrische Methode verwendet.

Kontinuierliche Endpunkte werden anhand eines stratifizierten Wilcoxon-Tests mit WHO-FK (Klasse II vs. Klasse III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie) als Strata analysiert. Als Schätzer für den medianen Unterschied zwischen den Behandlungsarmen wird der Hodges-Lehman Location Shift inklusive zugehörigem 95 %-KI mit zweiseitigem p-Wert dargestellt.

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 Effekten nach der Knapp-Hartung-Methode mit der Paule-Mandel-Methode zur Heterogenitätsschätzung⁵ erfolgen. Im Fall von sehr wenigen Studien ist die Heterogenität nicht verlässlich schätzbar. Liegen daher weniger als 5 Studien vor, ist auch die Anwendung eines Modells mit festem Effekt oder eine qualitative Zusammenfassung in Betracht zu ziehen. Kontextabhängig können auch alternative Verfahren wie z. B. Bayes'sche Verfahren oder Methoden aus dem Bereich der generalisierten

⁵ Veroniki AA, Jackson D, Viechtbauer W, Bender R, Knapp G, Kuss O et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. *Cochrane Database Syst Rev* 2015: 25-27.

linearen Modelle in Erwägung gezogen 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 verwendet⁶ werden.

Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand geeigneter statistische Maße auf Vorliegen von Heterogenität^{7,5} erfolgen. Die Heterogenitätsmaße sind unabhängig von dem Ergebnis der Untersuchung auf Heterogenität immer anzugeben. Ist die Heterogenität der Studienergebnisse nicht bedeutsam (z. B. p -Wert für Heterogenitätsstatistik $\geq 0,05$), 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.

Nicht zutreffend.

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.

⁶ 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.

⁷ Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-560.

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.

Mortalität

Die primäre Analyse der Mortalität wird auf Basis der Ergebnisse zu Woche 24 durchgeführt.

Verzerrende Aspekte nach Woche 24

Die Studie STELLAR bestand aus zwei konsekutiven Behandlungsphasen: die erste Phase war die 24 Wochen lange, doppelblinde und Placebo-kontrollierte Behandlungsphase (Double-Blind Placebo-Controlled, DBPC). Ihr folgte eine langfristige doppelblinde Behandlungsphase (Long-Term Double-Blind, LTDB). Patient:innen welche die DBPC-Phase durchlaufen haben, führten die Studie in der LTDB-Phase weiter, bis der/die letzte randomisierte Patient:in die 24 Wochen lange DBPC-Phase abgeschlossen hatte. Patient:innen konnten im Rahmen der offenen und nicht vergleichenden Long-Term-Extension (LTE)-Studie SOTERIA ihre Therapie mit Sotatercept weiterführen oder aus dem Kontrollarm die Therapie mit Sotatercept initiieren. Die LTE-Studie SOTERIA-Studie ist eine laufende Studie, die zuvor in anderen Sotatercept-Studien eingeschlossenen Patient:innen die Möglichkeit gibt, ihre Behandlung mit Sotatercept weiterzuführen oder zu initiieren. Der Übergang in die LTE-Studie SOTERIA war erst nach Abschluss der Studie STELLAR vorgesehen, konnte aber auch früher erfolgen, wenn:

- während der DBPC-Phase eine klinische Verschlechterung erfolgte, die keine Anpassung der medikamentösen Therapie nach sich zog, konnte nach Abschluss der DBPC-Phase direkt in die LTE-Studie SOTERIA gewechselt werden
- nach Abschluss der DBPC-Phase eine klinische Verschlechterung vorlag konnte direkt in die LTE-Studie SOTERIA gewechselt werden.

Das bedeutet, dass auch Patient:innen aus dem Kontrollarm mit einer Verschlechterung des Gesundheitszustandes frühzeitig auf die aktive Therapie mit Sotatercept im Rahmen der SOTERIA wechseln konnten. Patient:innen aus dem Sotatercept-Arm haben in SOTERIA ihre Therapie mit Sotatercept weitergeführt. Es erfolgt also ein Wechsel der Behandlung im Kontrollarm, jedoch nicht im Interventionsarm. Eine Time-to-Event (TTE) Analyse der STELLAR-Studie allein führt zu einem potentiell verzerrten Ergebnis, da Patient:innen aufgrund einer Krankheitsprogression früher als regulär zu Studienende in die Long-Term-Extension (LTE)-Studie SOTERIA wechseln konnten und hiervon insbesondere im Kontrollarm auch Gebrauch gemacht haben, vergleiche Abschnitt 4.3.1.2.1.

Zum einen wird der Effekt auf das Gesamtüberleben durch den Wechsel der Patient:innen mit klinischer Verschlechterung potentiell verzerrt, da anzunehmen ist, dass diese ein höheres Mortalitätsrisiko haben. Da Patient:innen aus dem Kontrollarm häufiger frühzeitig in die LTE-

Studie SOTERIA gewechselt haben, würde der wahre Behandlungseffekt zuungunsten des Interventionsarmes verzerrt.

Zum anderen ist basierend auf den Daten der Studie STELLAR anzunehmen, dass die Intervention zu einer Verlängerung des Überlebens führt, und daher wird bei nicht berücksichtigen des Therapiewechsels in der Analyse der wahre Behandlungseffekt unterschätzt.

Ergänzende Analyse und deren Sensitivitätsanalyse

Daher wird eine TTE-Analyse der kombinierten Daten aus der Studie STELLAR und SOTERIA vorgelegt, um die Langzeiteffekte zu untersuchen. Diese unterschätzt den wahren Behandlungseffekt jedoch potentiell. Daher wird zusätzlich eine weitere Analyse vorgelegt, die für den Behandlungswechsel korrigiert:

Im IQWiG-Arbeitspapier zu Treatment Switching in onkologischen Studien (27) diskutiert das IQWiG insbesondere zwei Methoden zur Korrektur bei Behandlungswechsel: Inverse Probability of Censoring Weighting (IPCW) und Rank-preserving structural Failure Time (RPSFT) als potentiell geeignete statistische Methoden.

Die IPCW-Methode setzt voraus, dass die Prognose den Therapiewechsel und den Endpunkt beeinflusst (Patient:innen mit einer „schlechten“ Prognose wechseln eher die Behandlung als solche mit einer „guten“ Prognose). Da in der STELLAR-Studie jedoch alle Patient:innen – unabhängig von ihrer Prognose – aus dem Placebo-Arm in die LTE-Studie SOTERIA wechseln konnten, ist diese Voraussetzung nur für einige Patient:innen (diejenigen, die aufgrund einer Krankheitsprogression frühzeitig in SOTERIA gewechselt haben) erfüllt. Somit ist die IPCW-Methode im vorliegenden Fall nicht geeignet.

„Die Motivation der Methode Rank-preserving structural Failure Time (RPSFT) besteht darin, den Therapieeffekt unter der Annahme, dass alle Patientinnen und Patienten über den gesamten Studienverlauf gemäß ihrer Randomisierung behandelt worden wären, zu schätzen“ (27). Zusätzlich zu der naiven Analyse (TTE-Analyse der kombinierten Daten aus der Studie STELLAR und SOTERIA), die den Behandlungswechsel ignoriert, wird eine TTE-Analyse mit Korrektur nach der Methode RPSFT als Sensitivitätsanalyse vorgelegt.

Zeit bis zur klinischen Verschlechterung oder Tod

Bei den Analysen des kombinierten Endpunktes „Zeit bis zur klinischen Verschlechterung oder Tod“ wird zur Überprüfung der Robustheit der Ergebnisse im Sinne einer Bestätigung der Patientenrelevanz gemäß der Einschätzung des G-BA eine Sensitivitätsanalyse auf Basis der Beratung des G-BA durchgeführt, vgl. Abschnitt 4.2.5.2.2. Dabei unterscheiden sich die Haupt- und Sensitivitätsanalyse in den Einzelkomponenten:

Tabelle 4-6: Sensitivitätsanalyse für den Endpunkt „Zeit bis zur klinischen Verschlechterung oder Tod“

Einzelkomponenten Hauptanalyse	Einzelkomponenten Sensitivitätsanalyse
Gesamtüberleben	Gesamtüberleben
Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression	Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression
Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Dosierung der Infusion mit Prostazyklin um 10 % oder mehr	- ^a
Notwendigkeit einer Atrioseptostomie	Notwendigkeit einer Atrioseptostomie
Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden)	Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden)
Verschlechterung der PAH definiert durch das Auftreten beider folgender Ereignisse, im Vergleich zu Baseline: <ul style="list-style-type: none"> ○ Verschlechterung der WHO-Funktionsklasse ○ Verschlechterung der 6MWD um $\geq 15\%$ bestätigt durch 2 Tests mit einem Zeitabstand von mindestens 4 Stunden, aber nicht mehr als 1 Woche 	Verschlechterung der PAH definiert durch die Verschlechterung der 6MWD um ≥ 40 Meter
<p>a: In der Studie STELLAR wurden die Symptome, die eine Therapieanpassung erforderten, nicht erfasst. 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation)</p>	

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 (zu ergänzenden Kriterien zur Darstellung siehe Abschnitt 4.3.1.3.2).

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 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. 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.

Das Vorgehen bei der Betrachtung von Subgruppen gliedert sich in zwei Schritte.

1. Im ersten Schritt wird ermittelt, welche Endpunkte und zugehörigen Subgruppenmerkmale in der vorliegenden Nutzenbewertung für die Betrachtung von Subgruppenanalysen herangezogen werden.
2. Der zweite Schritt befasst sich mit der Durchführung von Subgruppenanalysen und dem methodischen Vorgehen zur Bewertung möglicher Effektmodifikationen.

Schritt 1: Auswahl von Endpunkten (inkl. SOC und PT) und zugehörigen Subgruppenmerkmalen

Für die patientenrelevanten Endpunkte aus den Endpunktkategorien Mortalität, Morbidität, Gesundheitsbezogene Lebensqualität und Nebenwirkungen werden entsprechend der VerFO (2) des G-BA Subgruppen zu Alter, Geschlecht, Krankheitsschwere und Region dargestellt. Darüber hinaus werden für die patientenrelevanten Endpunkte aus den Endpunktkategorien Mortalität, Morbidität, Gesundheitsbezogene Lebensqualität und Nebenwirkungen zusätzlich die Subgruppen berichtet, die aus medizinischer Perspektive wichtige Indikatoren für die Erkrankung sind. Zudem werden alle a priori definierten Subgruppenmerkmale für den primären Studienendpunkt 6MWD auf mögliche Effektmodifikationen dargestellt.

SOC und PT werden für die Subgruppenanalyse in Betracht gezogen, falls diese zum einen die unten angegebenen Inzidenzgrenzen überschreiten und zum anderen ein statistisch signifikanter Unterschied der jeweiligen SOC bzw. PT bei Betrachtung der Zielpopulation vorliegt:

- Unerwünschte Ereignisse gesamt:
[Ereignisse bei mindestens 10 % der Patient:innen in einem Studienarm]
ODER
[Ereignisse bei mindestens 10 Patient:innen UND bei mindestens 1 % der Patient:innen in einem Studienarm]
- Schwerwiegende unerwünschte Ereignisse und Schwere unerwünschte Ereignisse:
[Ereignisse bei mindestens 5 % der Patient:innen in einem Studienarm]
ODER
[Ereignisse bei mindestens 10 Patient:innen UND bei mindestens 1 % der Patient:innen in einem Studienarm]

Für die daraus resultierenden SOC und PT werden Subgruppenmerkmale analog zu der zuvor aufgeführten Systematik für die Endpunktkategorien dargestellt. Subgruppen für den Endpunkt Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) werden nicht dargestellt, da diese wie gefordert deskriptiv dargestellt werden.

Schritt 2: Durchführung von Subgruppenanalysen und methodisches Vorgehen zur Bewertung möglicher Effektmodifikationen

Die Durchführung von Subgruppenanalysen dient der Identifikation möglicher Effekte, die zur Modifikation der Ergebnisse in Subgruppen führen könnten; es soll hiermit insbesondere die Fragestellung geklärt werden, ob für eine Subgruppe ein anders zu bewertender Zusatznutzen vorliegt (3). Die Ähnlichkeit der Subgruppen wird mithilfe eines mehrstufigen Verfahrens überprüft (siehe Abbildung 2).

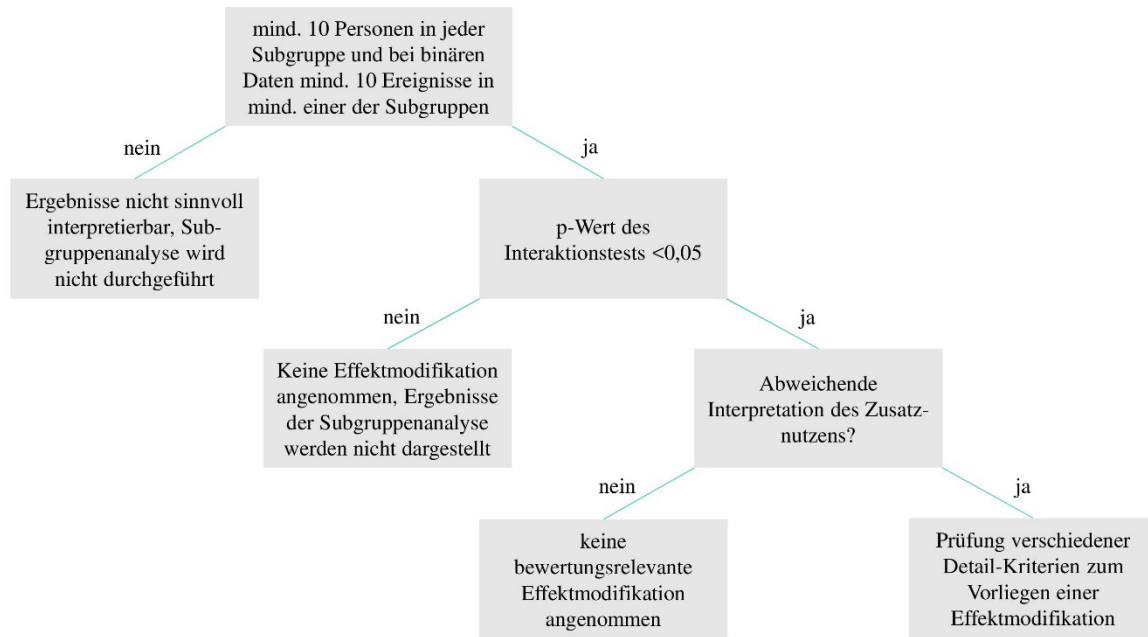


Abbildung 2: Mehrstufiges Vorgehen zur Durchführung von Subgruppenanalysen und methodisches Vorgehen zur Bewertung möglicher Effektmodifikationen (Eigene Darstellung)

Vor Durchführung des Interaktionstests wird geprüft, ob einerseits eine ausreichende Anzahl an Patient:innen in den Subgruppen eingeschlossen ist und andererseits genügend Ereignisse aufgetreten sind, um das Ergebnis der Subgruppenanalysen sinnvoll interpretieren zu können. In Anlehnung an das IQWiG-Methodenpapier 7.0 wird ein Interaktionstest folglich nur dann durchgeführt, wenn die in Tabelle 4-7 genannten Kriterien erfüllt sind (3). Sollten diese nicht erfüllt sein, erfolgt lediglich eine deskriptive Darstellung der Subgruppe.

Tabelle 4-7: Kriterien für die Berechnung der Interaktionstests in Anlehnung an das IQWiG-Methodenpapier 7.0 (3)

Kriterien				
<ul style="list-style-type: none"> • Es müssen mindestens 10 Patient:innen in jeder Kategorie enthalten sein: $N_1 + N_2 \geq 10$ und $N_3 + N_4 \geq 10$. • Des Weiteren müssen in mindestens einer Kategorie mindestens 10 Ereignisse enthalten sein: $n_1 + n_2 \geq 10$ oder $n_3 + n_4 \geq 10$. 				
Erläuterung der Kategorien				
	Sotatercept		Placebo	
	Anzahl der Patient:innen	Patient:innen mit Ereignis ^a	Anzahl der Patient:innen	Patient:innen mit Ereignis ^a
Kategorie 1	N ₁	n ₁	N ₂	n ₂
Kategorie 2	N ₃	n ₃	N ₄	n ₄
a: Zusätzliches Kriterium bei dichotomen und TTE-Endpunkten				

IQWiG: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; N: Anzahl der Patient:innen; n: Patient:innen mit Ereignis; TTE: Time-To-Event

Zur Überprüfung potenzieller Effektmodifikationen wird ein Interaktionstest durchgeführt. Für das vorliegende Dossier wird hierfür das jeweilige Modell um den Interaktionsterm *Behandlung*Subgruppe* erweitert. Anschließend erfolgt die Testung mittels des Likelihood-Ratio-Tests. Für die in der Studie STELLAR präspezifizierte Analyse der kontinuierlichen Endpunkte (Wilcoxon-Test) erfolgt die Testung auf potenzielle Interaktion mittels Cochran's Q-Test.

Die Ergebnisse der Interaktionstests werden für alle relevanten Endpunkte tabellarisch in Abschnitt 4.3.1.3.2 dargestellt. Für Subgruppenanalysen mit nicht signifikantem Interaktionstest ($p \geq 0,05$) wird davon ausgegangen, dass kein unterschiedlicher Effekt in den Subgruppen vorliegt. Aus diesem Grund ist ausschließlich das Ergebnis der Gesamtstudie relevant. Auf eine Darstellung der jeweiligen Subgruppenanalysen im Ergebnisteil des Dossiers wird in diesem Fall verzichtet. Zur Sicherstellung der formalen Vollständigkeit werden die Ergebnisse der Subgruppenanalysen mit nicht signifikantem Interaktionstest dem Dossier in Anhang 4-G beigelegt.

Ein zum Niveau $\alpha = 0,05$ signifikantes Ergebnis im Interaktionstest wird zunächst als statistisches Signal für einen Unterschied zwischen den Subgruppen gewertet, es erfolgt eine ausführliche Darstellung dieser Ergebnisse in Abschnitt 4.3.1.3.2.2. Bei der Testung einer größeren Anzahl statistischer Hypothesen ist jedoch zu beachten, dass eine gewisse Anzahl der durchgeführten Tests aufgrund der angenommenen Irrtumswahrscheinlichkeit, also per Definition, zu einem falsch-positiven Ergebnis führen. Ein statistisch signifikanter Interaktionstest ($p < 0,05$) wird daher im vorliegenden Dossier nicht automatisch mit dem Vorhandensein einer tatsächlichen Effektmodifikation gleichgesetzt. Vielmehr erfolgt in weiteren Schritten eine detaillierte Analyse zum Vorliegen einer tatsächlichen Effektmodifikation.

Hierbei wird zunächst geprüft, ob sich aufgrund der Lage der Effektschätzer und deren KI in den einzelnen Subgruppen Unterschiede in der Interpretation des Zusatznutzens zwischen den Subgruppenanalysen und den Analysen der Gesamtstudienpopulation ergeben. Diese Prüfung umfasst beispielsweise das Vorliegen eines Zusatznutzens in einer Subgruppe, in der zuvor kein Zusatznutzen vorlag, oder das Vorliegen eines unterschiedlichen Ausmaßes des Zusatznutzens in den Subgruppen. Ergibt die Prüfung der Lage der Effektschätzer und deren KI keine abweichende Interpretation des Zusatznutzens, so wird keine bewertungsrelevante Effektmodifikation angenommen.

Aufgrund des zuvor beschriebenen Problems des multiplen Testens muss für Subgruppenanalysen mit positivem Interaktionstest und abweichender Interpretation des Zusatznutzens abschließend eingeschätzt werden, ob es sich um eine tatsächliche Effektmodifikation oder um ein falsch positives Ergebnis des Interaktionstests handelt. Hierzu werden in Anlehnung an die EMA-Leitlinien zur Untersuchung von Subgruppen in

konfirmatorischen klinischen Studien sowie an ein Gutachten des medizinischen Dienstes der Spitzenverbände der Krankenkassen die folgenden Kriterien herangezogen (28, 29):

- Besteht eine medizinisch, biologisch oder physiologisch plausible Erklärung für die mögliche Effektmodifikation?
- Liegen konsistente Ergebnisse zur möglichen Effektmodifikation über mehrere Endpunkte vor?
- Liegen zur möglichen Effektmodifikation bereits ähnliche Ergebnisse aus anderen Studien vor?

Analysierte Subgruppen

Für alle patientenrelevanten Endpunkte und alle relevanten SOC/PT werden entsprechend der VerfO des G-BA (2) Ergebnisse zu Subgruppen nach

- Geschlecht (Weiblich vs. Männlich)
- Alter (< 65 Jahre vs. ≥ 65 Jahre)
- Krankheitsschwere (WHO-FK II vs. III)
- Region (WHO Stratum A vs. Rest der Welt [Rest of the World, RoW])

dargestellt. Für alle patientenrelevanten Endpunkte und alle relevanten SOC/PT wird zusätzlich die folgende Subgruppe von besonderem Interesse dargestellt:

- PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)

Für den primären Endpunkt der Studie STELLAR Veränderung der 6MWD werden zudem die präspezifizierten Subgruppenanalysen dargestellt:

- Ätiologische Subgruppen der PAH (idiopathische pulmonale arterielle Hypertonie [IPAH] vs. heritable pulmonale arterielle Hypertonie [hPAH] vs. Medikamenten-induzierte PAH vs. Bindegewebserkrankungen vs. angeborene Herzfehler)
- Pulmonary Vascular Resistance (PVR)-Wert zu Baseline (≤ 800 vs. > 800 dyn*sek/cm⁵)
- Prostazykline als Infusionstherapie zu Baseline (Ja vs. Nein)

4.2.5.6 Indirekte Vergleiche

Zurzeit sind international Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Es besteht dabei internationaler Konsens, dass Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen

Komparator (häufig als nicht adjustierte indirekte Vergleiche bezeichnet) regelhaft keine valide Analysemethoden darstellen⁸. Eine Ausnahme kann das Vorliegen von dramatischen Effekten sein. An Stelle von nicht adjustierten indirekten Vergleichen sollen je nach Datenlage einfache adjustierte indirekte Vergleiche⁹ oder komplexere Netzwerk-Meta-Analysen (auch als „Mixed Treatment Comparison [MTC] Meta-Analysen“ oder „Multiple Treatment Meta-Analysen“ bezeichnet) für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche berechnet werden. Aktuelle Verfahren wurden beispielsweise von Lu und Ades (2004)¹⁰ und Rücker (2012)¹¹ vorgestellt.

Alle Verfahren für indirekte Vergleiche gehen im Prinzip von den gleichen zentralen Annahmen aus. Hierbei handelt es sich um die Annahmen der Ähnlichkeit der eingeschlossenen Studien, der Homogenität der paarweisen Vergleiche und der Konsistenz zwischen direkter und indirekter Evidenz 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 ist¹².

Das Ergebnis eines indirekten Vergleichs kann maßgeblich von der Auswahl des Brückenkomparators bzw. der Brückenkomparatoren abhängen. Als Brückenkomparatoren sind dabei insbesondere Interventionen zu berücksichtigen, für die sowohl zum bewertenden Arzneimittel als auch zur zweckmäßigen Vergleichstherapie mindestens eine direkt vergleichende Studie vorliegt (Brückenkomparatoren ersten Grades). Insgesamt ist es notwendig, die zugrunde liegende Methodik für alle relevanten Endpunkte genau und reproduzierbar zu beschreiben und die zentralen Annahmen zu untersuchen^{13, 14, 15}

⁸ Bender R, Schwenke C, Schmoor C, Hauschke D. Stellenwert von Ergebnissen aus indirekten Vergleichen - Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR [online]. [Zugriff: 31.10.2016]. URL: http://www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWIG_GMDS_IBS_DR.pdf.

⁹ Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.

¹⁰ Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.

¹¹ Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012; 3(4): 312-324.

¹² Schöttker B, Lüthmann D, Boukhemair D, Raspe H. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.

¹³ 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.

¹⁴ Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study *BMJ* 2011; 343 :d4909

¹⁵ Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods* 2013; 4(4): 291-323.

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

- *Benennung aller potentiellen Brückenkomparatoren ersten Grades und ggf. Begründung für die Auswahl.*
- *Genaue Spezifikation des statistischen Modells inklusive aller Modellannahmen. Bei Verwendung eines Bayes'schen Modells sind dabei auch die angenommenen A-priori-Verteilungen (falls informative Verteilungen verwendet werden, mit Begründung), die Anzahl der Markov-Ketten, die Art der Untersuchung der Konvergenz der Markov-Ketten und deren Startwerte und Länge zu spezifizieren.*
- *Art der Prüfung der Ähnlichkeit der eingeschlossenen Studien.*
- *Art der Prüfung der Homogenität der Ergebnisse direkter paarweiser Vergleiche.*
- *Art der Prüfung der Konsistenzannahme im Netzwerk.*
- *Bilden Sie den Code des Computerprogramms inklusive der einzulesenden Daten 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.*

Nicht zutreffend.

4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

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)

Falls für die Bewertung des Zusatznutzens mehrere Komparatoren (z.B. Wirkstoffe) herangezogen werden, sind die Aussagen zum Zusatznutzen primär gegenüber der Gesamtheit der gewählten Komparatoren durchzuführen (z. B. basierend auf Meta-Analysen unter gemeinsamer Betrachtung aller direkt vergleichender Studien). Spezifische methodische Argumente, die gegen eine gemeinsame Analyse sprechen (z. B. statistische oder inhaltliche Heterogenität), sind davon unbenommen. Eine zusammenfassende Aussage zum Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ist in jedem Fall erforderlich.

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, Angabe zu geplanten und durchgeführten Datenschnitten und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-8: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen/abgebrochen/laufend)	Studiendauer ggf. Datenschnitt	Therapiearme
STELLAR ^a (NCT04576988)	ja	ja	abgeschlossen	24 Wochen	<u>Interventionsarm:</u> Initialdosis: 1 x 0,3 mg/kg s.c. Sotatercept Erhaltungsdosis: 0,7 mg/kg s.c. Sotatercept <u>Kontrollarm:</u> Placebo
ZENITH (NCT04896008)	Nein	Ja	Laufend	Ereignisgesteuert	<u>Interventionsarm:</u> Initialdosis: 1 x 0,3 mg/kg s.c. Sotatercept Erhaltungsdosis: 0,7 mg/kg s.c. Sotatercept <u>Kontrollarm:</u> Placebo
PULSAR (NCT03496207)	nein	ja	abgeschlossen	Placebo-kontrollierte Behandlungsphase: 24 Wochen Long-term Extension-Phase: 30 Monate	<u>Interventionsarm 1:</u> 0,3 mg/kg s.c. Sotatercept <u>Interventionsarm 2:</u> 0,7 mg/kg s.c. Sotatercept

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen/abgebrochen/laufend)	Studiendauer ggf. Datenschnitt	Therapiearme
					<u>Kontrollarm:</u> Placebo
HYPERION (NCT04811092)	nein	ja	laufend	Ereignis-gesteuert	<u>Interventionsarm:</u> Initialdosis: 1 x 0,3 mg/kg s.c. Sotatercept Erhaltungsdosis: 0,7 mg/kg s.c. Sotatercept <u>Kontrollarm:</u> Placebo
<p>A: Neben der Studie STELLAR als RCT wird zusätzlich die nicht randomisierte, nicht vergleichende Long-Term Extension Studie SOTERIA herangezogen (30). Die SOTERIA-Studie ist eine laufende Studie, die zuvor in anderen Sotatercept-Studien eingeschlossenen Patient:innen die Möglichkeit gibt, ihre Behandlung mit Sotatercept weiterzuführen oder zu initiieren. Für die vorliegende Nutzenbewertung wird die SOTERIA-Studie unterstützend für den Endpunkt Mortalität herangezogen.</p> <p>kg: Kilogramm; mg: Milligramm; NCT: National Clinical Trial; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie); s.c.: Subkutan</p>					

Geben Sie an, welchen Stand die Information in Tabelle 4-8 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 Dokuments maßgeblichen Zeitpunkt liegen.

Stand der Informationen: 07.08.2024.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-8 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-9: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
PULSAR (NCT03496207)	Nicht E2 (Falsche Intervention: zu hohe Initialdosis)
HYPERION (NCT04811092)	Nicht E7 (Keine Ergebnisse)
ZENITH (NCT04896008)	Nicht E7 (Keine Ergebnisse)
NCT: National Clinical Trial; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)	

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.

Die im Abschnitt 4.2.3.2 beschriebene systematische bibliografische Literaturrecherche wurde am 09.07.2024 in den Datenbanken MEDLINE, EMBASE und Cochrane auf der OVID-Plattform durchgeführt. Die Suche nach Studien mit dem zu bewertenden Arzneimittel Sotatercept ergab insgesamt 89 Treffer. Nach automatisiertem und händischem Ausschluss der Duplikate (n = 26) wurden die verbliebenen 63 Publikationen gemäß den definierten Einschlusskriterien in Abschnitt 4.2.2 (Tabelle 4-3) von zwei Personen unabhängig voneinander hinsichtlich ihrer Relevanz gesichtet.

Basierend auf den Informationen aus Titel und – soweit vorhanden – Abstracts wurden 61 Publikationen als nicht relevant eingestuft und ausgeschlossen. Die verbleibenden zwei Publikation wurde im Volltext gesichtet. Davon erfüllte eine Publikation zur Studie

STELLAR die Einschlusskriterien für die Patientenpopulation: Erwachsene Patient:innen mit PAH der WHO-FK II bis III (Abbildung 3) (31). Die Suchstrategien samt Trefferzahlen sind in Anhang 4-A1 angegeben. Die ausgeschlossenen Publikationen sind in Anhang 4-C1 unter Angabe des jeweiligen Ausschlussgrundes aufgeführt.

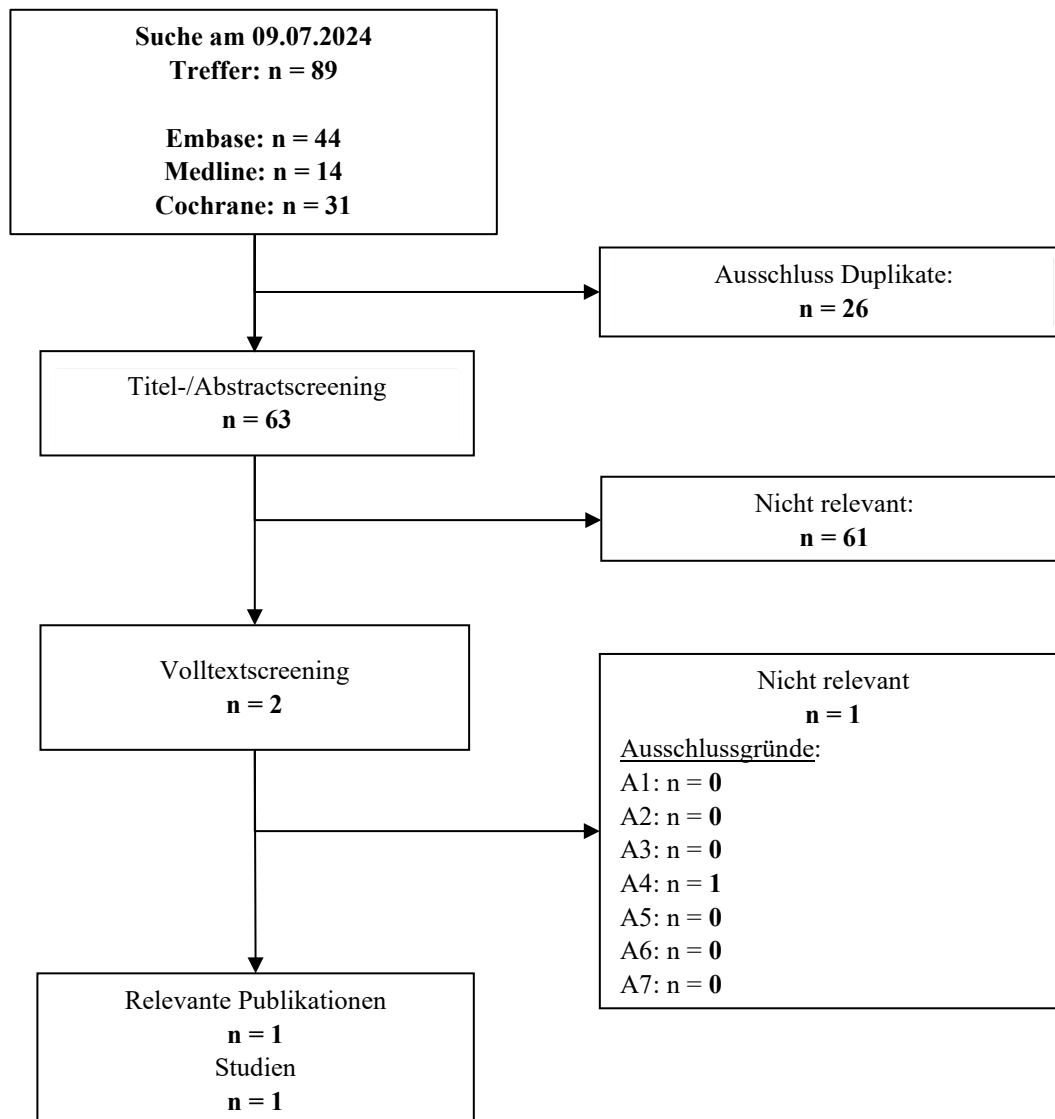


Abbildung 3: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel
FK: Funktionsklasse; WHO: World Health Organization (Weltgesundheitsorganisation)

4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern/ Studienergebnisdatenbanken identifiziert wurden. Geben Sie dabei an, in welchem Studienregister / Studienergebnisdatenbank 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 4-8) 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.

Die Suche erfolgte am 09.07.2024 in den in Abschnitt 4.2.3.3 genannten, öffentlich zugänglichen Studienregistern/Studienregisterdatenbanken. Die identifizierten Studien wurden von zwei Personen unabhängig voneinander gemäß den definierten Einschlusskriterien in Abschnitt 4.2.2 (Tabelle 4-3) auf ihre Relevanz bewertet.

Es wurde eine relevante Studie identifiziert (siehe Tabelle 4-10). Über das Suchportal der EMA sowie der AMIce-Datenbank konnten bei der Suche am 09.07.2024 keine zusätzlichen bewertungsrelevanten Dokumente als die bereits vorliegenden Registereinträge, die Publikation sowie der Studienbericht identifiziert werden.

Die Suchstrategien samt Trefferzahlen sind in Anhang 4-B1 angegeben. Die ausgeschlossenen Studien sind in Anhang 4-D1 unter Angabe des jeweiligen Ausschlussgrundes aufgeführt.

Tabelle 4-10: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern/ Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel

Studie	Identifikationsorte (Name des Studienregisters/der Studienergebnisdatenbank 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)
STELLAR ^b	ClinicalTrials.gov: NCT04576988 (32) EU-CTR: 2020-004142-11 (33) ICTRP: NCT04576988 (34) EUCTR2020-004142-11 (35)	ja	ja	abgeschlossen
<p>a: Zitat des Studienregistereintrags, die Studienregisternummer (NCT-Nummer, EudraCT-Nummer) sowie, falls vorhanden, der im Studienregister/in der Studienergebnisdatenbank aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p> <p>b: Neben der Studie STELLAR als RCT wird zusätzlich die nicht randomisierte, nicht vergleichende Long-Term Extension Studie SOTERIA herangezogen. Die SOTERIA-Studie ist eine laufende Studie, die zuvor in anderen Sotatercept-Studien eingeschlossenen Patient:innen die Möglichkeit gibt, ihre Behandlung mit Sotatercept weiterzuführen oder zu initiieren. Für die vorliegende Nutzenbewertung wird die SOTERIA-Studie unterstützend für den Endpunkt Mortalität herangezogen.</p> <p>EU-CTR: EU Clinical Trials Register; EudraCT: European Union Drug Regulating Authorities Clinical Trials; ICTRP: International Clinical Trials Registry Platform; NCT: National Clinical Trial; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)</p>				

Geben Sie an, welchen Stand die Information in Tabelle 4-10 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.

Der Stand der Informationen in Tabelle 4-10 ist der 09.07.2024.

4.3.1.1.4 Studien aus der Suche auf der Internetseite des G-BA

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Sichtung der Internetseite des G-BA identifiziert wurden. Geben Sie dabei an, welche Dokumente dort hinterlegt sind (z. B. Dossier eines anderen pharmazeutischen Unternehmers, IQWiG Nutzenbewertung). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-8) und ob die Studie auch durch die bibliografische Literaturrecherche bzw. Suche in Studienregistern/ Studienergebnisdatenbank identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein.

[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 4-11: Relevante Studien aus der Suche auf der Internetseite des G-BA – RCT mit dem zu bewertenden Arzneimittel

Studie	Relevante Quellen ^a	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Studie durch Suche in Studienregistern/ Studienergebnisdatenbanken identifiziert (ja/nein)
Es wurden keine relevanten Studien auf der Internetseite des G-BA identifiziert.				
a: Quellen aus der Suche auf der Internetseite des G-BA G-BA: Gemeinsamer Bundesausschuss; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)				

Geben Sie an, welchen Stand die Information in Tabelle 4-11 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.

Der Stand der Informationen in Tabelle 4-11 ist der 09.07.2024.

4.3.1.1.5 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, 4.3.1.1.3 und 4.3.1.1.4) 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 aller verfügbaren Quellen in Abschnitt 4.3.1.4 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 4-12: 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)	Studienberichte (ja/nein [Zitat])	Register-einträge ^c (ja/nein [Zitat])	Publikation und sonstige Quellen ^d (ja/nein [Zitat])
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
placebokontrolliert						
STELLAR ^e	ja	ja	nein	ja (36)	ja (32-35)	ja (31)
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						
Nicht zutreffend						
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregistereinträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.6 genannte Referenzliste. Darüber hinaus ist darauf zu achten, dass alle Quellen, auf die in dieser Tabelle verwiesen wird, auch in Abschnitt 4.3.1.4 (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> <p>d: Sonstige Quellen: Dokumente aus der Suche auf der Internetseite des G-BA.</p> <p>e: Neben der Studie STELLAR als RCT wird zusätzlich die nicht randomisierte, nicht vergleichende Long-Term Extension Studie SOTERIA herangezogen. Die SOTERIA-Studie ist eine laufende Studie, die zuvor in anderen Sotatercept-Studien eingeschlossenen Patient:innen die Möglichkeit gibt, ihre Behandlung mit Sotatercept weiterzuführen oder zu initiieren. Für die vorliegende Nutzenbewertung wird die SOTERIA-Studie unterstützend für den Endpunkt Mortalität herangezogen.</p> <p>G-BA: Gemeinsamer Bundesausschuss; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)</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. Falls Teilpopulationen berücksichtigt werden, ist die Charakterisierung der Studienpopulation auch für diese Teilpopulation durchzuführen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Geben Sie bei den Datenschnitten auch den Anlass des Datenschnittes an. Fügen Sie für jede Studie eine neue Zeile ein.

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

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-13: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	Studiendesign <RCT, doppelblind/einfach, verblindet/offen, parallel/cross-over etc.>	Population <relevante Charakteristika, z. B. Schweregrad>	Interventionen (Zahl der randomisierten Patienten)	Studiendauer/ Datenschnitte <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
STELLAR	RCT, doppelblind, parallel, multizentrisch, Phase III	Erwachsene Patient:innen mit PAH (WHO-Gruppe I) mit WHO-FK II oder III	Sotatercept + PAH- Hintergrundtherapie (n = 163) Placebo + PAH- Hintergrundtherapie (n = 160)	<u>Screening:</u> Bis zu 4 Wochen vor Studienbeginn <u>Behandlung:</u> 24 Wochen DBPC- Phase und bis zu 72 Wochen LTDB- Phase, beide Phasen wurden doppelblind durchgeführt. <u>Nachbeobachtung:</u> Mindestens 8 Wochen (mindestens zwei Visiten für Patient:innen, welche die Studie vorzeitig abgebrochen haben oder nicht in die Long-Term Extension-Studie SOTERIA wechselten)	<u>91 Studienzentren in 21 Ländern:</u> Argentinien (2) Australien (4) Belgien (2) Brasilien (3) Kanada (2) Tschechische Republik (2) Frankreich (10) Deutschland (9) Israel (4) Italien (1) Korea (1) Mexiko (2) Niederlande (2) Neuseeland (2) Polen (2) Serbien (1) Spanien (5) Schweden (2) Schweiz (1) Vereinigtes Königreich (3)	<u>Primäre Endpunkte:</u> 6MWD (Veränderung gegenüber Baseline zu Woche 24), <u>Sekundäre Endpunkte:</u> TTCW, French Risk Score, Symptom- und Impact-Items des PAH-SYMPACT <u>Explorative patientenrelevante Endpunkte:</u> Dyspnoe gemäß Borg CR10-Skala EQ-5D VAS UE

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

	<u>Datenschnitte:</u>	Vereinigte Staaten von Amerika (31)
	26.08.2022 (Zulassungs- begründender Datenschnitt)	<u>Studienperiode:</u>
	06.12.2022 (Finaler Datenschnitt)	25.01.2021- 06.12.2022
<p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); CR10-Skala: 10 Point Category Ratio Scale; DBPC: Double-Blind Placebo-Controlled (doppelblind Placebo-kontrolliert); EQ-5D: EuroQoL-5 Dimensions; FK: Funktionsklasse; LTDB: Long-Term Double-Blind; PAH: Pulmonale arterielle Hypertonie; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact; PVR: Pulmonary Vascular Resistance; RCT: Randomisierte kontrollierte Studie; TTCW: Zeit bis zur klinischen Verschlechterung oder Tod; UE: Unerwünschtes Ereignis; VAS: Visual Analog Scale (visuelle Analogskala); WHO: World Health Organization (Weltgesundheitsorganisation)</p>		

Tabelle 4-14: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	Sotatercept	Placebo	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in-Phase etc.</i>
STELLAR	Initialdosis 0,3 mg/kg Körpergewicht als subkutane Injektion, ab Woche 3 0,7 mg/kg Körpergewicht als subkutane Injektion alle 21 Tage + PAH- Hintergrundtherapie ^a	Subkutane Injektion alle 21 Tage + PAH- Hintergrundtherapie ^a	<u>Begleittherapien:</u> Während der Screening-Phase sowie während der Studie war es den Patient:innen erlaubt, Begleitmedikationen für chronische Begleiterkrankungen zu nehmen. Eine erforderliche Dosisanpassung oder der Bedarf einer neuen Therapie aufgrund einer neuen oder der Verschlechterung einer bereits existierenden Begleiterkrankung erfolgte nach Ermessen der Prüfärzt:innen. Die Weiterführung einer kardiopulmonalen Rehabilitation im Rahmen eines Übungsprogrammes entsprechend einer Physiotherapie war als Begleittherapie erlaubt.
a: Mono- oder Kombinationstherapie bestehend aus Endothelin-Rezeptor-Antagonisten, Phosphodiesterase-Typ-5-Inhibitoren, Stimulator der löslichen Guanylatzyklase, Prostazyklin-Analoga, Selektive Prostazyklin-Rezeptor-Agonisten kg: Kilogramm; mg: Milligramm; PAH: Pulmonale arterielle Hypertonie; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)			

Tabelle 4-15: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Charakteristikum	Studie: STELLAR (MK-7962-003) ^a	
	Sotatercept N ^b = 163	Placebo N ^b = 160
Geschlecht, n (%)		
Männlich	34 (20,9)	33 (20,6)
Weiblich	129 (79,1)	127 (79,4)
Alter (Jahre)		
Mittelwert (SD)	47,6 (14,1)	48,3 (15,5)
Median (Q1; Q3)	47,0 (37,0; 58,0)	49,0 (35,0; 60,5)
Min, Max	18,0; 81,0	18,0; 82,0
Altersgruppe, n (%)		
< 65 Jahre	138 (84,7)	131 (81,9)
≥ 65 Jahre	25 (15,3)	29 (18,1)
WHO-Funktionsklasse, n (%)		
Klasse II	79 (48,5)	78 (48,8)
Klasse III	84 (51,5)	82 (51,3)
Region, n (%)^c		
WHO Stratum A	135 (82,8)	127 (79,4)
Rest der Welt	28 (17,2)	33 (20,6)

Charakteristikum	Studie: STELLAR (MK-7962-003) ^a	
	Sotatercept N ^b = 163	Placebo N ^b = 160
Art der PAH-Hintergrundtherapie, n (%)		
Monotherapie	9 (5,5)	4 (2,5)
Zweifachtherapie	56 (34,4)	56 (35,0)
Dreifachtherapie	98 (60,1)	100 (62,5)
WHO Diagnose pulmonare Hypertonie Gruppe I: PAH, n (%)		
Idiopathische PAH	83 (50,9)	106 (66,3)
Heritable PAH	35 (21,5)	24 (15,0)
Arzneimittel/Toxin-induzierte PAH	7 (4,3)	4 (2,5)
PAH im Zusammenhang mit Bindegewebserkrankungen	29 (17,8)	19 (11,9)
PAH im Zusammenhang mit (einfachen) angeborenen systemisch-pulmonalen Shunts, mindestens (frühestens) ein Jahr nach Korrektur, z.B. nach Shuntkorrektur	9 (5,5)	7 (4,4)
Intravenöse Prostazyklin-Therapie, n (%)		
Ja	65 (39,9)	65 (40,6)
Nein	98 (60,1)	95 (59,4)
Baseline PVR (dynes*sec/cm⁵)		
Mittelwert (SD)	781,3 (398,5)	745,8 (313,5)
Median (Q1; Q3)	624,0 (504,0; 904,0)	684,0 (508,0; 908,0)
Min, Max	400,0; 2688,0	400,0; 2624,0
Abstammung, n (%)		
Weiß	147 (90,2)	141 (88,1)
Schwarz oder Afroamerikaner	2 (1,2)	5 (3,1)
Asiatisch	1 (0,6)	6 (3,8)
Indigene Bevölkerung Nordamerikas oder indigene Bevölkerung Alaskas	0 (0,0)	1 (0,6)
Indigene Bevölkerung Hawaiis oder anderer pazifischer Inseln	0 (0,0)	1 (0,6)
Mehrere	2 (1,2)	0 (0,0)
Andere	5 (3,1)	4 (2,5)
Fehlend	6 (3,7)	2 (1,3)
Ethnie, n (%)		
Hispanisch oder Latino	27 (16,6)	31 (19,4)
Nicht Hispanisch oder Latino	132 (81,0)	124 (77,5)
Nicht berichtet	4 (2,5)	5 (3,1)
Körpergröße (cm)		
Mittelwert (SD)	165,1 (8,9)	165,3 (9,1)
Median (Q1; Q3)	165,0 (160,0; 170,0)	165,0 (158,8; 170,0)
Min, Max	144,0; 193,0	147,3; 195,0
Körpergewicht (kg)		
Mittelwert (SD)	71,8 (19,0)	72,8 (17,6)
Median (Q1; Q3)	67,0 (58,1; 80,2)	70,2 (61,5; 82,1)
Min, Max	39,6; 135,0	38,0; 141,3
BMI (kg/m²)		
Mittelwert (SD)	26,1 (5,7)	26,6 (6,1)
Median (Q1; Q3)	25,3 (22,2; 28,9)	25,7 (22,3; 29,8)
Min, Max	15,1; 50,3	15,0; 54,2
a: Datenschnitt: 06. Dezember 2022		
b: Anzahl der Patient:innen: Full-Analysis-Set		
c: WHO Stratum A umfasst folgende Länder: Andorra, Australien, Österreich, Belgien, Brunei Darussalam, Kanada, Kroatien, Kuba, Zypern, Tschechische Republik, Dänemark, Finnland, Frankreich, Deutschland, Griechenland, Island, Israel, Italien, Japan, Luxemburg, Malta, Monaco, Niederlande, Neuseeland, Norwegen, Portugal, San Marino, Singapur, Slowenien, Spanien, Schweden, Schweiz, Vereinigtes Königreich und Vereinigte Staaten von Amerika		
BMI: Body Mass Index; PAH: Pulmonale arterielle Hypertonie; Placeholder; PVR: Pulmonary Vascular Resistance (pulmonaler		

Charakteristikum	Studie: STELLAR (MK-7962-003) ^a	
	Sotatercept N ^b = 163	Placebo N ^b = 160
Gefäßwiderstand); Q1: Erstes Quartil; Q3: Drittes Quartil; WHO FC: WHO-Funktionsklasse; WHO: World Health Organization		

Beschreiben Sie die Studien zusammenfassend. In der Beschreibung der Studien sollten Informationen zur Behandlungsdauer sowie zu geplanter und tatsächlicher Beobachtungsdauer enthalten sein. Sofern sich die Beobachtungsdauer zwischen den relevanten Endpunkten unterscheidet, sind diese unterschiedlichen Beobachtungsdauern endpunktbezogen anzugeben. Beschreiben Sie zudem, ob und aus welchem Anlass verschiedene Datenschnitte durchgeführt wurden oder noch geplant sind. Geben Sie dabei auch an, ob diese Datenschnitte jeweils vorab (d.h. im statistischen Analyseplan) geplant waren. In der Regel ist nur die Darstellung von a priori geplanten oder von Zulassungsbehörden geforderten Datenschnitten erforderlich. 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.

Studiendesign

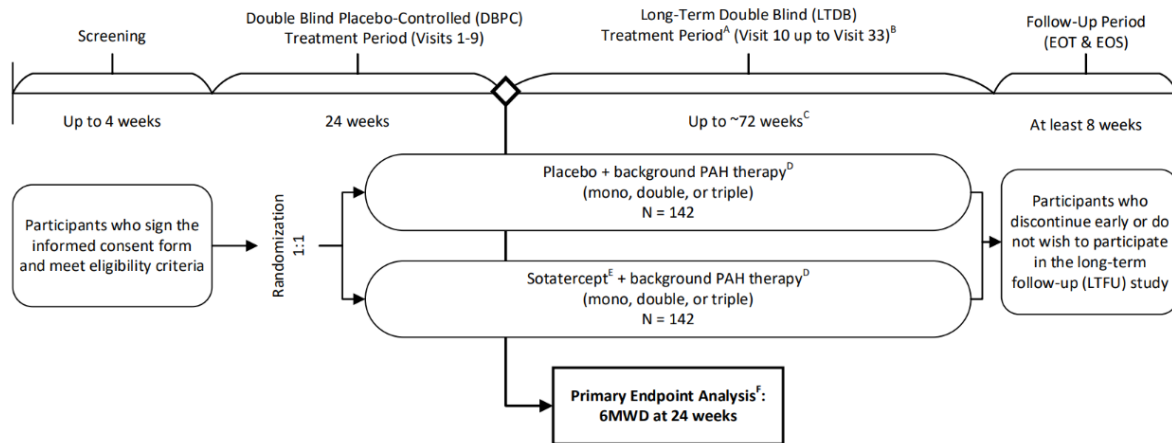
Für die vorliegende Nutzenbewertung wurde die Studie STELLAR identifiziert.

Bei der Studie STELLAR handelt es sich um eine randomisierte, multizentrische, doppelblinde Phase-III-Studie, zur Untersuchung der Wirksamkeit von Sotatercept zusätzlich zur PAH-Hintergrundtherapie gegenüber Placebo zusätzlich zur PAH-Hintergrundtherapie für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit.

Die Randomisierung erfolgte stratifiziert nach Baseline WHO-FK (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifachtherapie vs. Dreifachtherapie).

Die Studie bestand aus zwei konsekutiven Behandlungsphasen: die erste Phase war die 24 Wochen lange, doppelblinde und Placebo-kontrollierte Behandlungsphase (Double-Blind Placebo-Controlled, DBPC). Ihr folgte eine langfristige doppelblinde Behandlungsphase (Long-Term Double-Blind, LTDB). Patient:innen welche die DBPC-Phase durchlaufen haben, führten die Studie in der LTDB-Phase weiter, bis der/die letzte randomisierte Patient:in die 24 Wochen lange DBPC-Phase abgeschlossen hatte. Patient:innen konnten im Rahmen der offenen und nicht vergleichenden LTE-Studie SOTERIA ihre Therapie mit Sotatercept weiterführen oder aus dem Kontrollarm die Therapie mit Sotatercept initiieren. Der Übergang in die LTE-Studie SOTERIA war erst nach Abschluss der Studie STELLAR vorgesehen, konnte aber nach einer klinischen Verschlechterung auch früher erfolgen, siehe Abschnitt 4.2.5.4.

Darüber hinaus umfasst die Studie STELLAR eine Follow-Up-Phase (mindestens 8 Wochen mit mindestens zwei Visiten für Patient:innen, welche die Studie vorzeitig abgebrochen haben oder nicht in die Long-Term Extension-Studie SOTERIA wechselten).



^A During the LTDB treatment period, select study visits may be performed as home health care (HHC) visits.

^B LTDB treatment period will last until the last participant randomized completes the DBPC treatment period, at which point the study will be unblinded and participants may rollover into the LTFU study

^C LTDB treatment period duration is estimated based on projected enrollment duration and time required for the last participant to complete the DBPC treatment period.

^D Background PAH therapy refers to approved PAH-specific medications and may consist of monotherapy or combination therapy with ERA, PDE5 inhibitors, soluble guanylate cyclase stimulators, and/or prostacyclin analogues or receptor agonists.

^E Sotatercept at a starting dose of 0.3 mg/kg SC with a target dose of 0.7 mg/kg SC

^F Primary endpoint analysis will be completed after the last participant randomized completes the DBPC treatment period.

Abbildung 4: Schematische Darstellung des Studiendesigns - Studie STELLAR (Quelle: CSP)

6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); EOS: Studienende (end of study); EOT: Behandlungsende (end of treatment); PAH: Pulmonale arterielle Hypertonie

Behandlungsdauer und Beobachtungsdauer

In der Studie STELLAR haben im Interventionsarm von insgesamt 163 Patient:innen 155 Patient:innen (95,1 %) die Studie wie vorhergesehen beendet. Acht Patient:innen (4,9 %) haben die Studie vorzeitig beendet, davon drei Patient:innen aufgrund eines unerwünschten Ereignisses, zwei Patient:innen aufgrund von Tod und ein(e) Patient:in aufgrund von Nicht-einhaltung des Studienprotokolls. Zwei Patient:innen haben ihre Einverständniserklärung zurückgezogen. Im Kontrollarm haben von insgesamt 160 Patient:innen 142 Patient:innen (88,8 %) die Studie abgeschlossen. 18 Patient:innen (11,3 %) haben die Studie nicht abgeschlossen, davon ein(e) Patient:in aufgrund eines unerwünschten Ereignisses, sechs Patient:innen aufgrund von Tod, zwei Patient:innen aufgrund von Krankheitsprogression und zwei Patient:innen aufgrund von Nicht-einhaltung des Studienprotokolls. Sechs Patient:innen haben ihre Einverständniserklärung zurückgezogen und ein(e) Patient:in wegen sonstigen Gründen.

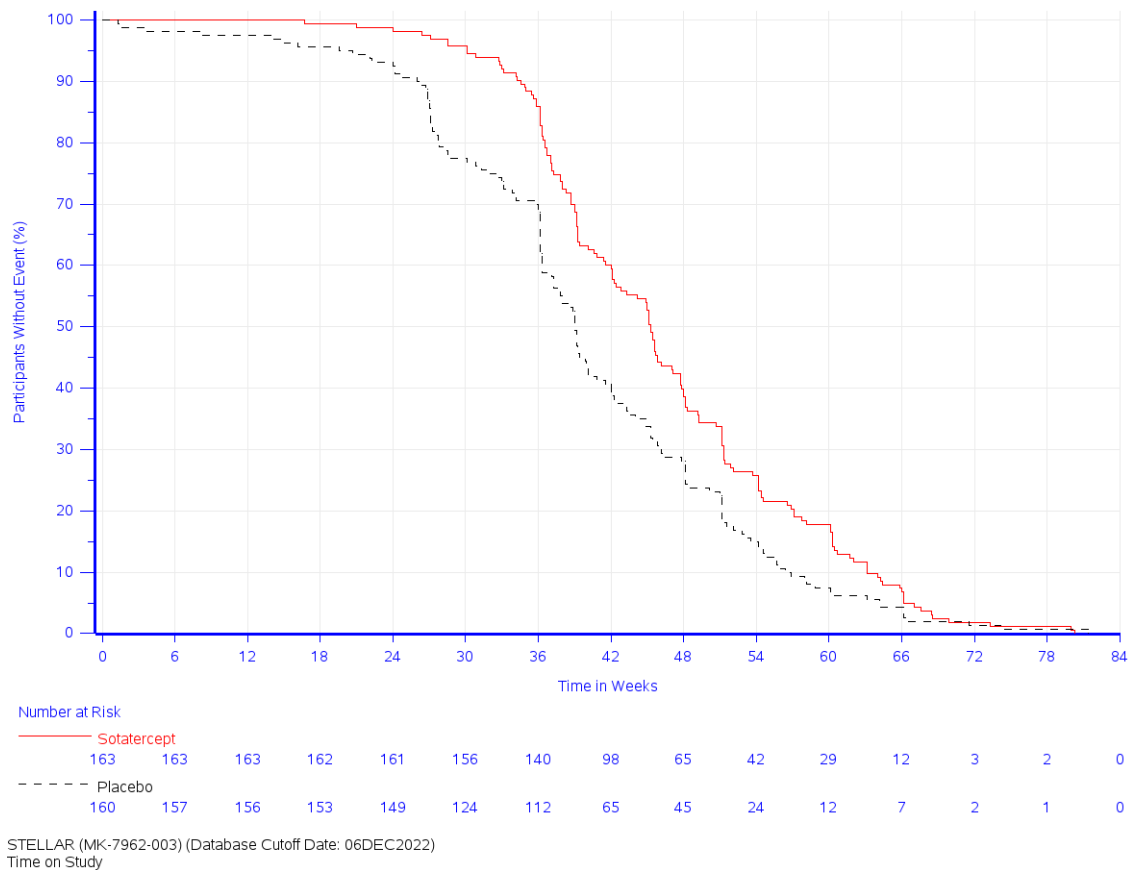


Abbildung 5: Kaplan-Meier-Kurve für die Zeit bis Studienende (STELLAR)

Auffällig ist, dass die Anzahl an Patient:innen in der Studie STELLAR aus dem Kontrollarm deutlich schneller abnimmt als die Anzahl an Patient:innen im Interventionsarm. Dies ist neben den oben genannten Gründen damit zu begründen, dass die Patient:innen aus dem Kontrollarm früher in die LTE-Studie SOTERIA wechselten (siehe hierzu Abbildung 5) und sich dadurch die Behandlungs- und Beobachtungszeiten nach Abschluss der DBPC-Phase unterscheiden.

Die Behandlungs- und Beobachtungsdauer der Endpunkte unterscheidet sich daher zwischen den Studienarmen, wobei bis Woche 24 sichergestellt ist, dass kein frühzeitiger Abbruch aufgrund einer klinischen Verschlechterung erfolgte. Die mediane Behandlungsdauer betrug 44,7 Wochen im Sotatercept-Arm und 39 Wochen im Placebo-Arm.

Die mediane Beobachtungsdauer des Endpunkts Mortalität sowie des Endpunkts Zeit bis zur klinischen Verschlechterung oder Tod betrug 45,3 Wochen im Interventionsarm und 39 Wochen im Kontrollarm.

Die mediane Beobachtungsdauer des Endpunkts Verbesserung der 6MWD betrug 45,1 Wochen im Interventionsarm und 39 Wochen im Kontrollarm.

Die mediane Beobachtungsdauer des Endpunkts Verbesserung der Dyspnoe gemäß Borg CR10-Skala betrug 45,1 Wochen im Interventionsarm und 30 Wochen im Kontrollarm.

Die mediane Beobachtungsdauer für die EQ-5D VAS betrug 39,3 Wochen im Interventionsarm und 37,3 Wochen im Placebo-Arm.

Die mediane Beobachtungsdauer für die Domänen des PAH-SYMPACT® betrug 42,1 Wochen im Interventionsarm und 38,9 Wochen im Kontrollarm.

Entsprechend der längeren Behandlungsdauer war die mediane Beobachtungsdauer mit 45,1 Wochen bei den unerwünschten Ereignissen und 49,1 Wochen bei den schwerwiegenden unerwünschten Ereignissen im Interventionsarm länger im Vergleich zum Kontrollarm mit 39 Wochen bei den unerwünschten Ereignissen und 35,9 Wochen bei den schwerwiegenden unerwünschten Ereignissen.

Datenschnitte

Es wurden zwei Datenschnitte präspezifiziert: der erste, zulassungsbegründende Datenschnitt erfolgte am 26. August 2022, als der/die letzte Patient:in die Visite zu Woche 24 abgeschlossen hatte. Der finale Datenschnitt erfolgte am 06. Dezember 2022. Der finale Datenschnitt beinhaltete eine längere Nachbeobachtungszeit der Patient:innen sowie Korrekturen in der Datenbank. Die Grundlage der vorliegenden Nutzenbewertung bildet der finale Datenschnitt (06. Dezember 2022).

Patientencharakteristika

Das Patientenkollektiv der Studie STELLAR bildet das zugelassene Anwendungsgebiet ab. Die Dosierung des zu bewertenden Arzneimittels erfolgte zulassungskonform. Die Hintergrundtherapie mit einer Mono- oder Kombinationstherapie bestehend aus Endothelin-Rezeptor-Antagonisten, Phosphodiesterase-Typ-5-Inhibitoren, Stimulator der löslichen Guanylatzyklase, Prostazyklin-Analoga sowie selektive Prostazyklin-Rezeptor-Agonisten bildet den deutschen Versorgungskontext und damit in Verbindung mit der Gabe von Placebo die zVT ab. Daher wird die gesamte Studienpopulation der Studie STELLAR als Grundlage für den Nachweis des medizinischen Zusatznutzens im vorliegenden Anwendungsgebiet herangezogen.

Insgesamt wurden in der Studie STELLAR 323 Patient:innen randomisiert, davon 163 Patient:innen (20,9 % Männer und 79,1 % Frauen) im Interventionsarm und 160 Patient:innen (20,6 % Männer und 79,4 % Frauen) im Vergleichsarm. In Bezug auf die Region waren die Patient:innen ebenfalls vergleichbar (WHO Stratum A 82,8 % im Sotatercept-Arm und 79,4 % im Placebo-Arm). Hinsichtlich der WHO-FK besteht ebenfalls kein Unterschied zwischen den Behandlungsarmen. Die meisten Patient:innen erhielten eine Dreifachtherapie (60,1 % im Sotatercept-Arm und 62,5 % im Placebo-Arm), gefolgt von der Zweifachtherapie mit 34,4 % im Sotatercept-Arm und 35 % im Placebo-Arm. Nur wenige Patient:innen erhielten eine Monotherapie (5,5 % im Sotatercept-Arm und 2,5 % im Placebo-Arm). Die Mehrheit der STELLAR Patient:innen wies eine idiopathische PAH auf (50,9 % im Sotatercept-Arm und 66,3 % im Placebo-Arm).

Auch für die weiteren untersuchten Patientenmerkmale zeigen sich keine wesentlichen Unterschiede zwischen den beiden Behandlungsgruppen (siehe Tabelle 4-15). Eine Verzerrung der Ergebnisse durch ungleiche Verteilung von Patientinnen unterschiedlicher Merkmalsausprägung in die beiden Behandlungsgruppen kann daher ausgeschlossen werden.

Übertragbarkeit auf den deutschen Versorgungskontext

Mit 91 Studienzentren in 21 Ländern handelt es sich bei der Studie STELLAR um eine internationale und multizentrische Studie. In Europa wurden Patient:innen in 40 Zentren eingeschlossen. Neun Zentren in Deutschland haben 71 Patient:innen eingeschlossen, dies entspricht 22% der Patient:innen der STELLAR Studie.

Die deutliche Mehrheit der Patient:innen ist weiß (90,2 % im Sotatercept-Arm und 88,1 % im Placebo-Arm). PAH betrifft weitaus häufiger Frauen, dementsprechend sind in der Studie mehrheitlich Frauen eingeschlossen. Die Patient:innen sind bis zu 40 Jahre an PAH erkrankt, die mediane Erkrankungsdauer beträgt 7,26 Jahre (CSR Final REF). Dies entspricht der großen Bandbreite, die sich auch bei deutschen PAH-Patient:innen zeigt, die für eine Behandlung mit Sotatercept gemäß Zulassung in Frage kommen. Insgesamt entsprechen die krankheitsspezifischen Charakteristika der Studienpopulation der deutschen Zielpopulation.

Die Erhebung der Endpunkte erfolgte standardisiert und mit niedrigem Verzerrungspotential.

Somit ist davon auszugehen, dass die in der Studie STELLAR beobachteten Behandlungseffekte 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 4-16: 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			
STELLAR	ja	ja	ja	ja	ja	ja	niedrig
RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)							

Begründen Sie für jede Studie die abschließende Einschätzung.

Die Bewertung des Verzerrungspotenzials der Studie STELLAR basiert auf den Ergebnissen der unter Anhang 4-F durchgeführten und durch Quellen belegten Einschätzung. Die Erzeugung der Randomisierungssequenz wurde in der Studie STELLAR adäquat durchgeführt, die Gruppenzuteilung fand verdeckt statt. In dieser doppelblinden Studie waren Patient:innen und behandelnde Personen verblindet. Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige, das Verzerrungspotenzial beeinflussende Faktoren wurden nicht identifiziert.

Somit wird das Verzerrungspotenzial für die Studie STELLAR auf Studienebene als niedrig eingestuft.

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 4-17: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Studie	Mortalität ^a	Morbidität ^b	Gesundheitsbezogene Lebensqualität ^c	Nebenwirkungen ^d
STELLAR	ja	ja	ja	ja
a: Mortalität zu Woche 24				
b: Verbesserung der 6MWD, Zeit bis zur klinischen Verschlechterung oder Tod, Verbesserung der Dyspnoe gemäß Borg CR10-Skala, Krankheitssymptomatik (PAH-SYMPACT) und Gesundheitszustand (EQ-5D VAS)				
c: Gesundheitsbezogene Lebensqualität (PAH-SYMPACT)				

d: Unerwünschte Ereignisse Gesamtraten, Unerwünschte Ereignisse (gegliedert nach SOC und PT) und Unerwünschte Ereignisse von Interesse

6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); CR10-Skala: 10 Point Category Ratio Scale;
EQ-5D: EuroQoL-5 Dimensions; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact;
RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie); VAS: Visual Analog Scale (visuelle Analogskala)

4.3.1.3.1 Endpunkte – 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 inkl. Angaben zur Häufigkeit von und zum Umgang mit nicht oder nicht vollständig beobachteten Patienten (bei Verlaufsbeobachtungen pro Messzeitpunkt)
- 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.

Unterschiedliche Beobachtungszeiten zwischen den Behandlungsgruppen sollen durch adäquate Analysen (z.B. Überlebenszeitanalysen) adressiert werden, und zwar für alle Endpunkte (einschließlich UE nach den nachfolgend genannten Kriterien), für die eine solche Analyse aufgrund deutlich unterschiedlicher Beobachtungszeiten erforderlich ist.

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden. Dabei ist für jeden Endpunkt, für den eine solche Analyse durchgeführt wird, eine separate Kaplan-Meier-Kurve darzustellen.

Zu mit Skalen erhobenen patientenberichteten Endpunkten (z.B. zur gesundheitsbezogenen Lebensqualität oder zu Symptomen) sind immer auch die Werte im Studienverlauf anzugeben, auch als grafische Darstellung, sowie eine Auswertung, die die über den Studienverlauf ermittelten Informationen vollständig berücksichtigt (z.B. als Symptomlast über die Zeit, geschätzt mittels MMRM-Analyse [falls aufgrund der Datenlage geeignet]). Die Auswertung von Responderanalysen mittels klinischer Relevanzschwellen bei komplexen Skalen soll nach dem folgenden Vorgehen erfolgen:

1. Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten

Erhebungsinstruments entspricht, sind diese Responderanalysen für die Bewertung darzustellen.

2. Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, bestehen in diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, aber stattdessen Analysen kontinuierlicher Daten zur Verfügung stehen, verschiedene Möglichkeiten. Entweder können post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % der Skalenspannweite dargestellt werden. Alternativ können Analysen der kontinuierlichen Daten dargestellt werden, für die Relevanzbewertung ist dabei auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückzugreifen. Dabei ist eine Irrelevanzschwelle als Intervall von -0,2 bis 0,2 zu verwenden: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig außerhalb dieses Irrelevanzbereichs, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als klein angesehen werden kann.

3. Liegen sowohl geeignete Responderanalysen (Responsekriterium präspezifiziert mindestens 15 % der Skalenspannweite oder post hoc genau 15 % der Skalenspannweite) als auch Analysen stetiger Daten vor, sind die Responderanalysen darzustellen.

Zu unerwünschten Ereignissen (UE) sind folgende Auswertungen vorzulegen:

1. Gesamtrate UE,
2. Gesamtrate schwerwiegender UE (SUE),
3. Gesamtrate der Abbrüche wegen UE,
4. Gesamtraten von UE differenziert nach Schweregrad, sofern dies in der/den relevante/n Studie/n erhoben wurde (z.B. gemäß CTCAE und/oder einer anderen etablierten bzw. validierten indikationsspezifischen Klassifikation) einschließlich einer Abgrenzung schwerer und nicht schwerer UE,
5. zu den unter 1, 2 und 4 genannten Kategorien (UE ohne weitere Differenzierung, SUE, UE differenziert nach Schweregrad) soll zusätzlich zu den Gesamtraten die Darstellung nach Organsystemen und Einzelereignissen (als System Organ Class [SOCs] und Preferred Terms [PT] nach MedDRA) jeweils nach folgenden Kriterien erfolgen:
 - UE (unabhängig vom Schweregrad): Ereignisse, die bei mindestens 10 % der Patienten in einem Studienarm aufgetreten sind
 - Schwere UE (z.B. CTCAE-Grad ≥ 3) und SUE: Ereignisse, die bei mindestens 5% der Patienten in einem Studienarm aufgetreten sind

- zusätzlich für alle Ereignisse unabhängig vom Schweregrad: Ereignisse, die bei mindestens 10 Patienten UND bei mindestens 1 % der Patienten in einem Studienarm aufgetreten sind.

6. A priori definierte UE von besonderem Interesse [AESI]) sowie prädefinierte SOC-übergreifende UE-Auswertungen (z.B. als Standardised MedDRA Queries, SMQs) sollen unabhängig von der Ereignisrate dargestellt werden und zwar differenziert nach Schweregrad (dargestellt als Gesamtrate und differenziert nach Schweregrad, nicht schwer, schwer, schwerwiegend).

7. zu Kategorie 3: Die Abbruchgründe auf SOC/PT-Ebene müssen vollständig, jedoch nur deskriptiv dargestellt werden.

Sofern bei der Erhebung unerwünschter Ereignisse erkrankungsbezogenen Ereignisse (z. B. Progression, Exazerbation) berücksichtigt werden (diese Ereignisse also in die UE-Erhebung eingehen), sollen für die Gesamtraten (UE, schwere UE und SUE) zusätzliche UE-Analysen durchgeführt werden, bei denen diese Ereignisse unberücksichtigt bleiben. Alle Auswertungen zu UE können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine Darstellung ausschließlich in Modul 5 ist nicht ausreichend. Davon unbenommen sind die Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE), sowie die für die Gesamtaussage zum Zusatznutzen herangezogenen Ergebnisse im vorliegenden Abschnitt darzustellen.

Auswertungen zu den im Abschnitt 4.3.1.2.1 aufgeführten Datenschnitten sollen vollständig, d.h. für alle erhobenen relevanten Endpunkte, durchgeführt und vorgelegt werden. Das gilt auch dann wenn ein Datenschnitt ursprünglich nur zur Auswertung einzelner Endpunkte geplant war. Auf die Darstellung der Ergebnisse einzelner Endpunkte eines Datenschnitts bzw. eines gesamten Datenschnitts kann verzichtet werden, wenn hierdurch kein wesentlicher Informationsgewinn gegenüber einem anderen Datenschnitt zu erwarten ist (z. B. wenn die Nachbeobachtung zu einem Endpunkt bereits zum vorhergehenden Datenschnitt nahezu vollständig war oder ein Datenschnitt in unmittelbarer zeitlicher Nähe zu einem anderen Datenschnitt liegt).

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.

4.3.1.3.1 Mortalität – RCT

4.3.1.3.1.1 Mortalität zu Woche 24

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

Tabelle 4-18: Operationalisierung des Endpunkts Mortalität zu Woche 24

Studie	Operationalisierung
STELLAR	<p>Hauptanalyse:</p> <p>Die Mortalität zu Woche 24 ist definiert als Anteil an verstorbenen Patient:innen, unabhängig von der zugrunde liegenden Ursache. Die primäre Analyse betrachtet den Anteil an Patient:innen, welche innerhalb der ersten 24 Wochen verstarben. Die Ableitung des Zusatznutzens erfolgt anhand des relativen Risikos bzw. Peto-Odds Ratios.</p> <p>Für die Auswertungen werden alle Todesfälle bis zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population.</p> <p>Ergänzende Analyse:</p> <p>Zur Untersuchung, ob die in den ersten 24 Wochen beobachteten Effekte auch nachhaltig sind, wird zusätzlich zur Hauptanalyse eine Ereigniszeitanalyse zum Gesamtüberleben dargestellt.</p> <p>Für die Auswertungen wird der gesamte Beobachtungszeitraum in der Studie STELLAR herangezogen und zusätzlich die Weiterbeobachtungen für alle STELLAR-Patient:innen, die in die Long-Term-Extension-Studie SOTERIA gewechselt sind, bis zum Datenschnitt am 08. November 2023.</p> <p>Das Gesamtüberleben ist definiert als Zeitraum von der Randomisierung bis zum Tod, unabhängig von der zugrunde liegenden Ursache. Patient:innen, die im Verlauf der Studie nicht versterben, werden zum Zeitpunkt des letzten Studienkontakts oder zum Zeitpunkt des Datenschnitts zensiert, je nachdem was zuerst eintritt.</p> <p>Sensitivitätsanalyse zur ergänzenden Analyse:</p> <p>Wie in Abschnitt 4.2.5.4 beschrieben wird zusätzlich für das Gesamtüberleben ein Rank Preserving Structural Failure Time Modell dargestellt. Diese adressiert die durch den Behandlungswechsel entstehende Verzerrung.</p>
FAS: Full Analysis Set	

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 4-19: Bewertung des Verzerrungspotenzials für Mortalität zu Woche 24 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
STELLAR	niedrig	ja	ja ^a	ja	ja	niedrig
a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß Intention-To-Treat (ITT)-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen bei der Betrachtung der Ergebnisse zu Woche 24 nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Mortalität zu Woche 24 für die Studie STELLAR ist somit als niedrig einzustufen.

Stellen Sie die Ergebnisse für den Endpunkt Mortalität zu Woche 24 für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Hauptanalyse

Tabelle 4-20: Ergebnisse für Mortalität zu Woche 24 aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	N ^b	Patient:innen mit Ereignis n (%)	N ^b	Patient:innen mit Ereignis n (%)	Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
Mortalität zu Woche 24	163	0 (0,0)	160	6 (3,8)	0,13 [0,03; 0,64]	0,013	-3,76 [-7,99; -1,06]

a: Datenschnitt: 06. Dezember 2022
b: Anzahl der Patient:innen: Full-Analysis-Set
c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)
d: Zweiseitiger p-Wert (Wald-Test)
e: Miettin und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization

Für den Endpunkt Mortalität zu Woche 24 zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (Peto OR [95 %-KI]: 0,13 [0,03; 0,64]; p = 0,013), also eine erhebliche Reduktion des Mortalitätsrisikos.

Ergänzend zu dieser Analyse werden im Folgenden Auswertungen für das Gesamtüberleben über den gesamten Beobachtungszeitraum in der Studie STELLAR zuzüglich der Weiterbeobachtungen in der LTE-Studie SOTERIA dargestellt (siehe Tabelle 4-21). Die Patient:innen aus dem Kontrollarm haben ab dem Wechsel in die Studie SOTERIA Sotatercept erhalten, wechselten also von Placebo auf die aktive Therapie mit Sotatercept. Zusätzlich wird für das Gesamtüberleben ein Rank Preserving Structural Failure Time Modell dargestellt. Diese adressiert die durch den Behandlungswechsel entstehende Verzerrung.

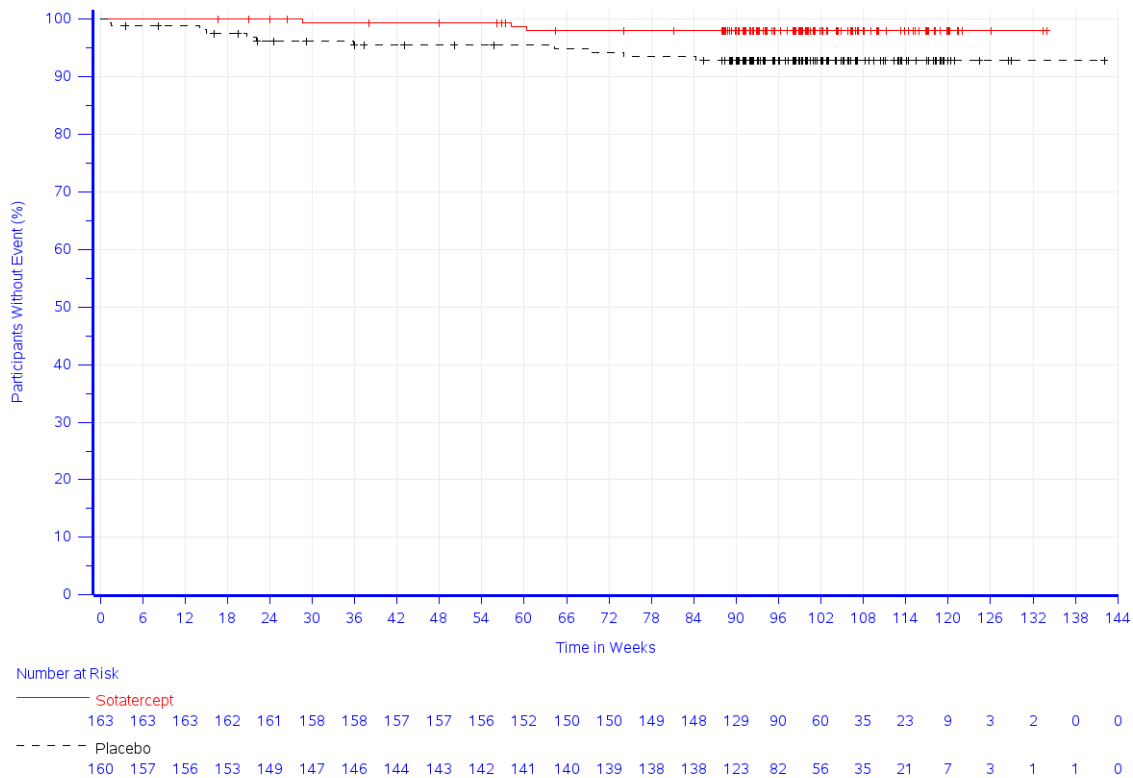
Zur Begründung für die Wahl der vorgelegten Analysen zum Gesamtüberleben siehe Abschnitt 4.2.5.4.

Ergänzende Analyse

Tabelle 4-21: Ergebnisse für Gesamtüberleben mit dem zu bewertenden Arzneimittel (Studie STELLAR und weiterführende Daten aus SOTERIA)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept			Placebo			Sotatercept vs. Placebo	
	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	Hazard Ratio [95 %-KI] ^d	p-Wert ^{d,e}
Gesamtüberleben	163	3 (1,8)	Nicht erreicht [-; -]	160	11 (6,9)	Nicht erreicht [-; -]	0,26 [0,07; 0,92]	0,037

a: Datenschnitt: 06. Dezember 2022 (STELLAR) bzw. 08. November 2023 (SOTERIA)
 b: Anzahl an Patient:innen: STELLAR Full Analysis Set Population. Die Patient:innen wurden zum Zeitpunkt des letzten Studienkontakts in STELLAR oder SOTERIA (sofern in SOTERIA eingeschlossen) zensiert ohne Adjustierung für den Behandlungswechsel von Placebo in STELLAR zu Sotatercept in SOTERIA
 c: Produkt-Limit (Kaplan-Meier) Methode für zensierte Daten
 d: Basierend auf einem Cox-Proportional-Hazards-Modell mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 e: Zweiseitiger p-Wert (Wald-Test; Score-Test im Fall von null Ereignissen in einem der Studienarme)
 KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization



Database Cutoff Date: STELLAR (MK-7962-003) - 06DEC2022/SOTERIA (MK-7962-004) - 08NOV2023
 Overall Survival
 Participants alive are censored at the date of last contact in STELLAR or SOTERIA (if enrolled in SOTERIA) without adjusting for treatment switch from placebo in STELLAR to sotatercept in SOTERIA

Abbildung 6: Überlebenszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Gesamtüberleben (Studie STELLAR und weiterführende Daten aus der LTE-Studie SOTERIA)

Tabelle 4-22: Ergebnisse für Gesamtüberleben mit dem zu bewertenden Arzneimittel (RPSFT-Methode, Studie STELLAR und weiterführende Daten aus der LTE-Studie SOTERIA)

Behandlung	N	Patient:innen mit Ereignis (%)	Personen - Wochen	Ereignisrate je 100 - Personen-Wochen (%)	Mediane Überlebenszeit ^a (Woche) [95 %-KI]	Sotatercept vs. Placebo	
						Hazard Ratio ^b [95 %-KI] ^c	p-Wert ^d
Placebo	160	11 (6,9)	14633,3	0,08	Nicht erreicht (---, ---)	0,17 (0,03, 0,90)	0,037
Placebo (RPSFT Adjustiert)	160	11 (6,9)	7571,5	0,15	Nicht erreicht (---, ---)		
Sotatercept	163	3 (1,8)	15680,4	0,02	Nicht erreicht (---, ---)		

Datenschnitt: 06. Dezember 2022 (STELLAR) bzw. 08. November 2023 (SOTERIA)

Es wurde das Rank-preserving structural failure time (RPSFT) Modell herangezogen um für den Cross-over Effekt beim Behandlungswechsel von Placebo in STELLAR zu Sotatercept in SOTERIA zu adjustieren

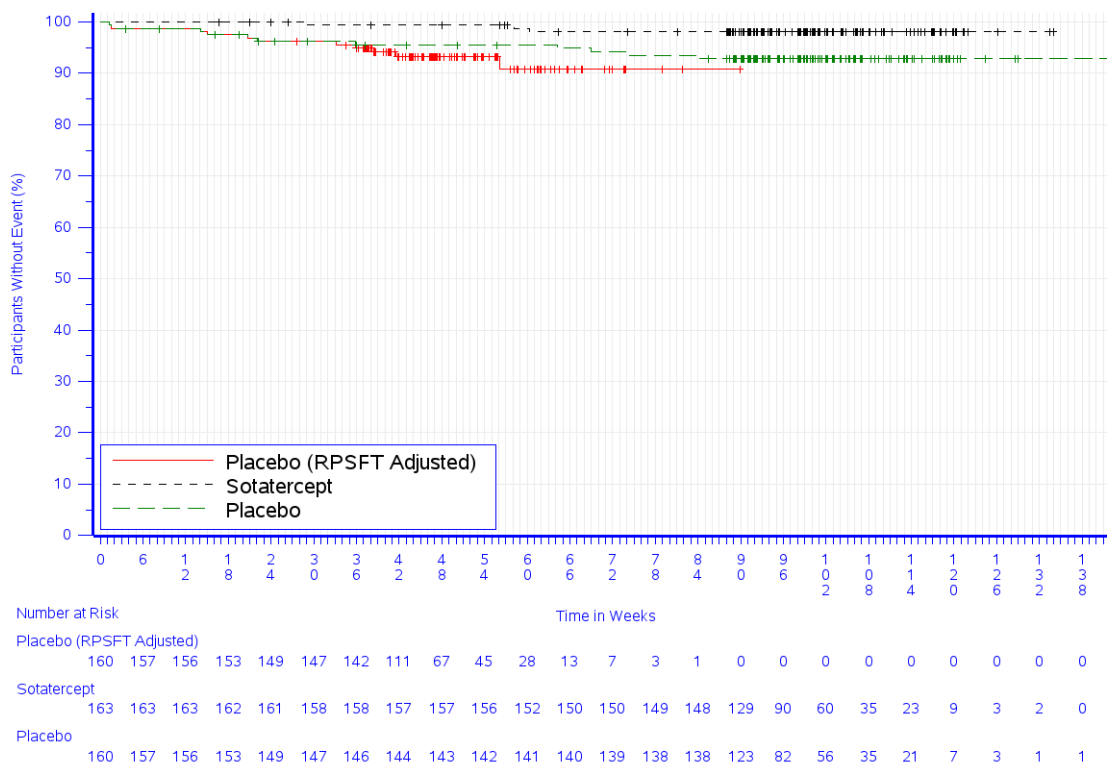
Anzahl an Patient:innen: STELLAR Full Analysis Set Population. Patient:innen welche noch am Leben waren und nicht in SOTERIA eingeschlossen wurden, wurden zum Zeitpunkt des letzten Studienkontakts in STELLAR zensiert. Patient:innen, welche sich im Sotatercept-Arm der STELLAR Studie befanden und in SOTERIA eingeschlossen wurden, wurden zum Zeitpunkt des letzten Studienkontakts in SOTERIA zensiert. Patient:innen, welche sich im Placebo-Arm der STELLAR Studie befanden und in SOTERIA eingeschlossen wurden, wurden zum Zeitpunkt des letzten Studienkontakts in SOTERIA zensiert jedoch wurde die Zeit ab Behandlungswechsel adjustiert.

a: Produkt-Limit (Kaplan-Meier) Methode für zensierte Daten

b: Basierend auf einem Cox Regressionsmodell korrigiert mittels RPSFT mit Behandlung als Kovariate stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (Klasse II vs. Klasse III) und PAH Hintergrundtherapie (Mono/Zweifachtherapie vs. Dreifachtherapie)

c: Berechnet mittels Inflation des Standardfehlers des Log-Hazard-Ratios um den auf dem Full-Analysis Set berchneten p-Wert des Cox-Models zu erhalten.

d: Zweiseitiger p-Wert mittels Wald-Test



Database Cutoff Date: STELLAR (MK-7962-003) - 06DEC2022/SOTERIA (MK-7962-004) - 08NOV2023
 Participants alive and did not enroll to SOTERIA are censored at the date of last contact in STELLAR.
 Participants in the STELLAR sotatercept arm alive and enrolled to SOTERIA are censored at the date of last contact in SOTERIA.
 Participants in the STELLAR placebo arm alive and enrolled to SOTERIA are censored at the date of last contact in SOTERIA but adjusted for the date they switched from placebo in STELLAR to sotatercept in SOTERIA

Abbildung 7: Überlebenszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Gesamtüberleben (RPSFT-Methode, Studie STELLAR und weiterführende Daten aus SOTERIA)

Die ergänzende Analyse und deren Sensitivitätsanalyse bestätigen, dass durch eine Behandlung mit Sotatercept eine bisher nicht erreichte große und auch nachhaltige und damit erhebliche Senkung des langfristigen Mortalitätsrisikos erreicht wird.

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.

Nicht zutreffend.

4.3.1.3.1.2 Morbidität – RCT

4.3.1.3.1.2.1 Verbesserung der 6MWD – RCT

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

Tabelle 4-23: Operationalisierung des Endpunkts Verbesserung der 6MWD

Studie	Operationalisierung
STELLAR	<p>Die Veränderung der 6MWD ist der primäre Endpunkt der Studie STELLAR.</p> <p>Die Verbesserung der 6MWD von Baseline zu Woche 24 wird mittels Responderanalysen unter Verwendung des Responsekriteriums von ≥ 40 Meter analysiert.</p> <p>Ergänzend dargestellt wird die mediane Differenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehmann Location Shift/Wilcoxon-Test). Dies ist die Operationalisierung des primären Endpunktes in der Studie STELLAR.</p> <p>Der 6MWD-Test wird vor Verabreichung der ersten Studienmedikation zum Zeitpunkt des Screenings (Baseline-Wert) sowie zu Woche 3, 12 und 24 erhoben. Darüber hinaus wird dieser alle 12 Wochen erhoben.</p> <p>Der 6MWD-Test wird im Gebäudeinneren entlang eines flachen und geraden Korridors mit harter Oberfläche durchgeführt. Die Länge des Korridors soll alle drei Meter Markierungen enthalten und die Wendepunkte soll. mit einem Kegel markiert sein. Eine Startlinie, die den Anfang und das Ende jeder 60-m-Runde markiert, soll auf dem Boden markiert sein. Für den Test sollen die Patient:innen komfortable Kleidung sowie geeignetes Schuhwerk tragen. Gehhilfen sollen, sofern diese von den Patient:innen üblicherweise gebraucht werden, während des Tests verwendet werden. Darüber hinaus sollten die Patient:innen ausgeruht sein und innerhalb von 2 Stunden vor Durchführung des Tests keinerlei Übungen oder Sport ausgeübt haben.</p> <p>Über alle Visiten hinweg sollte der Test ungefähr zur gleichen Uhrzeit stattfinden und entlang des gleichen Korridors durchgeführt werden. Die Patient:innen werden angehalten, bis an die Grenzen ihrer Leistungsfähigkeit zu gehen. Hierbei ist es ihnen freigestellt, zu pausieren oder das Tempo zu reduzieren, falls nötig. Eine Aufwärmphase ist nicht vorgesehen, jedoch sollen die Patient:innen mindestens für 10 Minuten kurz vor dem Start ruhen. Die verbalen Instruktionen, welche die Beobachter:innen den Patient:innen vor, während und nach dem Test geben, sind exakt vorgegeben.</p> <p>Auswertung über den Studienverlauf:</p> <p>In der vorliegenden Nutzenbewertung wird eine Auswertung über den Studienverlauf der einzelnen Erhebungszeitpunkte dargestellt. Hierzu werden zu jedem Erhebungszeitpunkt die Anzahl an verfügbaren Daten, der Mittelwert (inkl. Standardabweichung), der Median (inkl. Q1; Q3) und Minimum und Maximum zum jeweiligen Zeitpunkt deskriptiv dargestellt. Zusätzlich erfolgt eine grafische Darstellung des Mittelwerts und des Standardfehlers über den Studienverlauf. Die Auswertungen über den Studienverlauf mit dem zu bewertenden Arzneimittel werden in Anhang 4-G dargestellt.</p> <p>Umgang mit fehlenden Werten:</p> <p><u>Responderanalyse:</u></p> <p>Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p><u>Mittelwertdifferenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehman Location Shift/Wilcoxon-Test):</u></p> <p>Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde eine Verschlechterung um 2000 Meter zu Woche 24 im Vergleich zu Baseline angenommen, für Patient:innen mit fehlendem Wert aufgrund eines nicht-tödlichen Ereignisses eine Verschlechterung um 1000 Meter.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population</p>
6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); FAS: Full Analysis Set; COVID-19: Coronavirus Disease 2019	

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 4-24: Bewertung des Verzerrungspotenzials für Verbesserung der 6MWD 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
STELLAR	niedrig	ja	ja ^a	ja	ja	niedrig

a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip.
6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); FAS: Full Analysis Set; ITT: Intention-To-Treat;
RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Verbesserung der 6MWD in der Studie STELLAR ist somit als niedrig einzustufen.

Stellen Sie die Ergebnisse für den Endpunkt Verbesserung der 6MWD für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-25: Ergebnisse für Verbesserung der 6MWD aus RCT mit dem zu bewertenden Arzneimittel – Responderanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
N ^b		N ^b					
Verbesserung der 6MWD	163	72 (44,2)	159	29 (18,2)	2,40 [1,66; 3,48]	< 0,001	25,81 [15,91; 35,30]

Studie: STELLAR (MK-7962-003) ^a	Sotatercept	Placebo	Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%) N ^b	Patient:innen mit Ereignis n (%) N ^b	Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
a: Datenschnitt: 06. Dezember 2022					
b: Anzahl der Patient:innen: Full Analysis Set Population. Patient:innen mit fehlender Erhebung zu Woche 24 und nicht aufgrund von COVID-19 werden als Non-Responder gewertet. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden ausgeschlossen					
c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)					
d: Zweiseitiger p-Wert (Wald-Test)					
e: Miittinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)					
COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Tabelle 4-26: Ergebnisse für Veränderung der 6MWD aus RCT mit dem zu bewertenden Arzneimittel – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der 6-Minuten Gehstrecke (in Metern) zu Woche 24					
Sotatercept	163	34,3 (33,0, 35,5)	40,40 (27,28, 53,53)	6,70	<.001
Placebo	160	1,0 (-1,0, 3,5)			
a: Datenschnitt: 06. Dezember 2022					
b: Anzahl der Patient:innen: Full-Analysis-Set					
c: Mittelwert, Minimum und Maximum der Mediane der imputierten Daten sofern fehlende Werte imputiert wurden. Für Patient:innen mit fehlenden Werten für den 6MWD zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -2000 Metern zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -1000 Metern zugeordnet um den nächstniedrigeren Rang zu erhalten. N.a. wenn keine Imputation durchgeführt wurde.					
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen					
e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)					
6MWD: 6-Minuten Gehstrecke; ASE: Asymptotischer Standardfehler; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Für den Endpunkt Verbesserung der 6MWD zeigt sich ein statistisch signifikanter Unterschied zugunsten von Sotatercept (siehe Tabelle 4-25, (RR [95 %-KI]: 2,40 [1,66; 3,48]; $p < 0,001$). 44,2 % der Patient:innen im Sotatercept-Arm konnten ihre innerhalb von 6 Minuten zurückgelegte Gehdistanz um mindestens 40 Meter gegenüber ihrem Baseline-Wert verbessern, im Placebo-Arm lediglich 18,2 %.

Auch die Analyse zur Veränderung der 6MWD zeigt eine signifikante Verbesserung von 40,4 Metern (Hodge-Lehman Location-Shift).

Dies zeigt, dass Sotatercept die körperliche Belastbarkeit der Patient:innen, welche sich unmittelbar in alltäglichen Funktionen und Aktivitäten widerspiegelt, erheblich verbessert.

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.

Nicht zutreffend.

4.3.1.3.1.2.2 Zeit bis zur klinischen Verschlechterung oder Tod – RCT

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

Tabelle 4-27: Operationalisierung von Zeit bis zur klinischen Verschlechterung oder Tod

Studie	Operationalisierung
STELLAR	<p>Hauptanalyse:</p> <p>Der kombinierte Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod (TTCW) wird zu Woche 3, 12 sowie 24 und darüber hinaus alle 12 Wochen erhoben.</p> <p>TTCW ist definiert als Zeit von der Randomisierung bis zum ersten Auftreten eines der folgenden Mortalitäts- oder Morbiditätsereignisse:</p> <ul style="list-style-type: none"> • Tod • Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression • Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Dosierung der Infusion mit Prostazyklin um 10 % oder mehr • Notwendigkeit einer Atrioseptostomie • Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) • Verschlechterung der PAH definiert durch das Auftreten beider folgender Ereignisse, unabhängig der Reihenfolge, im Vergleich zu Baseline: <ul style="list-style-type: none"> ○ Verschlechterung der WHO-Funktionsklasse ○ Verschlechterung der 6-Minuten-Gehstrecke um ≥ 5 % bestätigt durch 2 Tests mit einem Zeitabstand von mindestens 4 Stunden, aber nicht mehr als 1 Woche <p>Sensitivitätsanalyse:</p> <p>Die Haupt- und Sensitivitätsanalyse unterscheiden sich in der Definition der Einzelkomponenten des kombinierten Endpunktes TTCW wie folgt:</p> <ul style="list-style-type: none"> • Tod • Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression • Notwendigkeit einer Atrioseptostomie • Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) • Verschlechterung der PAH definiert durch Verschlechterung der 6MWD um ≥ 40 m <p>Es wird für alle Auswertungen das jeweils zuerst auftretende Ereignis gezählt. Patient:innen ohne Ereignis werden zum Zeitpunkt der letzten Visite zensiert. Die Beurteilung der Einzelkomponenten des Endpunktes erfolgt durch die Prüffärztin/den Prüffarzt.</p> <p>Für die Auswertungen wird der gesamte Beobachtungszeitraum der Studie bis zum finalen Datenschnitt (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population.</p>
<p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); FAS: Full Analysis Set; m: Meter; PAH: Pulmonale arterielle Hypertonie; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie); TTCW: Zeit bis zur klinischen Verschlechterung oder Tod; WHO: World Health Organization (Weltgesundheitsorganisation)</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 4-28: Bewertung des Verzerrungspotenzials für Zeit bis zur klinischen Verschlechterung oder Tod in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ^a	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
STELLAR	niedrig	ja	ja	ja	ja	niedrig
a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod für die Studie STELLAR ist somit als niedrig einzustufen.

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

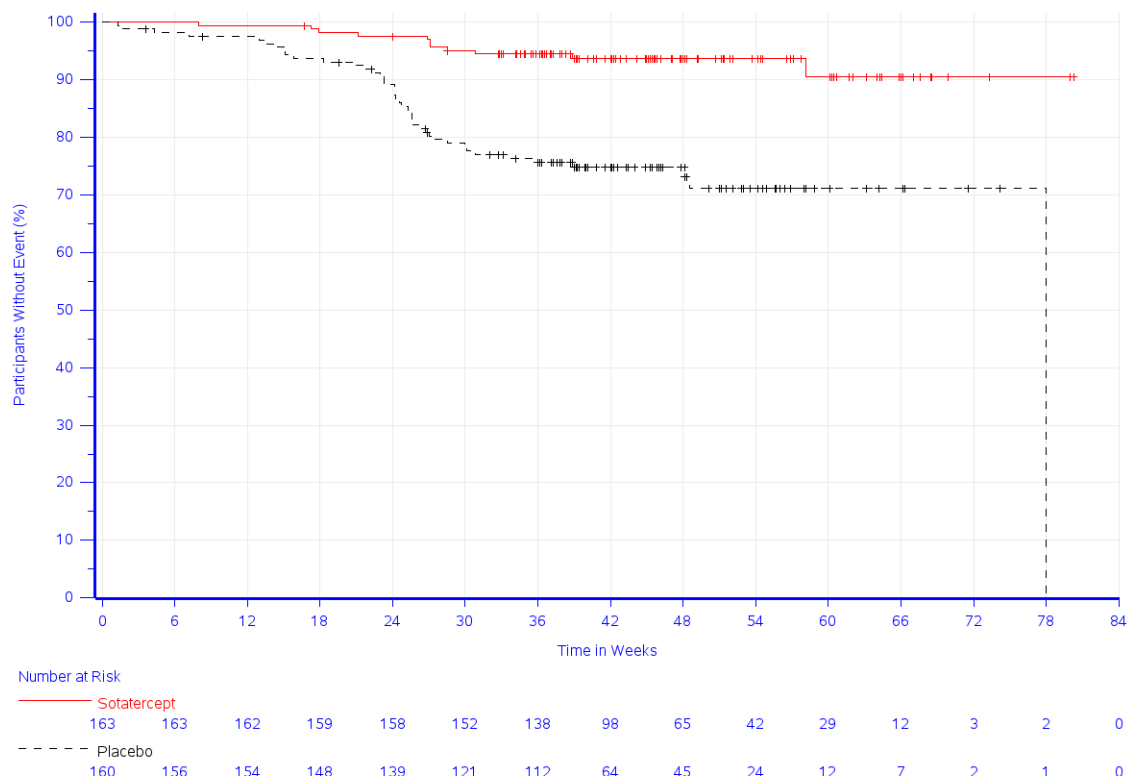
Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Hauptanalyse

Tabelle 4-29: Ergebnisse für Zeit bis zur klinischen Verschlechterung oder Tod aus RCT mit dem zu bewertenden Arzneimittel – Hauptanalyse

	Sotatercept			Placebo			Sotatercept vs. Placebo	
	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	Hazard Ratio [95 %-KI] ^d	p-Wert ^{d,e}
Studie: STELLAR (MK-7962-003)^a								
Zeit bis zur ersten klinischen Verschlechterung oder Tod ^f	163	11 (6,7)	Nicht erreicht [-; -]	160	42 (26,3)	78,00 [-; -]	0,23 [0,12; 0,44]	< 0,001
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Produkt-Limit (Kaplan-Meier) Methode für zensierte Daten</p> <p>d: Basierend auf einem Cox Regressionsmodell mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)</p> <p>e: Zweiseitiger p-Wert (Wald-Test; Score-Test im Fall von null Ereignissen in einem der Studienarme)</p> <p>f: Die klinische Verschlechterung umfasst Endpunkte zu Tod, Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder Erhöhung der Prostazyklindosis um 10% oder mehr, Notwendigkeit einer Atrioseptostomie, Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) und Verschlechterung der PAH definiert durch WHO-Funktionsklasse und 6MWD</p> <p>6MWD: 6-Minuten Gehstrecke; G-BA: Gemeinsamer Bundesausschuss; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization</p>								

Für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod zeigt sich ein statistisch signifikanter Unterschied zugunsten von Sotatercept (HR [95 %-KI]: 0,23 [0,12; 0,44]; p = <0,001). Das Risiko zu versterben oder eine klinische Verschlechterung zu erfahren beträgt etwa ein Viertel für die Patient:innen im Sotatercept-Arm im Vergleich zum Kontrollarm.



STELLAR (MK-7962-003) (Database Cutoff Date: 06DEC2022)
 Time to First Clinical Worsening or Death (includes endpoints of death, worsening related listing for lung and/or heart transplant, need to initiate rescue therapy with an approved PAH therapy or the need to increase the dose of infusion prostacyclin by 10% or more, need for atrial septostomy, hospitalization for worsening of PAH (≥ 24 hours) and deterioration of PAH based on WHO FC and 6MWD)

Abbildung 8: Ereigniszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod der Studie STELLAR – Hauptanalyse

6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation); WHO-FC: WHO-Funktionsklasse

Aus der Ereigniszeitanalyse ist ersichtlich, dass sich die Kaplan-Meier-Kurven bereits sehr früh nach Initiierung der Therapie mit Sotatercept trennen. Im weiteren Studienverlauf verbleiben die Ereigniszahlen für die Patient:innen unter Sotatercept auf einem sehr niedrigen Niveau.

Übersicht der Einzelkomponenten des Endpunkts Zeit bis zur klinischen Verschlechterung oder Tod

Nachfolgend sind die Einzelkomponenten (Tabelle 4-30) für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod aufgeführt.

Tabelle 4-30: Ergebnisse Zeit bis zur klinischen Verschlechterung oder Tod getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Hauptanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)
Erstes Ereignis zur klinischen Verschlechterung oder Tod^{c,d}, n (%)		
Kein Ereignis	152 (93,3)	118 (73,8)
Ereignis	11 (6,7)	42 (26,3)
Todesfall	2 (1,2)	6 (3,8)
Intensivierung der PAH-Hintergrundtherapie ^e	5 (3,1)	13 (8,1)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)
Erstes Ereignis zur klinischen Verschlechterung oder Tod^{c,d}, n (%)		
Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder Erhöhung der Prostazyklindosis um 10% oder mehr	2 (1,2)	14 (8,8)
Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder Erhöhung der Prostazyklindosis um ≥ 10%, PAH-spezifische Hospitalisierung (≥ 24 Stunden), Verschlechterung der PAH	0 (0,0)	1 (0,6)
Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder Erhöhung der Prostazyklindosis um 10% oder mehr, Verschlechterung der PAH ^e	0 (0,0)	1 (0,6)
PAH-spezifische Hospitalisierung (≥ 24 Stunden)	1 (0,6)	6 (3,8)
Listung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression	1 (0,6)	0 (0,0)
Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder Erhöhung der Prostazyklindosis um ≥ 10%	0 (0,0)	1 (0,6)
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Ein(e) Patient:inn kann für das erste Ereignis für klinische Verschlechterung oder Tod mehrere dokumentierte Erhebungen aufweisen. Die Ereignisse werden mittels Kommasetzung in den entsprechenden Zeilen separiert von einander dargestellt</p> <p>d: Die klinische Verschlechterung umfasst Endpunkte zu Tod, Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Dosierung der Infusion mit Prostazyklin um 10% oder mehr, Notwendigkeit einer Atrioseptostomie, Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) und Verschlechterung der PAH definiert durch WHO-Funktionsklasse und 6MWD</p> <p>e: Verschlechterung der PAH definiert durch das Auftreten beider folgender Ereignisse, unabhängig der Reihenfolge, im Vergleich zu Baseline; Verschlechterung der WHO-Funktionsklasse und Verschlechterung der 6MWD um ≥ 15% bestätigt durch 2 Tests mit einem Zeitabstand von mindestens 4 Stunden, aber nicht mehr als 1 Woche</p> <p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation); WHO FC: WHO-Funktionsklasse;</p>		

Im Sotatercept-Arm kam es bei elf von 163 (6,7 %) Patient:innen zum Auftreten eines TTCW-Ereignisses. Zwei Patient:innen verstarben vor Eintritt eines Ereignisses zur klinischen Verschlechterung. Fünf Patient:innen hatten ein Ereignis zur Verschlechterung der PAH. Bei zwei Patient:innen trat die Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Prostazyklindosis um 10 % oder mehr auf. Ein:e Patient:in hatte ein Ereignis zur Hospitalisierung aufgrund von Krankheitsprogression und ein:e Patient:in hatte ein Ereignis zur Listung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression.

Im Placebo-Arm kam es bei 42 von 160 (26,3 %) Patient:innen zum Auftreten eines TTCW-Ereignisses. Sechs Patient:innen verstarben vor Eintritt eines Ereignisses zur klinischen Verschlechterung. 13 Patient:innen hatten ein Ereignis zur Verschlechterung der PAH. Bei 14 Patient:innen trat die Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Prostazyklindosis um 10 % oder mehr auf. Ein:e Patient:in hatte am selben Tag sowohl ein Ereignis zur Therapieeskalation als auch zur Hospitalisierung aufgrund von Krankheitsprogression und ein:e Patient:in ein Ereignis zur Therapieeskalation und ein Ereignis zur Verschlechterung der PAH. Sechs Patient:innen wurden aufgrund von Krankheitsprogression hospitalisiert und ein:e Patient:in hatte am selben Tag eine Listung für eine Lungen und/oder Herztransplantation und eine Therapieeskalation.

Es zeigt sich eine vergleichbare Verteilung der TTCW-Ereignisse in beiden Armen, jedoch traten deutlich mehr Ereignisse im Placebo-Arm auf.

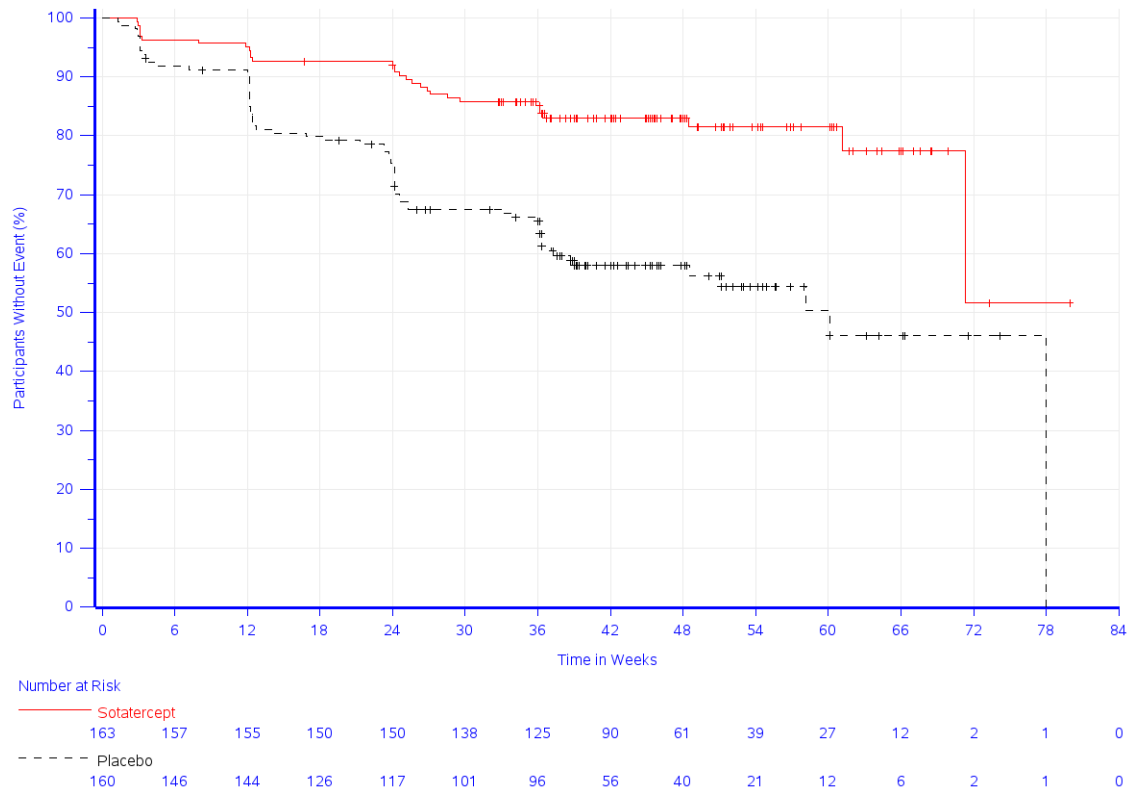
Sensitivitätsanalyse

Tabelle 4-31: Ergebnisse für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod aus RCT mit dem zu bewertenden Arzneimittel – Sensitivitätsanalyse

	Sotatercept			Placebo			Sotatercept vs. Placebo	
	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	N ^b	Patient:innen mit Ereignis n (%)	Mediane Ereigniszeit ^c in Wochen [95 %-KI]	Hazard Ratio [95 %-KI] ^d	p-Wert ^{d,e}
Studie: STELLAR (MK-7962-003)^a								
Zeit bis zur ersten klinischen Verschlechterung oder Tod ^f	163	30 (18,4)	Nicht erreicht [71,29; -]	160	69 (43,1)	60,14 [39,00; -]	0,34 [0,22; 0,52]	< 0,001

a: Datenschnitt: 06. Dezember 2022
 b: Anzahl der Patient:innen: Full-Analysis-Set
 c: Produkt-Limit (Kaplan-Meier) Methode für zensierte Daten
 d: Basierend auf einem Cox Regressionsmodell mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 e: Zweiseitiger p-Wert (Wald-Test; Score-Test im Fall von null Ereignissen in einem der Studienarme)
 f: Die klinische Verschlechterung umfasst folgende Komponenten: Listung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Atrioseptostomie, Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) und Verschlechterung der 6MWD um ≥ 40 Meter
 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); G-BA: Gemeinsamer Bundesausschuss; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization (Weltgesundheitsorganisation)

Die Sensitivitätsanalyse für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod zeigt einen statistisch signifikanten Unterschied zugunsten von Sotatercept (HR [95 %-KI]: 0,34 [0,22; 0,52]; p = < 0,001). Die Sensitivitätsanalyse bestätigt die Ergebnisse der Hauptanalyse und zeigt eine signifikante Überlegenheit von Sotatercept.



STELLAR (MK-7962-003) (Database Cutoff Date: 06DEC2022)
 Time to First Clinical Worsening or Death (includes endpoints of death, worsening related listing for lung and/or heart transplant, need for atrial septostomy, hospitalization for worsening of PAH (≥ 24 hours) and deterioration in 6MWD by ≥ 40m)

Abbildung 9: Ereigniszeitanalyse: Kaplan-Meier-Kurve für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod der Studie STELLAR – Sensitivitätsanalyse

6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie

Für die Kaplan-Meier-Kurven der Sensitivitätsanalyse zeichnet sich ein vergleichbares Bild zur Hauptanalyse ab: die Kaplan-Meier-Kurven trennen sich bereits sehr früh nach Initiierung der Therapie mit Sotatercept und auch im weiteren Studienverlauf verbleiben die Ereigniszahlen im Sotatercept-Arm auf einem niedrigen Niveau.

Übersicht der Einzelkomponenten des Endpunkts Zeit bis zur klinischen Verschlechterung oder Tod – Sensitivitätsanalyse

Nachfolgend sind die Einzelkomponenten (Tabelle 4-30) für die Sensitivitätsanalyse des Endpunkts Zeit bis zur klinischen Verschlechterung oder Tod aufgeführt.

Tabelle 4-32: Ergebnisse für Zeit bis zur klinischen Verschlechterung oder Tod getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Sensitivitätsanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)
Erstes Ereignis zur klinischen Verschlechterung oder Tod^{c,d}, n (%)		
Kein Ereignis	133 (81,6)	91 (56,9)
Ereignis	30 (18,4)	69 (43,1)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)
Erstes Ereignis zur klinischen Verschlechterung oder Tod ^{c,d}, n (%)		
Todesfall	1 (0,6)	4 (2,5)
Verschlechterung der körperlichen Funktion - 6MWD (Verschlechterung um ≥ 40 Meter gegenüber Baseline)	28 (17,2)	60 (37,5)
PAH-spezifische Hospitalisierung (≥ 24 Stunden)	1 (0,6)	3 (1,9)
PAH-spezifische Hospitalisierung (≥ 24 Stunden), Verschlechterung der körperlichen Funktion - 6MWD (Verschlechterung um ≥ 40 M gegenüber Baseline)	0 (0,0)	1 (0,6)
Listung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression	0 (0,0)	1 (0,6)
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Ein(e) Patient:inn kann für das erste Ereignis für klinische Verschlechterung oder Tod mehrere dokumentierte Erhebungen aufweisen. Die Ereignisse werden mittels Kommasetzung in den entsprechenden Zeilen separiert voneinander dargestellt</p> <p>d: Die klinische Verschlechterung umfasst folgende Komponenten: Listung für eine Lungen- und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Atrioseptostomie, PAH-spezifische Hospitalisierung (≥ 24 Stunden) und Verschlechterung der 6MWD um ≥ 40 Meter</p> <p>Verschlechterung der PAH definiert durch das Auftreten beider folgender Ereignisse, unabhängig der Reihenfolge, im Vergleich zu Baseline; Verschlechterung der WHO-Funktionsklasse und Verschlechterung der 6MWD um $\geq 15\%$ bestätigt durch 2 Tests mit einem Zeitabstand von mindestens 4 Stunden, aber nicht mehr als 1 Woche</p> <p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization (Weltgesundheitsorganisation); WHO FC: WHO-Funktionsklasse</p>		

Im Sotatercept-Arm kam es bei 30 von 163 (18,4 %) Patient:innen zum Auftreten eines TTCW-Ereignisses. Ein:e Patient:in verstarb vor Eintritt eines Ereignisses zur klinischen Verschlechterung. 28 Patient:innen hatten ein Ereignis zur Verschlechterung der PAH. Ein:e Patient:in hatte ein Ereignis zur Hospitalisierung aufgrund von Krankheitsprogression.

Im Placebo-Arm kam es bei 69 von 160 (43,1 %) Patient:innen zum Auftreten eines TTCW-Ereignisses. Vier Patient:innen verstarben vor Eintritt eines Ereignisses zur klinischen Verschlechterung. 60 Patient:innen hatten ein Ereignis zur Verschlechterung der PAH. Drei Patient:innen wurden aufgrund von Krankheitsprogression hospitalisiert. Ein:e Patient:in hatte am selben Tag ein Ereignis zur Hospitalisierung und zur Verschlechterung der PAH. Ein:e Patient:in wurde für eine Lungen und/oder Herztransplantation gelistet.

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.

Nicht zutreffend.

4.3.1.3.1.2.3 Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test – RCT

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 des Endpunkts Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test

Studie	Operationalisierung
STELLAR	<p>Die Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test wird mittels Responderanalyse analysiert. Ergänzend wird anhand der medianen Differenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehmann Location Shift/Wilcoxon-Test) die Veränderung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test dargestellt, vergleiche hierzu Abschnitt 4.2.5.2.2</p> <p>Die Borg CR10-Skala wird im Rahmen der 6MWD jeweils zum Zeitpunkt des Screenings (Baseline-Wert) sowie zu Woche 3, 12 und 24 erhoben. Darüber hinaus wird dieser alle 12 Wochen erhoben.</p> <p>Die Borg CR10-Skala ist ein Bewertungsverfahren zur Einteilung des Schweregrads einer Dyspnoe. Die CR10-Skala ist eine kategorische Skala mit Skalenwerten von 0 bis 10, wobei der Wert 0 einer normalen Atmung entspricht und 10 einer extremen Atemnot. Die Borg CR10-Skala wird vor und nach Erhebung der 6-Minuten Gehstrecke erhoben. Dabei wird die Skala den Patient:innen gezeigt und diese aufgefordert, ihre Atemnot zu beschreiben. Die verbalen Instruktionen, welche die Beobachter den Patient:innen geben, sind exakt vorgegeben.</p> <p>Aufgrund der methodischen Vorgaben des G-BA wird die Analyse mit 15 % der Skalenspannweite dargestellt (37).</p> <p>Auswertung über den Studienverlauf:</p> <p>In der vorliegenden Nutzenbewertung wird eine Auswertung über den Studienverlauf der einzelnen Erhebungszeitpunkte dargestellt. Hierzu werden zu jedem Erhebungszeitpunkt die Anzahl an verfügbaren Daten, der Mittelwert (inkl. Standardabweichung), der Median (inkl. Q1; Q3) und Minimum und Maximum zum jeweiligen Zeitpunkt deskriptiv dargestellt. Zusätzlich wird eine grafische Darstellung des Mittelwerts und des Standardfehlers über den Studienverlauf in Anhang 4-G dargestellt.</p> <p>Umgang mit fehlenden Werten:</p> <p><u>Responderanalyse:</u></p> <p>Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p><u>Mittelwertdifferenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehmann Location Shift/Wilcoxon-Test):</u></p> <p>Es werden keine Imputationen durchgeführt.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population</p>
<p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); Covid-19: Coronavirus Disease 2019; CR10-Skala: 10 Point Category Ratio Scale; FAS: Full Analysis Set; G-BA: Gemeinsamer Bundesausschuss</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 4-34: Bewertung des Verzerrungspotenzials für Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test in RCT mit dem zu bewertenden Arzneimittel

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

a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip.
 CR10-Skala: 10 Point Category Ratio Scale; FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test für die Studie STELLAR ist somit als niedrig einzustufen.

Eine detaillierte Übersicht zu den Rücklaufquoten, getrennt nach Behandlungsarmen und Erhebungszeitpunkten befindet sich in Anhang 4-G.

Stellen Sie die Ergebnisse für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test 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 Verbesserung der Dyspnoe um 15 % der Skalenspannweite gemäß Borg CR10-Skala pre 6MWD-Test aus RCT mit dem zu bewertenden Arzneimittel – Responderanalysen

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	N ^b	Patient:innen mit Ereignis n (%)	N ^b	Patient:innen mit Ereignis n (%)	Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
Verbesserung der Pre-6MWD-Test Dyspnoe gemäß Borg-Skala ^f	160	38 (23,8)	159	37 (23,3)	1,02 [0,69; 1,51]	0,918	0,49 [-8,89; 9,78]

a: Datenschnitt: 06. Dezember 2022
b: Anzahl Patient:innen:

- mit Erhebung sowohl zu Baseline als auch zu Woche 24
- sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19.
- Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt.

Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.
c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)
d: Zweiseitiger p-Wert (Wald-Test)
e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
f: Erfüllt, sofern eine Patient:in eine Verbesserung gegenüber Baseline für die Pre-6MWD-Test Dyspnoe gemäß Borg Skala zu Woche 24 erreicht (Verringerung um $\geq 15\%$)
6MWD: 6-Minuten Gehstrecke; COVID-19: Coronavirus Disease of 2019; CR10: 10 Punkte Category Ratio scale; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization

Für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test zeigt sich kein Unterschied zwischen Sotatercept und Placebo.

Tabelle 4-36: Ergebnisse für Veränderung der Dyspnoe gemäß Borg CR10-Skala aus RCT mit dem zu bewertenden Arzneimittel – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensätze der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline des Pre-6MWD Dyspnoe Wertes gemäß Borg-Skala zu Woche 24					
Sotatercept	155	n.a.	0,00 (-0,250, 0,000)	0,064	0,9093
Placebo	146	n.a.			

a: Datenschnitt: 06. Dezember 2022
b: Anzahl an Patient:innen: Full Analysis Set Population, eingeschränkt auf Patient:innen mit Wert zu Baseline und Woche 24
c: Mittelwert, Minimum und Maximum der Mediane der imputierten Datensätze sofern fehlende Werte imputiert wurden. N.a. wenn keine Imputation durchgeführt wurde.
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen
e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)
6MWD: 6-Minuten Gehstrecke; ASE: Asymptotischer Standardfehler; CR10: 10 Punkte Category Ratio scale; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization

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.

Nicht zutreffend.

4.3.1.3.1.2.4 Mehrkomponentenverbesserung

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

Tabelle 4-37: Operationalisierung des Endpunkts Mehrkomponentenverbesserung

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Mehrkomponentenverbesserung ist definiert als der Anteil an Patient:innen, welche zu Woche 24 in allen der folgenden Einzelkomponenten ein Ereignis aufweisen:</p> <ul style="list-style-type: none"> • Verbesserung der 6MWD (≥ 30 m) • Verbesserung in NT-proBNP (Verringerung des NT-proBNP Wertes um ≥ 30 %) oder Erhaltung bzw. Erreichung eines NT-proBNP Wertes von < 300 ng/L) • Verbesserung der WHO-FK oder Erhaltung der WHO-Funktionsklasse II <p>Die Einzelkomponenten 6MWD sowie die WHO-FK werden zu Woche 3, 12 und 24 sowie danach alle 12 Wochen erhoben. Die Einzelkomponente NT-proBNP wird zu Woche 1, 3, 6, 9 und 12 sowie Woche 24 und danach alle 12 Wochen erhoben.</p> <p>Umgang mit fehlenden Werten:</p> <p>Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population.</p>
<p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); COVID-19: Coronavirus Disease 2019; FAS: Full Analysis Set; L: Liter; m: Meter; ng: Nanogramm, NT-proBNP: N-Terminal Pro-B-Type Natriuretic Peptide; WHO: World Health Organization (Weltgesundheitsorganisation); WHO-FK: WHO-Funktionsklasse</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 4-38: Bewertung des Verzerrungspotenzials für Mehrkomponentenverbesserung in RCT mit dem zu bewertenden Arzneimittel

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

a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip
 FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Mehrkomponentenverbesserung für die Studie STELLAR ist somit als niedrig einzustufen.

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

Tabelle 4-39: Ergebnisse für Mehrkomponentenverbesserung aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	N ^b	Patient:innen mit Ereignis n (%)	N ^b	Patient:innen mit Ereignis n (%)	Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
Mehrkomponentenverbesserung ^f	162	64 (39,5)	159	17 (10,7)	3,70 [2,28; 6,00]	< 0,001	28,91 [20,31; 37,62]

a: Datenschnitt: 06. Dezember 2022
 b: Anzahl der Patient:innen: Full Analysis Set Population. Patient:innen mit fehlender Erhebung zu Woche 24 und nicht aufgrund von COVID-19 werden als Non-Responder gewertet. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden ausgeschlossen
 c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)
 d: Zweiseitiger p-Wert (Wald-Test)
 e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 f: Erfüllt, sofern ein:e Patient:in allen der folgenden Parameter zu Woche 24 gegenüber Baseline erreicht: Verbesserung der 6MWD (Verbesserung ≥ 30 Meter), Verbesserung des NT-proBNP (Verringerung des NT-proBNP-Wertes $\geq 30\%$) oder Erhaltung/Erreichen eines NT-proBNP Wertes von < 300 ng/L sowie Verbesserung der WHO-Funktionsklasse oder Erhaltung der WHO-Funktionsklasse II
 6MWD: 6-Minuten Gehstrecke; COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; NT-proBNP: N-terminal Prohormone of Brain Natriuretic Peptide; PAH: Pulmonale arterielle Hypertonie; WHO FC: WHO-Funktionsklasse; WHO: World Health Organization

Für den Endpunkt Mehrkomponentenverbesserung zeigt sich ein statistisch signifikanter Unterschied zugunsten von Sotatercept (siehe Tabelle 4-39, RR [95 %-KI]: 3,70 [2,28; 6,00]; $p = < 0,001$). Fast viermal so viele Patient:innen unter Sotatercept als Patient:innen unter Placebo erzielten eine klinisch relevante Verbesserung der Erkrankung.

Übersicht der Einzelkomponenten des Endpunkts Mehrkomponentenverbesserung

Nachfolgend sind die Einzelkomponenten (Tabelle 4-40 bis Tabelle 4-42) für den Endpunkt Mehrkomponentenverbesserung aufgeführt.

Tabelle 4-40: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung der 6MWD (≥ 30 m)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)	Total (N ^b = 323)
Verbesserung der 6MWD gegenüber Baseline (≥ 30m) zu Woche 24, n (%)			
Ja	88 (54,0)	35 (21,9)	123 (38,1)
Nein	70 (42,9)	112 (70,0)	182 (56,3)
Fehlende Erhebung aufgrund von COVID-19	0 (0,0)	1 (0,6)	1 (0,3)
Fehlende Erhebung aufgrund sonstiger Gründen	5 (3,1)	12 (7,5)	17 (5,3)
a: Datenschnitt: 06. Dezember 2022			
b: Anzahl der Patient:innen: Full-Analysis-Set			
6MWD: 6-Minuten Gehstrecke; Covid-19: Coronavirus Erkrankung verursacht durch schweres akutes Atemwegssyndrom			

Tabelle 4-41: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung in NT-proBNP (Verringerung des NT-proBNP Wertes um ≥ 30 %) oder Erhaltung bzw. Erreichung eines NT-proBNP Wertes von < 300 ng/L)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)	Total (N ^b = 323)
Verbesserung des NT-proBNP gegenüber Baseline (≥ 30%) oder Erhaltung/Verbesserung des NT-proBNP Wertes < 300 ng/L zu Woche 24, n (%)			
Ja	138 (84,7)	66 (41,3)	204 (63,2)
Nein	11 (6,7)	73 (45,6)	84 (26,0)
Fehlende Erhebung aufgrund von COVID-19	1 (0,6)	1 (0,6)	2 (0,6)
Fehlende Erhebung aufgrund sonstiger Gründen	13 (8,0)	20 (12,5)	33 (10,2)
a: Datenschnitt: 06. Dezember 2022			
b: Anzahl der Patient:innen: Full-Analysis-Set			
Covid-19: Coronavirus Erkrankung verursacht durch schweres akutes Atemwegssyndrom; NT-proBNP: N-terminal Prohormone of Brain Natriuretic Peptide			

Tabelle 4-42: Ergebnisse Mehrkomponentenverbesserung getrennt nach den Einzelkomponenten aus RCT mit dem zu bewertenden Arzneimittel in der Studie STELLAR – Verbesserung oder Erhaltung der WHO-FK

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)	Total (N ^b = 323)
Verbesserung der WHO-Funktionsklasse gegenüber Baseline oder Erhaltung der WHO-Funktionsklasse II zu Woche 24, n (%)			
Ja	115 (70,6)	82 (51,3)	197 (61,0)
Nein	44 (27,0)	65 (40,6)	109 (33,7)
Fehlende Erhebung aufgrund von COVID-19	0 (0,0)	1 (0,6)	1 (0,3)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept (N ^b = 163)	Placebo (N ^b = 160)	Total (N ^b = 323)
Verbesserung der WHO-Funktionsklasse gegenüber Baseline oder Erhaltung der WHO-Funktionsklasse II zu Woche 24, n (%)			
Fehlende Erhebung aufgrund sonstiger Gründen	4 (2,5)	12 (7,5)	16 (5,0)
a: Datenschnitt: 06. Dezember 2022			
b: Anzahl der Patient:innen: Full-Analysis-Set			
Covid-19: Coronavirus Erkrankung verursacht durch schweres akutes Atemwegssyndrom; WHO: World Health Organization			

Im Sotatercept-Arm konnten 115 von 163 (70,6 %) der Patient:innen eine Verbesserung der WHO-FK oder Erhaltung der WHO-FK II erzielen. 138 von 163 Patient:innen (84,7 %) konnten ihren NT-ProBNP um mindestens 30 % verbessern oder ihren Wert auf unter 300 ng/L bringen. 88 (54,0 %) der Patient:innen konnten ihre zurückgelegte Gehstrecke innerhalb von 6 Minuten gegenüber Baseline um mindestens 30 Meter verbessern.

Im Placebo-Arm konnten 82 von 160 (51,3 %) der Patient:innen eine Verbesserung oder Erhaltung ihrer WHO-FK erzielen. 66 von 160 (41,3 %) Patient:innen konnten ihren NT-ProBNP um mindestens 30 % verbessern oder ihren Wert auf unter 300 ng/L bringen. 35 (21,9 %) der 160 Patient:innen konnten ihre zurückgelegte Gehstrecke innerhalb von 6 Minuten gegenüber Baseline um mindestens 30 Meter verbessern.

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.

Nicht zutreffend.

4.3.1.3.1.2.5 Krankheitssymptomatik und Gesundheitszustand – RCT

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

Tabelle 4-43: Operationalisierung des Endpunkts Krankheitssymptomatik und Gesundheitszustand

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Krankheitssymptomatik und Gesundheitszustand (Verbesserung der Krankheitssymptomatik und des Gesundheitszustands) wird mittels Responderanalysen analysiert. Ergänzend dargestellt wird die mediane Differenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehmann Location Shift/Wilcoxon-Test).</p> <p>Krankheitssymptomatik:</p> <p>Zur Einschätzung der Krankheitssymptomatik werden die Symptom-Items des Pulmonary Arterial Hypertension – Symptoms and Impact (PAH-SYMPACT)-Fragebogens (18) herangezogen.</p> <p>Die Symptom-Items des PAH-SYMPACT umfassen insgesamt 12 Fragen, von denen 6 Fragen zu kardiopulmonalen Symptomen und 5 Fragen zu kardiovaskulären Symptomen zuzuordnen sind, sowie 1 Frage zum Sauerstoff-Gebrauch. Zu den kardiopulmonalen Symptomen gehören: Kurzatmigkeit, Erschöpfung, Antriebslosigkeit, Schwellungen in den Knöcheln oder Beinen, Schwellungen im Bereich des Magens sowie Husten. Zu den kardiovaskulären Symptomen gehören: Herzflattern, Herzrasen, Brustschmerz, Gefühl der Brustenge sowie Gefühl der Benommenheit. Zur Beantwortung der Fragen stehen die Antwortkategorien entsprechend einer 5-Punkte Likert Skala [z. B. „Überhaupt nicht“ (= 0), „Mild“ (= 1), „Mäßig“ (= 2), „Schwer“ (= 3) und „Sehr schwer“ (= 4)] zur Verfügung.</p> <p>Aufgrund der methodischen Vorgaben des G-BA wird die Analyse mit 15 % der Skalenspannweite dargestellt (37).</p> <p>Gesundheitszustand:</p> <p>Zur Einschätzung des Gesundheitszustandes wird die VAS des EQ-5D herangezogen.</p> <p>Der EQ-5D ist ein krankheitsübergreifender, validierter Fragebogen zur Erfassung des Gesundheitszustandes (19). Er besteht aus 2 Teilen – dem deskriptiven System und der VAS, wobei für die vorliegende Nutzenbewertung lediglich die VAS relevant ist und dargestellt wird. Anhand der VAS schätzen die Patient:innen ihren Gesundheitszustand auf einer Skala von 0 (schlechtest möglicher aktueller Gesundheitszustand) bis 100 (bestmöglicher aktueller Gesundheitszustand) ein (22). Aufgrund der methodischen Vorgaben des G-BA wird die Analyse mit 15 % der Skalenspannweite dargestellt (37).</p> <p>Auswertung über den Studienverlauf:</p> <p>In der vorliegenden Nutzenbewertung wird eine Auswertung über den Studienverlauf der einzelnen Symptom-Items und der VAS dargestellt. Hierzu werden zu jedem Erhebungszeitpunkt die Anzahl an verfügbaren Daten, der Mittelwert (inkl. Standardabweichung), der Median (inkl. Q1; Q3) und Minimum und Maximum zum jeweiligen Zeitpunkt deskriptiv dargestellt. Zusätzlich wird eine grafische Darstellung des Mittelwerts und des Standardfehlers über den Studienverlauf in Anhang 4-G dargestellt.</p> <p>Der PAH-SYMPACT und EQ-5D-Fragebogen wurde gemäß Studienprotokoll zu Woche 1 (Baseline-Wert) und zu Woche 24 erhoben.</p>

Studie	Operationalisierung
	<p><i>Umgang mit fehlenden Werten (PAH-SYMPACT):</i></p> <p><u>Responderanalyse:</u> Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p><u>Mittelwertdifferenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehman Location Shift/Wilcoxon-Test):</u> Standard multiple Imputation mit Logit-Transformation wie in der Studie präspezifiziert</p> <p><i>Umgang mit fehlenden Werten (EQ-5D VAS):</i></p> <p><u>Responderanalyse:</u> Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p><u>Mittelwertdifferenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehman Location Shift/Wilcoxon-Test):</u> Es werden keine Imputationen durchgeführt.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population.</p>
	<p>Covid-19: Coronavirus Disease 2019; EQ-5D: EuroQoL-5 Dimensions; FAS: Full Analysis Set; G-BA: Gemeinsamer Bundesausschuss; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact; Q1: Erstes Quartil; Q3: Drittes Quartil; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie); VAS: Visual Analog Scale (visuelle Analogskala)</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 4-44: Bewertung des Verzerrungspotenzials für Krankheitssymptomatik und Gesundheitszustand in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ^a	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
STELLAR	niedrig	ja	ja	ja	ja	niedrig
a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Krankheitssymptomatik und Gesundheitszustand für die Studie STELLAR ist somit als niedrig einzustufen.

Die Rücklaufquoten für den PAH-SYMPACT (für die Domänen Kardiopulmonale Symptome sowie Kardiovaskuläre Symptome) betrug in beiden Armen zu Baseline über 70 %. Für den EQ-5D VAS betrug die Rücklaufquote in beiden Armen zu Baseline über 75 %. Es ist davon auszugehen, dass die Werte zu Baseline zufällig fehlen (Missing at Random). Für die Verbesserung zu Woche 24 werden für beide Fragebögen (PAH-SYMPACT mit den Domänen Kardiopulmonale Symptome sowie Kardiovaskuläre Symptome und EQ-5D VAS) über 70 % der Patient:innen in beiden Armen mit vorliegendem Baselinewert mit einbezogen.

Eine detaillierte Übersicht zu den Rücklaufquoten, getrennt nach Fragebögen, Behandlungsarmen und Erhebungszeitpunkten befindet sich in Anhang 4-G.

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

Tabelle 4-45: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiopulmonale Symptome) – Responderanalyse

Studie: STELLAR (MK-7962-003)^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio^c [95 %-KI]	p-Wert^d	Differenz in %^e [95 %-KI]
	N^b		N^b				
Verbesserung der PAH-SYMPACT Domäne: kardiopulmonale Symptome ^f	115	47 (40,9)	117	35 (29,9)	1,35 [0,95; 1,93]	0,095	10,52 [-1,75; 22,55]

a: Datenschnitt: 06. Dezember 2022
b: Anzahl Patient:innen:

- mit Erhebung sowohl zu Baseline als auch zu Woche 24
- sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19.
- Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt.

Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.
c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)
d: Zweiseitiger p-Wert (Wald-Test)
e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
f: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kardiopulmonale Symptome zu Woche 24 erreicht (Verringerung um $\geq 15\%$)
COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization

Für die kardiopulmonalen Symptome (Kurzatmigkeit, Erschöpfung, Antriebslosigkeit, Schwellungen in den Knöcheln oder Beinen, Schwellungen im Bereich des Magens sowie Husten) werden im Sotatercept-Arm mehr Patient:innen mit Verbesserung beobachtet. Der Effekt erreicht jedoch keine statistische Signifikanz.

Tabelle 4-46: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiovaskuläre Symptome) – Responderanalyse

Studie: STELLAR (MK-7962-003)^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio^c [95 %-KI]	p-Wert^d	Differenz in %^e [95 %-KI]
	N^b		N^b				
Verbesserung der PAH-SYMPACT Domäne: kardiovaskuläre Symptome ^f	115	49 (42,6)	117	34 (29,1)	1,48 [1,04; 2,11]	0,030	13,86 [1,53; 25,83]

a: Datenschnitt: 06. Dezember 2022
b: Anzahl Patient:innen:

- mit Erhebung sowohl zu Baseline als auch zu Woche 24

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo	
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	Differenz in % ^e [95 %-KI]
	N ^b		N ^b		p-Wert ^d	
<ul style="list-style-type: none"> sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt. Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet. c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie) d: Zweiseitiger p-Wert (Wald-Test) e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) f: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kardiovaskuläre Symptome zu Woche 24 erreicht (Verringerung um $\geq 15\%$) COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization						

Für die kardiovaskulären Symptome (Herzflattern, Herzrasen, Brustschmerz, Gefühl der Brustenge sowie Gefühl der Benommenheit) zeigt sich eine statistisch signifikante Verbesserung für die Patient:innen unter der Behandlung mit Sotatercept gegenüber den Patient:innen aus dem Placebo-Arm (RR [95%-KI]; p-Wert: 1,48 [1,04; 2,11]; 0,030).

Tabelle 4-47: Ergebnisse für Verbesserung der Krankheitssymptomatik und Gesundheitszustand um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (EQ-5D VAS) – Responderanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo	
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	Differenz in % ^e [95 %-KI]
	N ^b		N ^b		p-Wert ^d	
Verbesserung des EQ-5D VAS ^f	124	29 (23,4)	126	20 (15,9)	1,49 [0,89; 2,49]	0,131 7,74 [-2,16; 17,62]
a: Datenschnitt: 06. Dezember 2022 b: Anzahl Patient:innen: <ul style="list-style-type: none"> mit Erhebung sowohl zu Baseline als auch zu Woche 24 sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt. Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet. c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie) d: Zweiseitiger p-Wert (Wald-Test) e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) f: Erfüllt, sofern ein:e Patientin eine Verbesserung gegenüber Baseline für den EQ-5D VAS zu Woche 24 erreicht (Verbesserung um $\geq 15\%$) COVID-19: Coronavirus Disease of 2019; EQ-5D: European Quality of Life 5 Dimensions; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; VAS: Visuelle Analogskala; WHO: World Health Organization						

Das Ergebnis für die VAS des EQ-5D zeigt ebenfalls einen Unterschied zugunsten der Behandlung mit Sotatercept, der jedoch keine statistische Signifikanz erreicht.

Tabelle 4-48: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiopulmonale Symptome) – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der PAH-SYMPACT Domäne kardiopulmonale Symptome zu Woche 24					
Sotatercept	163	-0,11 (-0,14, -0,06)	-0,13 (-0,249, -0,007)	0,062	0,0370
Placebo	160	-0,01 (-0,03, 0,02)			
Datenschnitt: 06. Dezember 2022					
b: Anzahl der Patient:innen: Full-Analysis-Set					
c: Mittelwert, Minimum und Maximum der Mediane der imputierten Daten sofern fehlende Werte imputiert wurden. Für Patient:innen mit fehlenden Werten für den PAH SYMPACT Score zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 200 zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert für den PAH SYMPACT Score aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 150 zugeordnet um den nächstniedrigsten Rang zu bekommen. N.a. wenn keine Imputation durchgeführt wurde.					
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen					
e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)					
ASE: Asymptotischer Standardfehler; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Tabelle 4-49: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kardiovaskuläre Symptome) – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der PAH-SYMPACT Domäne kardiovaskuläre Symptome zu Woche 24					
Sotatercept	163	-0,08 (-0,13, -0,03)	-0,13 (-0,243, -0,014)	0,058	0,0219
Placebo	160	0,00 (0,00, 0,00)			
a: Datenschnitt: 06. Dezember 2022					
b: Anzahl der Patient:innen: Full-Analysis-Set					
c: Mittelwert, Minimum und Maximum der Mediane der imputierten Daten sofern fehlende Werte imputiert wurden. Für Patient:innen mit fehlenden Werten für den PAH SYMPACT Score zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 200 zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert für den PAH SYMPACT Score aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 150 zugeordnet um den nächstniedrigsten Rang zu bekommen. N.a. wenn keine Imputation durchgeführt wurde.					
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen					
e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)					
ASE: Asymptotischer Standardfehler; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Tabelle 4-50: Ergebnisse für Krankheitssymptomatik und Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel (EQ-5D VAS) – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der EQ-5D VAS zu Woche 24					
Sotatercept	90	n.a.	4,50 (0,500, 9,000)	2,168	0,0327

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensätze der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Placebo	89	n.a.			

a: Datenschnitt: 06. Dezember 2022
b: Anzahl an Patient:innen: Full Analysis Set Population, eingeschränkt auf Patient:innen mit Wert zu Baseline und Woche 24
c: Mittelwert, Minimum und Maximum der Mediane der imputierten Datensätze sofern fehlende Werte imputiert wurden. N.a. wenn keine Imputation durchgeführt wurde.
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen
e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)
ASE: Asymptotischer Standardfehler; EQ-5D: European Quality of Life 5 Dimensions; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH: Pulmonale arterielle Hypertonie ; VAS: Visuelle Analogskala; WHO: World Health Organization

Die Ergebnisse der für die Studie STELLAR präspezifizierten Analyse zeigen für alle drei Endpunkte eine statistisch signifikant größere mediane Verbesserung unter Sotatercept im Vergleich zum Placebo-Arm.

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.

Nicht zutreffend.

4.3.1.3.1.3 Gesundheitsbezogene Lebensqualität – RCT

4.3.1.3.1.3.1 Gesundheitsbezogene Lebensqualität – RCT

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

Tabelle 4-51: Operationalisierung des Endpunkts Gesundheitsbezogene Lebensqualität

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Gesundheitsbezogene Lebensqualität wird mittels Responderanalysen analysiert (Verbesserung der Gesundheitsbezogenen Lebensqualität). Ergänzend dargestellt wird die mediane Differenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehmann Location Shift/Wilcoxon-Test).</p> <p>Zur Einschätzung der Lebensqualität werden die Impact-Items des Pulmonary Arterial Hypertension – Symptoms and Impact (PAH-SYMPACT)-Fragebogens (18) herangezogen.</p> <p>Der PAH-SYMPACT umfasst insgesamt 11 Fragen zu den Impact-Items: „Sich langsam auf einer ebenen Oberfläche zu bewegen“, „Sich schnell auf einer ebenen Fläche zu bewegen“, „Bergauf gehen“, „Dinge zu tragen“, „leichte Arbeit im Haushalt“, „sich waschen oder ankleiden“, „benötigte Hilfe“, „Fähigkeit klar zu denken“, „Trauergefühl“, „Unruhe“, und „Frustration“. Zur Beantwortung der Fragen, die in die Items eingehen, stehen die Antwortkategorien entsprechend einer 5-Punkte Likert-Skala [z. B. „Überhaupt nicht“ (= 0), „Mild“ (= 1), „Mäßig“ (= 2), „Schwer“ (= 3) und „Sehr schwer“ (= 4)] zur Verfügung.</p> <p>Aufgrund der methodischen Vorgaben des G-BA wird die Analyse mit 15 % der Skalenspannweite dargestellt (37).</p> <p>Auswertung über den Studienverlauf:</p> <p>In der vorliegenden Nutzenbewertung wird eine Auswertung über den Studienverlauf der einzelnen Impact-Items dargestellt. Hierzu werden zu jedem Erhebungszeitpunkt die Anzahl an verfügbaren Daten, der Mittelwert (inkl. Standardabweichung), der Median (inkl. Q1; Q3) und Minimum und Maximum zum jeweiligen Zeitpunkt tabellarisch dargestellt. Zusätzlich wird eine grafische Darstellung des Mittelwerts und des Standardfehlers über den Studienverlauf in Anhang 4-G dargestellt.</p> <p>Der PAH-SYMPACT wurde gemäß Studienprotokoll zu Woche 1 (Baseline-Wert) und zu Woche 24 erhoben.</p> <p>Umgang mit fehlenden Werten:</p> <p><u>Responderanalyse:</u></p> <p>Fehlende Werte aufgrund einer COVID-19 Erkrankung werden nicht berücksichtigt. Fehlende Werte, die nicht auf COVID-19 zurückzuführen sind, werden als Non-Responder gewertet.</p> <p><u>Mittelwertdifferenz zwischen den Studienarmen im Vergleich zu Baseline (Hodge Lehman Location Shift/Wilcoxon-Test):</u></p> <p>Standard multiple Imputation mit Logit-Transformation</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der FAS-Population.</p>

Studie	Operationalisierung
	Covid-19: Coronavirus Disease 2019; FAS: Full Analysis Set; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact; Q1: Erstes Quartil; Q3: Drittes Quartil; G-BA: Gemeinsamer Bundesausschuss

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 4-52: Bewertung des Verzerrungspotenzials für Krankheitssymptomatik und Gesundheitszustand in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ^a	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
STELLAR	niedrig	ja	ja	ja	ja	niedrig
a: Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip FAS: Full Analysis Set; ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren, liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Gesundheitsbezogene Lebensqualität für die Studie STELLAR ist somit als niedrig einzustufen.

Die Rücklaufquoten für den PAH-SYMPACT (für die Domänen Kognitive/Emotionale Auswirkungen sowie Körperliche Auswirkungen) betrug in beiden Armen zu Baseline über 70 %. Es ist davon auszugehen, dass die Werte zu Baseline zufällig fehlen (Missing at Random). Für die Verbesserung zu Woche 24 werden für den PAH-SYMPACT (für beide Domänen) über 70 % der Patient:innen in beiden Armen mit vorliegendem Baselinewert mit einbezogen.

Eine detaillierte Übersicht zu den Rücklaufquoten, getrennt nach Behandlungsarmen und Erhebungszeitpunkten, befindet sich in Anhang 4-G.

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

Tabelle 4-53: Ergebnisse für Verbesserung der Gesundheitsbezogenen Lebensqualität um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Körperliche Auswirkungen) – Responderanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
	N ^b		N ^b				
Verbesserung der PAH-SYMPACT Domäne: Auswirkungen auf die körperliche Funktion ^f	117	39 (33,3)	123	31 (25,2)	1,31 [0,87; 1,96]	0,193	7,73 [-3,86; 19,27]

a: Datenschnitt: 06. Dezember 2022
b: Anzahl Patient:innen:

- mit Erhebung sowohl zu Baseline als auch zu Woche 24
- sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19.
- Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt.

Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.
c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz ≤ 1 % oder ≥ 99 % in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)
d: Zweiseitiger p-Wert (Wald-Test)
e: Miittinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
f: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Auswirkungen auf die körperliche Funktion zu Woche 24 erreicht (Verringerung um ≥ 15%)
COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization

Für die Auswirkungen auf die körperliche Funktion berichten Patient:innen unter Sotatercept häufiger Verbesserungen als im Placebo-Arm. Diese erreichen jedoch keine statistische Signifikanz.

Tabelle 4-54: Ergebnisse für Verbesserung der Gesundheitsbezogenen Lebensqualität um 15 % der Skalenspannweite aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kognitive/Emotionale Auswirkungen) – Responderanalyse

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%)		Patient:innen mit Ereignis n (%)		Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
	N ^b		N ^b				
Verbesserung der PAH-SYMPACT Domäne:	117	30 (25,6)	123	30 (24,4)	1,04 [0,67; 1,60]	0,866	0,88 [-10,09; 11,84]

Studie: STELLAR (MK-7962-003) ^a	Sotatercept	Placebo	Sotatercept vs. Placebo		
	Patient:innen mit Ereignis n (%) N ^b	Patient:innen mit Ereignis n (%) N ^b	Relatives Risiko/ Peto-Odds Ratio ^c [95 %-KI]	p-Wert ^d	Differenz in % ^e [95 %-KI]
kognitive/emotionale Auswirkungen ^f					
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl Patient:innen:</p> <ul style="list-style-type: none"> mit Erhebung sowohl zu Baseline als auch zu Woche 24 sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt. <p>Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.</p> <p>c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)</p> <p>d: Zweiseitiger p-Wert (Wald-Test)</p> <p>e: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)</p> <p>f: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kognitive/Emotionale Auswirkungen zu Woche 24 erreicht (Verringerung um $\geq 15\%$)</p> <p>COVID-19: Coronavirus Disease of 2019; KI: Konfidenzintervall; PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization</p>					

Für die kognitiven/emotionalen Auswirkungen lässt sich kein Unterschied zwischen der Behandlung mit Sotatercept und Placebo feststellen.

Tabelle 4-55: Ergebnisse für Gesundheitsbezogene Lebensqualität aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Körperliche Auswirkungen) – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der PAH-SYMPACT Domäne körperlicher Einfluss zu Woche 24					
Sotatercept	163	-0,12 (-0,15, 0,00)	-0,26 (-0,482, -0,034)	0,114	0,0116
Placebo	160	0,01 (0,00, 0,14)			
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Mittelwert, Minimum und Maximum der Mediane der imputierten Daten sofern fehlende Werte imputiert wurden. Für Patient:innen mit fehlenden Werten für den PAH SYMPACT Score zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 200 zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert für den PAH SYMPACT Score aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 150 zugeordnet um den nächstniedrigsten Rang zu bekommen. N.a. wenn keine Imputation durchgeführt wurde.</p> <p>d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen</p> <p>e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)</p> <p>ASE: Asymptotischer Standardfehler; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization</p>					

Tabelle 4-56: Ergebnisse für Gesundheitsbezogene Lebensqualität aus RCT mit dem zu bewertenden Arzneimittel (PAH-SYMPACT Kognitive/Emotionale Auswirkungen) – Veränderung gegenüber Baseline

Studie: STELLAR (MK-7962-003) ^a	N ^b	Median	Sotatercept vs. Placebo (Hodges-Lehman Location Shift)		
		Imputierter Median (Min, Max) über die imputierten Datensets der geschätzten Mediane hinweg. ^c	Effektschätzer (95 %-KI) ^d	ASE	p-Wert ^e
Veränderung gegenüber Baseline der PAH-SYMPACT Domäne kognitiver/emotionaler Einfluss zu Woche 24					
Sotatercept	163	0,00 (0,00, 0,00)	-0,15 (-0,383, 0,091)	0,121	0,1863
Placebo	160	0,00 (0,00, 0,00)			
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Mittelwert, Minimum und Maximum der Mediane der imputierten Daten sofern fehlende Werte imputiert wurden. Für Patient:innen mit fehlenden Werten für den PAH SYMPACT Score zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 200 zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert für den PAH SYMPACT Score aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von 150 zugeordnet um den nächstniedrigsten Rang zu bekommen. N.a. wenn keine Imputation durchgeführt wurde.</p> <p>d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen</p> <p>e: Zweiseitiger p-Wert basierend auf einem stratifizierten Wilcoxon-Test mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifach- vs. Dreifachtherapie)</p> <p>ASE: Asymptotischer Standardfehler; KI: Konfidenzintervall; Max: Maximum; Min: Minimum; n.a.: nicht anwendbar (wenn keine Imputation durchgeführt wurde); PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization</p>					

Die Ergebnisse der für die Studie STELLAR präspezifizierten Analyse zeigen ein ähnliches Bild: das Impact-Item „körperliche Auswirkungen“ zeigt eine signifikante Verbesserung für die Patient:innen unter Behandlung mit Sotatercept, für das Impact-Item „kognitive/emotionale Auswirkungen“ lässt sich kein Unterschied feststellen.

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.

Nicht zutreffend.

4.3.1.3.1.4 Nebenwirkungen – RCT

4.3.1.3.1.4.1 Unerwünschte Ereignisse Gesamtraten – RCT

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

Tabelle 4-57: Operationalisierung des Endpunkts Unerwünschte Ereignisse Gesamtraten zu Woche 24

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Unerwünschte Ereignisse Gesamtraten zu Woche 24 wird mittels Responderanalysen analysiert. Die Analyse betrachtet den Anteil an Patient:innen, welche innerhalb der ersten 24 Wochen ein unerwünschtes Ereignis aufwiesen. Die Ableitung des Zusatznutzens erfolgt anhand des relativen Risikos.</p> <p>Folgende Endpunkte werden unter dem Endpunkt Unerwünschte Ereignisse Gesamtraten zu Woche 24 zusammengefasst:</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse gesamt • Schwerwiegende unerwünschte Ereignisse • Schwere unerwünschte Ereignisse • Therapieabbruch wegen unerwünschter Ereignisse <p><u>Unerwünschte Ereignisse gesamt</u></p> <p>Ein unerwünschtes Ereignis ist definiert als jedes unvorhergesehene medizinische Ereignis bei den Patient:innen, denen ein Arzneimittel verabreicht wurde, dieses muss aber nicht notwendigerweise in kausalem Zusammenhang zur Therapie stehen. Ein unerwünschtes Ereignis kann daher jedes nachteilige und unbeabsichtigte Anzeichen (einschließlich z. B. abnormer Laborergebnisse), Symptom oder jede Krankheit sein, die in zeitlichem Zusammenhang mit der Verwendung des Arzneimittels steht – unabhängig von seinem Zusammenhang mit dem Arzneimittel.</p> <p><u>Schwerwiegende unerwünschte Ereignisse</u></p> <p>Ein schwerwiegendes unerwünschtes Ereignis ist jegliches unerwünschtes Ereignis, das mindestens eines der folgenden Kriterien erfüllt:</p> <ul style="list-style-type: none"> • Es ist tödlich • Es ist lebensbedrohlich • Es führt zur Hospitalisierung der Patient:innen oder zur Verlängerung eines bereits bestehenden stationären Aufenthalts • Es führt zu anhaltender oder bedeutender Einschränkung oder Behinderung • Es zieht eine angeborene Anomalie oder einen Geburtsfehler nach sich • Es ist aus einem anderen Grund ein medizinisch bedeutsames Ereignis <p><u>Schwere unerwünschte Ereignisse</u></p> <p>Ein schweres unerwünschtes Ereignis ist jegliches unerwünschte Ereignis, welches anhand folgender Kriterien durch den Prüfarzt als „Schwer“ eingeordnet wird. Folgende Klassifizierung wurde in der Studie präspezifiziert und von den Prüfarzt:innen angewendet:</p> <ul style="list-style-type: none"> • Mild: asymptomatische oder milde Symptome, rein klinische oder diagnostische Beobachtungen, keine Intervention indiziert. • Moderat: minimale, lokale oder nicht invasive Eingriffe indiziert, Beeinträchtigung der alltäglichen Aktivitäten • Schwer: schwer, medizinisch signifikant aber nicht unmittelbar lebensbedrohlich, Hospitalisierung erforderlich, Verhinderung der alltäglichen Aktivitäten. <p><u>Therapieabbruch wegen unerwünschter Ereignisse</u></p>

Studie	Operationalisierung
	<p>Der Endpunkt Therapieabbruch wegen unerwünschter Ereignisse ist definiert als jegliches unerwünschte Ereignis, das zum permanenten Absetzen der Studienmedikation führt.</p> <p>Der Beobachtungszeitraum für alle unerwünschten Ereignisse ist definiert als Zeit von der ersten Dosis der Studienmedikation bis zu 8 Wochen nach der letzten Gabe der Studienmedikation.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der Safety-Analysis-Set-Population und werden gemäß Medical Dictionary for Regulatory Activities (MedDRA, Version 25.0 oder höher) kodiert.</p> <p>Analysen ohne erkrankungsbezogene Ereignisse (z. B. Progression, Exazerbation) wurden nicht durchgeführt, da bei den aufgetretenen Ereignissen keines identifiziert werden konnte, das eindeutig eine Progression oder Exazerbation der Erkrankung darstellt.</p>
MedDRA: Medical Dictionary for Regulatory Activities	

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 4-58: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse Gesamtraten in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ^a	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
STELLAR	niedrig	ja	ja	ja	ja	niedrig
<p>a: Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population ITT: Intention-To-Treat; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population. Diese umfasst alle Patient:innen, die mindestens eine Dosis

der Studienmedikation erhalten haben. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Unerwünschte Ereignisse Gesamtraten für die Studie STELLAR ist somit als niedrig einzustufen.

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

Tabelle 4-59: Ergebnisse für Unerwünschte Ereignisse Gesamtraten aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo	
	Sotatercept (N ^b =163)	Placebo (N ^b =160)	Relatives Risiko [95 %-KI] ^c	p-Wert ^d
Unerwünschte Ereignisse	151 (92,64)	149 (93,13)	0,99 [0,94; 1,06]	0,8651
Schwerwiegende unerwünschte Ereignisse	40 (24,54)	47 (29,38)	0,84 [0,58; 1,20]	0,3281
Schwere unerwünschte Ereignisse	24 (14,72)	33 (20,63)	0,71 [0,44; 1,15]	0,1649
Therapieabbruch wegen Unerwünschter Ereignisse	6 (3,68)	11 (6,88)	0,54 [0,20; 1,41]	0,1994

a: Datenschnitt: 06. Dezember 2022
b: Anzahl der Patient:innen: Safety Set Population
c: 2x2 Kontingenztabelle mit Konfidenzintervall nach Wald. Bei keinem:keiner Patient:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztabelle, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppe kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet
d: Basierend auf einen Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet
KI: Konfidenzintervall; n.a.: nicht anwendbar

In der Studie STELLAR zeigt sich bei den Gesamtraten der unerwünschten Ereignisse kein relevanter Unterschied zwischen Sotatercept und Placebo (siehe Tabelle 4-59).

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.

Nicht zutreffend

4.3.1.3.1.4.2 Unerwünschte Ereignisse (gegliedert nach SOC und PT) – RCT

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

Tabelle 4-60: Operationalisierung des Endpunkts Unerwünschte Ereignisse (gegliedert nach SOC und PT)

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Unerwünschte Ereignisse zu Woche 24 (gegliedert nach SOC und PT) wird mittels Responderanalysen analysiert. Die Analyse betrachtet den Anteil an Patient:innen, welche innerhalb der ersten 24 Wochen ein unerwünschtes Ereignis (gegliedert nach SOC und PT) aufwiesen. Die Ableitung des Zusatznutzens erfolgt anhand des relativen Risikos.</p> <p>Folgende Endpunkte werden unter dem Endpunkt Unerwünschte Ereignisse zu Woche 24 (gegliedert nach SOC und PT) zusammengefasst:</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse gesamt (SOC und PT) • Schwerwiegende unerwünschte Ereignisse (SOC und PT) • Schwere unerwünschte Ereignisse (SOC und PT) • Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) <p>Die Operationalisierung der einzelnen Endpunkte der Endpunktkategorie Unerwünschte Ereignisse (gegliedert nach SOC und PT) entspricht der in Tabelle 4-57 beschriebenen Vorgehensweise bei den Gesamtraten der UE.</p> <p>Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet.</p> <p>Alle Auswertungen erfolgen auf Grundlage der Safety-Analysis-Set-Population und werden gemäß Medical Dictionary for Regulatory Activities (MedDRA, Version 25.0 oder höher) kodiert.</p>
<p>MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Terms; SOC: System Organ Class (Systemorganklasse); UE: Unerwünschtes Ereignis</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 4-61: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse (gegliedert nach SOC und PT) in RCT mit dem zu bewertenden Arzneimittel

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

a: Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population
ITT: Intention-To-Treat; PT: Preferred Terms; RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie);
SOC: System Organ Class (Systemorganklasse)

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der Safety-Analysis-Set-Population. Diese umfasst alle Patient:innen, die mindestens eine Dosis der Studienmedikation erhalten haben. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für die Studie STELLAR ist somit als niedrig einzustufen.

Stellen Sie die Ergebnisse für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Im Folgenden werden die Ergebnisse für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) ergänzend dargestellt. Die Systematik zur Darstellung des Endpunkts im Dossier richtet sich nach dem in Tabelle 4-5 beschriebenen Vorgehen.

Unerwünschte Ereignisse gesamt (SOC und PT)

Tabelle 4-62: Ergebnisse für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Unerwünschte Ereignisse gesamt (SOC und PT) ^b					
Erkrankungen des Blutes und des Lymphsystems	29 (17,79)	25 (15,63)	1,14 [0,70; 1,86]	0,6024	0,7631
Thrombozytopenie	16 (9,82)	3 (1,88)	5,24 [1,56; 17,62]	0,0025	0,0034
Herzkrankungen	30 (18,40)	30 (18,75)	0,98 [0,62; 1,55]	0,9365	0,9658
Erkrankungen des Ohrs und des Labyrinths	13 (7,98)	14 (8,75)	0,91 [0,44; 1,88]	0,8018	0,8961

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Unerwünschte Ereignisse gesamt (SOC und PT) ^b					
Augenerkrankungen	21 (12,88)	7 (4,38)	2,94 [1,29; 6,73]	0,0067	0,0627
Erkrankungen des Gastrointestinaltrakts	68 (41,72)	51 (31,88)	1,31 [0,98; 1,75]	0,0671	0,2358
Diarrhoe	25 (15,34)	16 (10,00)	1,53 [0,85; 2,76]	0,1503	n.s.
Übelkeit	23 (14,11)	19 (11,88)	1,19 [0,67; 2,10]	0,5510	n.s.
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	61 (37,42)	50 (31,25)	1,20 [0,88; 1,62]	0,2435	0,3856
Erschöpfung	23 (14,11)	16 (10,00)	1,41 [0,77; 2,57]	0,2577	n.s.
Schmerz an der Injektionsstelle	11 (6,75)	11 (6,88)	0,98 [0,44; 2,20]	0,9641	n.s.
Peripheres Ödem	14 (8,59)	12 (7,50)	1,15 [0,55; 2,40]	0,7195	n.s.
Erkrankungen des Immunsystems	11 (6,75)	7 (4,38)	1,54 [0,61; 3,88]	0,3533	0,4794
Infektionen und parasitäre Erkrankungen	105 (64,42)	89 (55,63)	1,16 [0,97; 1,39]	0,1073	0,2548
COVID-19	48 (29,45)	42 (26,25)	1,12 [0,79; 1,60]	0,5222	n.s.
Nasopharyngitis	11 (6,75)	13 (8,13)	0,83 [0,38; 1,80]	0,6377	n.s.
Infektion der oberen Atemwege	9 (5,52)	11 (6,88)	0,80 [0,34; 1,89]	0,6144	n.s.
Harnwegsinfekt	11 (6,75)	6 (3,75)	1,80 [0,68; 4,75]	0,2283	n.s.
Verletzung, Vergiftung und durch Eingriffe bedingte Komplikationen	27 (16,56)	19 (11,88)	1,39 [0,81; 2,41]	0,2286	0,3856
Untersuchungen	33 (20,25)	21 (13,13)	1,54 [0,93; 2,55]	0,0869	0,2358
Erhöhter Hämoglobin-Wert	10 (6,13)	0	20,61 [1,22; 348,86]	0,0015	0,0026
Stoffwechsel- und Ernährungsstörungen	37 (22,70)	33 (20,63)	1,10 [0,73; 1,67]	0,6515	0,7736
Hypokaliämie	14 (8,59)	6 (3,75)	2,29 [0,90; 5,81]	0,0716	n.s.
Eisenmangel	11 (6,75)	9 (5,63)	1,20 [0,51; 2,82]	0,6758	n.s.
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	50 (30,67)	30 (18,75)	1,64 [1,10; 2,43]	0,0132	0,0627
Erkrankungen des Nervensystems	62 (38,04)	40 (25,00)	1,52 [1,09; 2,12]	0,0119	0,0627
Schwindel	24 (14,72)	10 (6,25)	2,36 [1,16; 4,77]	0,0132	n.s.
Kopfschmerzen	40 (24,54)	28 (17,50)	1,40 [0,91; 2,16]	0,1213	n.s.
Psychiatrische Erkrankungen	13 (7,98)	8 (5,00)	1,60 [0,68; 3,74]	0,2789	0,4077
Erkrankungen der Nieren und Harnwege	7 (4,29)	12 (7,50)	0,57 [0,23; 1,42]	0,2216	0,3856
Erkrankungen der Geschlechtsorgane und der Brustdrüse	10 (6,13)	10 (6,25)	0,98 [0,42; 2,29]	0,9658	0,9658
Erkrankungen der Atemwege, des Brustraums und Mediastinums	59 (36,20)	48 (30,00)	1,21 [0,88; 1,65]	0,2376	0,3856
Dyspnoe	5 (3,07)	17 (10,63)	0,29 [0,11; 0,76]	0,0071	0,0083
Nasenbluten	36 (22,09)	3 (1,88)	11,78 [3,70; 37,48]	<0,0001	<0,0001
Nasale Kongestion	10 (6,13)	0	20,61 [1,22; 348,86]	0,0015	0,0026
Erkrankungen der Haut und des Unterhautzellgewebes	65 (39,88)	31 (19,38)	2,06 [1,42; 2,97]	0,0001	0,0011
Ausschlag	13 (7,98)	6 (3,75)	2,13 [0,83; 5,46]	0,1071	0,1071
Teleangiektasie	27 (16,56)	7 (4,38)	3,79 [1,70; 8,44]	0,0004	0,0013
Gefäßkrankungen	27 (16,56)	16 (10,00)	1,66 [0,93; 2,95]	0,0830	0,2358
Spülung	10 (6,13)	4 (2,50)	2,45 [0,79; 7,66]	0,1093	n.s.

a: Datenschnitt: 06. Dezember 2022

b: Eine Systemorganklasse oder ein spezifisches unerwünschtes Ereignis werden dargestellt, wenn Ereignisse bei mindestens 10 % der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen und bei mindestens 1 % der Patient:innen in einem Studienarm auftreten

c: Anzahl der Patient:innen: Safety Set Population

d: 2x2 Kontingenztabelle mit Konfidenzintervall nach Wald. Bei keinem/keiner Patient:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztabelle, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppe kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet

e: Basierend auf einen Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet

f: Adjustierter p-Wert für Unerwünschte Ereignisse auf SOC-Ebene anhand der FDR Methodik, bzw. auf Ebene der PT anhand der Double FDR (dFDR) Methodik. Nicht signifikant (i.e., 'n.s.')

wird für PT berichtet, wenn die übergeordnete SOC das Kriterium für den p-Wert im ersten Schritt der dFDR Methodik nicht erfüllt. Der adjustierte p-Wert kann Hinweise auf falsch-positive (signifikante) Ergebnisse des nominalen p-Wertes geben und kann somit den Einfluss des multiplen Testens kontrollieren

FDR: False Discovery Rate; KI: Konfidenzintervall; n.a.: nicht anwendbar; n.s.: nicht signifikant (adjustierter p-Wert $\geq 0,05$); PT: Preferred

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Unerwünschte Ereignisse gesamt (SOC und PT) ^b					
Terms; SOC: Systemorganklasse					

In der SOC Erkrankungen des Blutes und des Lymphsystems zeigt sich kein signifikanter Unterschied zwischen Sotatercept und Placebo. Innerhalb der SOC zeigt sich ein signifikanter Unterschied hinsichtlich des PT Thrombozytopenie (RR [95 %-KI]: 5,24 [1,56; 17,62]; p = 0,0025; p_{adj} = 0,0034) zwischen den Studienarmen.

In der SOC Augenerkrankungen zeigt sich ein signifikanter Unterschied (RR [95 %-KI]: 2,94 [1,29; 6,73]; p = 0,0067; p_{adj} = 0,0627) zwischen den Studienarmen.

In der SOC Untersuchungen zeigt sich kein signifikanter Unterschied zwischen Sotatercept und Placebo. Innerhalb der SOC zeigt sich ein signifikanter Unterschied hinsichtlich des PT Erhöhter Hämoglobin (Hb)-Wert (RR [95 %-KI]: 20,61 [1,22; 348,86]; p = 0,0015; p_{adj} = 0,0026) zwischen den Studienarmen.

In der SOC Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen zeigt sich ein signifikanter Unterschied (RR [95 %-KI]: 1,64 [1,10; 2,43]; p = 0,0132; p_{adj} = 0,0627) zwischen den Studienarmen.

In der SOC Erkrankungen des Nervensystems zeigt sich ein signifikanter Unterschied (RR [95 %-KI]: 1,52 [1,09; 2,12]; p = 0,0119; p_{adj} = 0,0627) zwischen den Studienarmen. Innerhalb der SOC zeigt sich ein signifikanter Unterschied hinsichtlich des PT Schwindel (RR [95 %-KI]: 2,36 [1,16; 4,77]; p = 0,0132; p_{adj} = n. s.).

In der SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums zeigt sich kein signifikanter Unterschied zwischen Sotatercept und Placebo. Innerhalb der SOC zeigt sich ein signifikanter Unterschied hinsichtlich folgender PT:

- Dyspnoe (RR [95 %-KI]: 0,29 [0,11; 0,76]; p = 0,0071; p_{adj} = 0,0083)
- Nasenbluten (RR [95 %-KI]: 11,78 [3,70; 37,48]; p = < 0,0001; p_{adj} = < 0,0001)
- Nasale Kongestion (RR [95 %-KI]: 20,61 [1,22; 348,86]; p = < 0,0015; p_{adj} = < 0,00026)

In der SOC Erkrankungen der Haut und des Unterhautzellgewebes zeigt sich ein signifikanter Unterschied (RR [95 %-KI]: 2,06 [1,42; 2,97]; p = 0,0001; p_{adj} = 0,0011) zwischen den Studienarmen. Innerhalb der SOC zeigt sich ein signifikanter Unterschied hinsichtlich des PT Teleangiectasien (RR [95 %-KI]: 3,79 [1,70; 8,44]; p = 0,0004; p_{adj} = 0,0013).

Insgesamt zeichnet sich auf SOC- und PT-Ebene ein gut verträgliches und behandelbares Nebenwirkungsprofil für Sotatercept ab, welches vergleichbar mit anderen PAH-Therapien ist.

Die Ereignisse für Nasenbluten und Teleangiektasien sind bis auf wenige Einzelfälle mild, es traten keine schweren Ereignisse auf.

Schwerwiegende Unerwünschte Ereignisse (SOC und PT)

Tabelle 4-63: Ergebnisse für den Endpunkt Schwerwiegende Unerwünschte Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Schwerwiegende unerwünschte Ereignisse (SOC und PT)^b					
Herzerkrankungen	6 (3,68)	11 (6,88)	0,54 [0,20; 1,41]	0,1994	0,3018
Infektionen und parasitäre Erkrankungen	14 (8,59)	8 (5,00)	1,72 [0,74; 3,98]	0,2012	0,3018
Erkrankungen der Atemwege, des Brustraums und Mediastinums	7 (4,29)	11 (6,88)	0,62 [0,25; 1,57]	0,3128	0,3128

a: Datenschnitt: 06. Dezember 2022
b: Eine Systemorganklasse oder ein spezifisches unerwünschtes Ereignis werden dargestellt, wenn Ereignisse bei mindestens 5 % der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen und bei mindestens 1 % der Patient:innen in einem Studienarm auftreten
c: Anzahl der Patient:innen: Safety Set Population
d: 2x2 Kontingenztafel mit Konfidenzintervall nach Wald. Bei keinem:keiner Patienten:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztafel, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppen kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet
e: Basierend auf einem Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet
f: Adjustierter p-Wert für Unerwünschte Ereignisse auf SOC-Ebene anhand der FDR Methodik, bzw. auf Ebene der PT anhand der Double FDR (dFDR) Methodik. Nicht signifikant (i.e., 'n.s.') wird für PT berichtet, wenn die übergeordnete SOC das Kriterium für den p-Wert im ersten Schritt der dFDR Methode nicht erfüllt. Der adjustierte p-Wert kann Hinweise auf falsch-positive (signifikante) Ergebnisse des nominalen p-Wertes geben und kann somit den Einfluss des multiplen Testens kontrollieren
FDR: False Discovery Rate; KI: Konfidenzintervall; n.a.: nicht anwendbar; n.s.: nicht signifikant (adjustierter p-Wert ≥ 0,05); PT: Preferred Terms; SOC: Systemorganklasse

In der Studie STELLAR zeigen sich für den Endpunkt Schwerwiegende unerwünschte Ereignisse (SOC und PT) keine Unterschiede zwischen den Behandlungsgruppen.

Schwere Unerwünschte Ereignisse (SOC und PT)

Tabelle 4-64: Ergebnisse für den Endpunkt Schwere Unerwünschte Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Schwere unerwünschte Ereignisse (SOC und PT)^b					
Herzerkrankungen	3 (1,84)	8 (5,00)	0,37 [0,10; 1,36]	0,1181	0,1181
Erkrankungen der Atemwege, des Brustraums und Mediastinums	5 (3,07)	11 (6,88)	0,45 [0,16; 1,26]	0,1154	0,1181

a: Datenschnitt: 06. Dezember 2022
b: Eine Systemorganklasse oder ein spezifisches unerwünschtes Ereignis werden dargestellt, wenn Ereignisse bei mindestens 5 % der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen und bei mindestens 1 % der Patient:innen in einem Studienarm auftreten
c: Anzahl der Patient:innen: Safety Set Population
d: 2x2 Kontingenztafel mit Konfidenzintervall nach Wald. Bei keinem:keiner Patienten:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztafel, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppen kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet
e: Basierend auf einem Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo		
	Sotatercept (N ^c =163)	Placebo (N ^c =160)	Relatives Risiko [95 %-KI] ^d	p-Wert ^e	Adjustierter p-Wert ^f
Schwere unerwünschte Ereignisse (SOC und PT)^b					
f: Adjustierter p-Wert für Unerwünschte Ereignisse auf SOC-Ebene anhand der FDR Methodik, bzw. auf Ebene der PT anhand der Double FDR (dFDR) Methodik. Nicht signifikant (i.e., 'n.s.') wird für PT berichtet, wenn die übergeordnete SOC das Kriterium für den p-Wert im ersten Schritt der dFDR Methode nicht erfüllt. Der adjustierte p-Wert kann Hinweise auf falsch-positive (signifikante) Ergebnisse des nominalen p-Wertes geben und kann somit den Einfluss des multiplen Testens kontrollieren					
FDR: False Discovery Rate; KI: Konfidenzintervall; n.a.: nicht anwendbar; n.s.: nicht signifikant (adjustierter p-Wert ≥ 0,05); PT: Preferred Terms; SOC: Systemorganklasse					

In der Studie STELLAR zeigen sich für den Endpunkt Schwere unerwünschte Ereignisse (SOC und PT) keine Unterschiede zwischen den Behandlungsgruppen.

Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT)

Tabelle 4-65: Ergebnisse für den Endpunkt Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)	
	Sotatercept (N ^c = 163)	Placebo (N ^c = 160)
Therapieabbruch wegen Unerwünschter Ereignisse (SOC und PT)^b		
Patient:innen mit mindestens einem unerwünschten Ereignis	6 (3,68)	11 (6,88)
Herzerkrankungen	1 (0,61)	3 (1,88)
Herzstillstand	0 (0,00)	2 (1,25)
Rechtsherzversagen	0 (0,00)	1 (0,63)
Akuter Myokardinfarkt	1 (0,61)	0 (0,00)
Erkrankungen des Immunsystems	1 (0,61)	0 (0,00)
Sarkoidose	1 (0,61)	0 (0,00)
Infektionen und parasitäre Erkrankungen	0 (0,00)	2 (1,25)
COVID-19-Lungenentzündung	0 (0,00)	1 (0,63)
Sepsis	0 (0,00)	1 (0,63)
Stoffwechsel- und Ernährungsstörungen	0 (0,00)	1 (0,63)
Mangelernährung	0 (0,00)	1 (0,63)
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	1 (0,61)	0 (0,00)
Arthralgie	1 (0,61)	0 (0,00)
Erkrankungen des Nervensystems	1 (0,61)	0 (0,00)
Intrakranielle Blutung	1 (0,61)	0 (0,00)
Schwangerschaft, Wochenbett und perinatale Bedingungen	0 (0,00)	1 (0,63)
Schwangerschaftsabbruch	0 (0,00)	1 (0,63)
Erkrankungen der Atemwege, des Brustraums und Mediastinums	2 (1,23)	4 (2,50)
Pulmonale arterielle Hypertonie	0 (0,00)	2 (1,25)
Pulmonale Hypertonie	0 (0,00)	1 (0,63)
Respiratorische Insuffizienz	0 (0,00)	1 (0,63)
Nasenbluten	1 (0,61)	0 (0,00)
Hämoptoe	1 (0,61)	0 (0,00)
Erkrankungen der Haut und des Unterhautzellgewebes	1 (0,61)	0 (0,00)
Teleangiektasie	1 (0,61)	0 (0,00)
a: Datenschnitt: 06. Dezember 2022		
b: Eine SOC oder ein spezifisches unerwünschtes Ereignis erscheint nur, wenn die Inzidenz in einem Studienarm > 0 % ist		
c: Anzahl der Patient:innen: Safety Set Population		
PT: Preferred Terms; SOC: Systemorganklasse		

In der Studie STELLAR zeigt sich hinsichtlich der Häufigkeit der Therapieabbrüche wegen unerwünschter Ereignisse (SOC und PT) ein Unterschied. Die Inzidenzen betragen im Interventionsarm 3,68 % und im Kontrollarm 6,88 %. Bezüglich der angegebenen Gründe für

den Therapieabbruch aufgrund unerwünschter Ereignisse (UE) zeigen sich keine auffälligen Unterschiede (Tabelle 4-65).

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.

Nicht zutreffend.

4.3.1.3.1.4.3 Unerwünschte Ereignisse von Interesse – RCT

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

Tabelle 4-66: Operationalisierung des Endpunkts Unerwünschte Ereignisse von besonderem Interesse

Studie	Operationalisierung
STELLAR	<p>Der Endpunkt Unerwünschte Ereignisse von Interesse zu Woche 24 wird mittels Responderanalysen analysiert. Die Analyse betrachtet den Anteil an Patient:innen, welche innerhalb der ersten 24 Wochen ein Unerwünschtes Ereignis von Interesse aufwiesen. Die Ableitung des Zusatznutzens erfolgt anhand des relativen Risikos.</p> <p>Folgende Endpunkte werden unter dem Endpunkt Unerwünschte Ereignisse von Interesse zu Woche 24 zusammengefasst:</p> <ul style="list-style-type: none"> • Schwerwiegende Unerwünschte Ereignisse von Interesse (AEOI) • Schwere Unerwünschte Ereignisse von Interesse (AEOI) • Schwerwiegende Unerwünschte Ereignisse von besonderem Interesse (AEOSI) • Schwere Unerwünschte Ereignisse von besonderem Interesse (AEOSI) <p>Die Operationalisierung der einzelnen Endpunkte der Endpunktkategorie Unerwünschte Ereignisse von Interesse entspricht der in Tabelle 4-57 beschriebenen Vorgehensweise.</p> <p>Die AEOI wurden anhand einer präspezifizierten Liste von SOC/PT ermittelt und umfassen folgende unerwünschte Ereignisse: erhöhter Hämoglobinwert, Thrombozytopenie, Neutropenie (inklusive febriler Neutropenie), embryofetale Toxizität, hepatische Toxizität, kardiologische Ereignisse, thrombotische Ereignisse, Ereignisse betreffend die Niere, Blutungen, erhöhter Blutdruck und supprimierter FSH-Wert.</p> <p>Als AEOSI wurde ausschließlich die Teleangiektasie definiert.</p>

Studie	Operationalisierung
	Für die Auswertungen werden Daten zu Woche 24 auf Basis des finalen Datenschnitts (06. Dezember 2022) verwendet. Alle Auswertungen erfolgen auf Grundlage der Safety-Analysis-Set-Population und werden gemäß Medical Dictionary for Regulatory Activities (MedDRA, Version 25.0 oder höher) kodiert.
AEOI: Adverse Events of Interest; AEOSI: Adverse Events of Special Interest; FSH: Follikel-stimulierendes Hormon; MedDRA: Medical Dictionary for Regulatory Activities; PT: Preferred Terms; SOC: System Organ Class (Systemorganklasse)	

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 4-67: Bewertung des Verzerrungspotenzials für Unerwünschte Ereignisse von Interesse in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips ^a	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
STELLAR	niedrig	ja	ja	ja	ja	niedrig
<p>a: Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population AEOI: Adverse Events of Interest (Unerwünschte Ereignisse von Interesse); RCT: Randomized Controlled Trial (randomisierte kontrollierte Studie)</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial auf Studienebene wird für die Studie STELLAR als niedrig eingestuft. Die Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population. Diese umfasst alle Patient:innen, die mindestens eine Dosis der Studienmedikation erhalten haben. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Unerwünschte Ereignisse von Interesse für die Studie STELLAR ist somit als niedrig einzustufen.

Stellen Sie die Ergebnisse für den Endpunkt Unerwünschte Ereignisse von Interesse (AEOI) für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-68: Ergebnisse für Unerwünschte Ereignisse von Interesse aus RCT mit dem zu bewertenden Arzneimittel

Studie: STELLAR (MK-7962-003) ^a	Patient:innen mit Ereignis n (%)		Sotatercept vs. Placebo	
	Sotatercept (N ^b =163)	Placebo (N ^b =160)	Relatives Risiko [95 %-KI] ^c	p-Wert ^d
Schwerwiegende AEOI	0	0	n.a. [n.a.; n.a.]	n.a.
Schwere AEOI	0	0	n.a. [n.a.; n.a.]	n.a.
Schwerwiegende AEOI	16 (9,82)	12 (7,50)	1,31 [0,64; 2,68]	0,4602
Schwere AEOI	8 (4,91)	10 (6,25)	0,79 [0,32; 1,94]	0,5997

a: Datenschnitt: 06. Dezember 2022
b: Anzahl der Patient:innen: Safety Set Population
c: 2x2 Kontingenztabelle mit Konfidenzintervall nach Wald. Bei keinem:keiner Patient:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztabelle, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppe kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet
d: Basierend auf einen Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet
AEOI: Unerwünschte Ereignisse von Interesse; AESI: Unerwünschte Ereignisse von besonderem Interesse; KI: Konfidenzintervall; n.a.: nicht anwendbar

In der Studie STELLAR traten in keinem Studienarm schwere und schwerwiegende Ereignisse für AEOI auf. Die in der Studie STELLAR aufgetretenen Ereignisse für Teleangiektasien sind bis auf je ein moderates Ereignis in beiden Behandlungsarmen mild (siehe Unerwünschte Ereignisse (gliedert nach SOC und PT) – RCT).

Für die AEOI zeigt sich kein statistisch signifikanter Unterschied zwischen Sotatercept und Placebo.

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.

Nicht zutreffend.

4.3.1.3.2 Subgruppenanalysen – RCT

Für die Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.3.1.3.1.¹⁶

¹⁶ unbesetzt

Darüber hinaus sind folgende Kriterien zu berücksichtigen:

- Subgruppenanalysen sind nur für die Merkmale (z.B. Alter) durchzuführen, bei denen die resultierenden Subgruppen jeweils mindestens 10 Patienten umfassen.
- Subgruppenanalysen sind für binäre Ereignisse je Merkmal nur dann durchzuführen, wenn in einer der Subgruppen mindestens 10 Ereignisse aufgetreten sind.
- Für Überlebenszeitanalysen müssen Kaplan-Meier-Kurven zu den einzelnen Subgruppen nur für Subgruppenanalysen mit statistisch signifikantem Interaktionsterm ($p < 0,05$) dargestellt werden.
- Ergebnisse zu UE nach SOC und PT müssen nur dargestellt werden, wenn das jeweilige Ergebnis für die Gesamtpopulation statistisch signifikant ist. Zu a priori definierten Ereignissen (z.B. AESI, SMQs) sowie den UE-Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE) müssen Subgruppenanalysen unabhängig vom Vorliegen statistischer Signifikanz in der Gesamtpopulation dargestellt werden.
- Bei Vorliegen mehrerer Studien und Durchführung von Metaanalysen zu diesen Studien gelten die zuvor genannten Kriterien für die jeweilige Metaanalyse, nicht für die Einzelstudien.
- Für Studien des pharmazeutischen Unternehmers sind entsprechende Analysen für alle benannten Effektmodifikatoren zu allen relevanten Endpunkten nach den zuvor genannten Kriterien vorzulegen und daher ggf. posthoc durchzuführen.
- Wird für die Nutzenbewertung nur die Teilpopulation einer Studie herangezogen (z.B. wegen Zulassungsbeschränkungen, aufgrund von durch den G-BA bestimmte Teilpopulationen), so gelten die genannten Kriterien für diese Teilpopulation, und die Subgruppenanalysen sind für die Teilpopulation und nicht für die Gesamtpopulation der Studie durchzuführen.
- Subgruppenanalysen, bei denen der Interaktionsterm nicht statistisch signifikant ist, können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine ausschließliche Darstellung in Modul 5 ist aber nicht ausreichend.

Beschreiben Sie die Ergebnisse von Subgruppenanalysen. Stellen Sie dabei zunächst tabellarisch dar, zu welchen der in Abschnitt 4.2.5.5 genannten Effektmodifikatoren Subgruppenanalysen zu den relevanten Endpunkten vorliegen, und ob diese a priori geplant und im Studienprotokoll festgelegt waren oder posthoc durchgeführt wurden.

Orientieren Sie sich an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-69 Matrix der durchgeführten Subgruppenanalysen

Endpunktkategorie Endpunkt	Alter (< 65 Jahre vs. ≥ 65 Jahre)	Geschlecht (männlich vs. weiblich)	Krankheitsschwere (WHO-FK II vs. WHO-FK III)	Region (WHO Stratum A vs. RoW)	PAH-Hintergrundtherapie (Mono/Zweifach vs. Dreifach)	Ätiologische Subgruppen der PAH ^a	Prostazyklin Infusion zu Baseline (ja vs. nein)	PVR-Wert zu Baseline (≤ 800 vs. > 800 dyn*sek/cm ²)
STELLAR								
Mortalität								
Mortalität zu Woche 24	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Morbidität								
Verbesserung der 6MWD	○	○	○	○	○	n.c.	n.c.	n.c.
Veränderung der 6MWD	○	●	●	○	●	●	●	●
Zeit bis zur klinischen Verschlechterung oder Tod	○	○	○	○	○	n.c.	n.c.	n.c.
Verbesserung der Dyspnoe gemäß Borg CR10-Skala	○	○	○	○	○	n.c.	n.c.	n.c.
Mehrkomponenten- verbesserung	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
Krankheitssymptomatik und Gesundheitszustand	○	○	○	○	○	n.c.	n.c.	n.c.
Gesundheitsbezogene Lebensqualität								
Gesundheitsbezogene Lebensqualität	○	○	○	○	○	n.c.	n.c.	n.c.
Nebenwirkungen								
Unerwünschte Ereignisse Gesamtraten ^b	○	○	○	○	○	n.c.	n.c.	n.c.
Unerwünschte Ereignisse (gegliedert nach SOC und PT) ^c	○ ^d	○ ^d	○ ^d	○ ^d	○ ^d	n.c.	n.c.	n.c.
Unerwünschte Ereignisse von Interesse ^e	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
<p>●: A priori geplante Subgruppenanalyse. ○: Posthoc durchgeführte Subgruppenanalyse. n.c.: not calculated (nicht berechnet)</p> <p>a: PAH Ätiologische Subgruppen (PAH Etiological Subgroups [idiopathische pulmonale arterielle Hypertonie, IPAH] vs. heritable pulmonale arterielle Hypertonie [hPAH] vs. Medikamenten-induzierte PAH vs. Bindegewebserkrankungen vs. angeborene Herzfehler)</p> <p>b: Unerwünschte Ereignisse Gesamtraten: Unerwünschte Ereignisse gesamt, Schwerwiegende unerwünschte Ereignisse, Schwere unerwünschte Ereignisse und Therapieabbruch wegen unerwünschter Ereignisse. Der Endpunkt Unerwünschte Ereignisse gesamt wird ergänzend dargestellt.</p> <p>c: Unerwünschte Ereignisse (gegliedert nach SOC und PT): Unerwünschte Ereignisse gesamt (SOC und PT), Schwerwiegende unerwünschte Ereignisse (SOC und PT) und Schwere unerwünschte Ereignisse (CTCAE-Grad 3-5) (SOC und PT). Auf eine Darstellung von Subgruppen für den Endpunkt Therapieabbruch wegen unerwünschter Ereignisse (SOC und PT) wird verzichtet</p>								

Endpunktkategorie Endpunkt	Alter (< 65 Jahre vs. ≥ 65 Jahre)	Geschlecht (männlich vs. weiblich)	Krankheitsschwere (WHO-FK II vs. WHO-FK III)	Region (WHO Stratum A vs. RoW)	PAH-Hintergrundtherapie (Mono/Zweifach vs. Dreifach)	Ätiologische Subgruppen der PAH ^a	Prostazyklin Infusion zu Baseline (ja vs. nein)	PVR-Wert zu Baseline (≤ 800 vs. > 800 dyn*sek/cm ²)
STELLAR								
<p>d: Für SOC/PT werden Subgruppen dargestellt, falls folgende Kriterien erfüllt sind (Subgruppen für a priori definierte Ereignisse (z. B. SMQ) werden unabhängig von den beiden Kriterien berichtet):</p> <ul style="list-style-type: none"> - Jeweilige SOC/PT überschreitet Inzidenzgrenzen (siehe Abschnitt 4.2.5.5) <p>UND</p> <ul style="list-style-type: none"> - Signifikanter Unterschied der jeweiligen SOC/PT bei Betrachtung der Zielpopulation <p>e: Unerwünschte Ereignisse von Interesse: Schwerwiegende und Schwere Unerwünschte Ereignisse von Interesse. Der Endpunkt wird ergänzend dargestellt, auf Subgruppenanalysen wird verzichtet</p> <p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); CR10: 10 Point Category Ratio Scale; CTCAE: Common Terminology Criteria for Adverse Events; hPAH: Heritable pulmonale arterielle Hypertonie; IPAH: Idiopathische pulmonale arterielle Hypertonie; SMQ: Standardised MedDRA Queries; SOC: System Organ Class (Systemorganklasse); PT: Preferred Terms; PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation); WHO-FK: WHO-Funktionsklasse; PVR: Pulmonary Vascular Resistance; RoW: Rest der Welt</p>								

Für die Studie STELLAR wurden umfangreiche Subgruppenanalysen durchgeführt. Eine detaillierte Übersicht über das Vorgehen zur Auswahl der Endpunkte und der zugehörigen Subgruppenmerkmale ist Tabelle 4-7 zu entnehmen. Die Matrix der Subgruppenanalysen ist in Tabelle 4-69 dargestellt.

Die Prüfung des Vorliegens einer möglichen Effektmodifikation erfolgte, wie in Abschnitt 4.2.5.5 beschrieben, innerhalb eines strukturierten, mehrstufigen Prozesses, dessen zentraler Ausgangspunkt die Durchführung eines geeigneten Interaktions- oder Homogenitätstests ist.

Bei der Interpretation der Ergebnisse der Interaktionstests ist das Problem der Multiplizität zu beachten. Dies bedeutet, dass bei der Testung einer größeren Anzahl statistischer Hypothesen – wie hier durch die Durchführung der Interaktionstests – eine gewisse Anzahl der Tests bereits aufgrund der angenommenen Irrtumswahrscheinlichkeit von $\alpha = 5\%$ zu einem falsch positiven Ergebnis führt.

Im Folgenden werden in Abschnitt 4.3.1.3.2.1 zunächst die Ergebnisse der Interaktionstests für alle untersuchten Subgruppen dargestellt. Entsprechend dem strukturierten, mehrstufigen Vorgehen zur Bewertung einer möglichen Effektmodifikation (siehe Abbildung 2) erfolgt anschließend in Abschnitt 4.3.1.3.2.2 eine ausführliche Ergebnisdarstellung für diejenigen Subgruppen, für die ein signifikanter Interaktionstest ($p < 0,05$) vorliegt. Die Ergebnisse der Subgruppen mit nicht signifikantem Interaktionstests ($p \geq 0,05$) werden der Vollständigkeit halber dem Nutzendossier in Anhang 4-G beigelegt.

4.3.1.3.2.1 Überblick über Ergebnisse der Interaktionstests aus Subgruppenanalysen

Stellen Sie anschließend in Tabelle 4-70 bis Tabelle 4-81 die Ergebnisse der Interaktionsterme für alle Subgruppenanalysen je Endpunkt in tabellarischer Form dar, und zwar für jede einzelne Studie separat. Kennzeichnen Sie dabei statistisch signifikante ($p < 0,05$) Interaktionsterme.

Die folgenden Tabellen (Tabelle 4-70 bis Tabelle 4-81) zeigen in einer Übersicht die Ergebnisse sämtlicher Interaktionstests, die für Subgruppenanalysen der Endpunkte aus der Studie STELLAR durchgeführt wurden.

Insgesamt wurden neun positive Interaktionstests beobachtet:

- Subgruppe Krankheitsschwere (WHO-FK II vs. III):
Positiver Interaktionstest bei der 6MWD Veränderung gegenüber Baseline zu Woche 24 ($p = 0,002$).
Positiver Interaktionstest beim EQ-5D VAS ($p = 0,026$).
- Subgruppe Geschlecht (Männlich vs. Weiblich):
Positiver Interaktionstest bei den unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für die SOC Erkrankungen des Blutes und des Lymphsystems ($p = 0,013$).
- Subgruppe Alter (< 65 Jahre vs. ≥ 65 Jahre):
Positiver Interaktionstest bei den unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für den PT Schwindelgefühl (SOC Erkrankungen des Nervensystems) ($p = 0,025$).
- Subgruppe Region (WHO-Stratum A vs. RoW):
Positiver Interaktionstest bei den Unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für die SOC Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen ($p = 0,045$).
Positiver Interaktionstest bei den unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für den PT Nasenbluten (SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums) ($p = 0,024$).
- Subgruppe PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie):
Positiver Interaktionstest bei der EQ-5D VAS ($p = 0,020$).
Positiver Interaktionstest bei den unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für die SOC Erkrankungen des Blutes und des Lymphsystems ($p = 0,028$).

Positiver Interaktionstest bei den unerwünschten Ereignissen gesamt (gegliedert nach SOC und PT) für die SOC Skelettmuskulatur-, Bindegewebs- und Knochen-erkrankungen ($p = 0,021$).

Gesamtmortalität

Mortalität zu Woche 24

Es wurden keine Subgruppenanalysen durchgeführt, da die Anzahl an Ereignissen unterhalb des Schwellenwerts von 10 Ereignissen, ab denen Subgruppenanalysen regelhaft durchgeführt werden sollen, liegt.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Morbidität

Verbesserung der 6MWD

Tabelle 4-70: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Verbesserung der 6MWD – Responderanalyse

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH-Hintergrundtherapie
Morbidität					
Verbesserung der 6MWD	0,880	0,149	0,204	0,918	0,339
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie; WHO: World Health Organization (Weltgesundheitsorganisation)					

Tabelle 4-71: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt 6MWD – Veränderung gegenüber Baseline (Teil 1)

Studie: STELLAR (MK-7962-003) ^a	p-Werte für Behandlung*Subgruppen Interaktionstest ^b			
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region
Morbidität				
Veränderung gegenüber Baseline der 6-Minuten Gehstrecke (in Metern) zu Woche 24 ^c	0,303	0,088	0,002^d	0,661
a: Datenschnitt: 06. Dezember 2022 b: Cochran’s Q (Chi-quadrat) zweiseitiger p-Wert für die Heterogenität zwischen den Subgruppen c: Für Patient:innen mit fehlenden Werten für den 6MWD zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -2000 Metern zugeordnet um den schlechtmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -1000 Metern zugeordnet um den nächstniedrigeren Rang zu erhalten. d: p-Wert des Interaktionstests kleiner als 0,05 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie ; PVR: Pulmonary Vascular Resistance (pulmonaler Gefäßwiderstand)				

Tabelle 4-72: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt 6MWD – Veränderung gegenüber Baseline (Teil 2)

Studie: STELLAR (MK-7962-003) ^a	p-Werte für Behandlung*Subgruppen Interaktionstest ^b			
	PAH-Hintergrundtherapie	Ätiologische Subgruppen der PAH	Prostazyklin Infusion zu Baseline	Baseline PVR
Morbidität				
Veränderung gegenüber Baseline der 6-Minuten Gehstrecke (in Metern) zu Woche 24^c	0,614	0,103	0,776	0,052
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Cochran's Q (Chi-quadrat) zweiseitiger p-Wert für die Heterogenität zwischen den Subgruppen</p> <p>c: Für Patient:innen mit fehlenden Werten für den 6MWD zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -2000 Metern zugeordnet um den schlechtestmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -1000 Metern zugeordnet um den nächstniedrigeren Rang zu erhalten.</p> <p>d: p-Wert des Interaktionstests kleiner als 0,05</p> <p>6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie ; PVR: Pulmonary Vascular Resistance (pulmonaler Gefäßwiderstand)</p>				

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Zeit bis zur klinischen Verschlechterung oder Tod

Tabelle 4-73: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^c				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Morbidität					
Zeit bis zur ersten klinischen Verschlechterung oder Tod ^b	0,356	0,082	0,700	0,756	0,258
a: Datenschnitt: 06. Dezember 2022 b: Die klinische Verschlechterung umfasst Endpunkte zu Tod, Listung für eine Lungen und/oder Herztransplantation aufgrund von Krankheitsprogression, Notwendigkeit einer Therapieeskalation mit einer zugelassenen PAH-Hintergrundtherapie oder die Erhöhung der Dosierung der Infusion mit Prostazyklin um 10% oder mehr, Notwendigkeit einer Atrioseptostomie, Hospitalisierung aufgrund von Krankheitsprogression (≥ 24 Stunden) und Verschlechterung der PAH definiert durch WHO-Funktionsklasse und 6MWD c: Basierend auf einem Cox Modell mit Behandlung und Subgruppe als Kovariaten sowie der Interaktion Behandlung*Subgruppen (p-Wert des Likelihood Ratio Tests für den Interaktionstest) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); PAH: Pulmonale arterielle Hypertonie					

Verbesserung der Dyspnoe gemäß Borg CR10-Skala

Tabelle 4-74: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung der Pre-6MWD-Test Dyspnoe gemäß Borg-Skala^c					
BORG Skala: Dyspnoe (Verbesserung gegenüber Baseline ≥ 15%)	0,885	0,781	0,821	0,541	0,135
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die Pre-6MWD-Test Dyspnoe gemäß Borg Skala zu Woche 24 erreicht (Verringerung um ≥ 15%) 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); CR10: 10 Punkte Category Ratio scale; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Krankheitssymptomatik und Gesundheitszustand

Tabelle 4-75: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (PAH-SYMPACT – Kardiopulmonale Symptome)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung der PAH-SYMPACT Domäne: kardiopulmonale Symptome^c					
PAH-Sympact: Kardiopulmonare Symptome (Verbesserung gegenüber Baseline \geq 15%)	0,381	0,270	0,439	0,585	0,175
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kardiopulmonale Symptome zu Woche 24 erreicht (Verringerung um \geq 15%) PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Tabelle 4-76: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (PAH-SYMPACT – Kardiovaskuläre Symptome)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung der PAH-SYMPACT Domäne: kardiovaskuläre Symptome^c					
PAH-Sympact: Kardiovaskuläre Symptome (Verbesserung gegenüber Baseline \geq 15%)	0,318	0,461	0,584	0,708	0,418
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kardiovaskuläre Symptome zu Woche 24 erreicht (Verringerung um \geq 15%) PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-77: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für die Endpunkte Krankheitssymptomatik und Gesundheitszustand (EQ-5D VAS)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung des EQ-5D VAS^c					
EQ-5D VAS (Verbesserung gegenüber Baseline \geq 15%)	0,758	0,884	0,026^d	0,109	0,020^d
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patientin eine Verbesserung gegenüber Baseline für den EQ-5D VAS zu Woche 24 erreicht (Verbesserung um \geq 15%) d: p-Wert des Interaktionstests kleiner als 0,05 EQ-5D: European Quality of Life 5 Dimensions; PAH: Pulmonale arterielle Hypertonie ; VAS: Visuelle Analogskala; WHO: World Health Organization					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Gesundheitsbezogene Lebensqualität

Gesundheitsbezogene Lebensqualität

Tabelle 4-78: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Gesundheitsbezogene Lebensqualität (PAH-SYMPACT – Körperliche Auswirkungen)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung der PAH-SYMPACT Domäne: Auswirkungen auf die körperliche Funktion^c					
PAH-Sympact: Körperliche Auswirkungen (Verbesserung gegenüber Baseline \geq 15%)	0,345	0,716	0,231	0,668	0,848
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Auswirkungen auf die körperliche Funktion zu Woche 24 erreicht (Verringerung um \geq 15%) PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Tabelle 4-79: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Gesundheitsbezogene Lebensqualität (PAH-SYMPACT – Kognitive/Emotionale Auswirkungen)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH Hintergrundtherapie
Verbesserung der PAH-SYMPACT Domäne: kognitive/emotionale Auswirkungen^c					
PAH-Sympact: Kognitive/Emotionale Auswirkungen (Verbesserung gegenüber Baseline \geq 15%)	0,686	0,226	0,456	0,804	0,321
a: Datenschnitt: 06. Dezember 2022 b: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) c: Erfüllt, sofern ein:e Patient:in eine Verbesserung gegenüber Baseline für die PAH-SYMPACT Domäne Kognitive/Emotionale Auswirkungen zu Woche 24 erreicht (Verringerung um \geq 15%) PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Nebenwirkungen

Unerwünschte Ereignisse Gesamtraten

Tabelle 4-80: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Unerwünschte Ereignisse Gesamtraten

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH-Hintergrundtherapie
Unerwünschte Ereignisse					
Unerwünschte Ereignisse	0,243	0,532	0,486	0,602	0,124
Schwerwiegende unerwünschte Ereignisse	0,411	0,885	0,595	0,982	0,936
Schwere unerwünschte Ereignisse	0,707	0,964	0,495	0,281	0,415
Therapieabbruch wegen Unerwünschter Ereignisse	0,810	0,275	0,621	0,312	0,131
a: Datenschnitt: 06. Dezember 2022					
b: Basierend auf einem Breslow-Day Test mit Subgruppen als Stratifizierungsfaktor. Falls in einer Subgruppenkategorie kein Ereignis auftrat oder alle Patient:innen einer Subgruppenkategorie Ereignisse aufweisen wird die Subgruppenkategorie vom Interaktionstest ausgeschlossen, wenn nur noch eine Subgruppenkategorie verbleibt wird kein Interaktionstest durchgeführt.					
n.a.: nicht anwendbar; PAH: Pulmonale arterielle Hypertonie					

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Unerwünschte Ereignisse (gegliedert nach SOC und PT)

Tabelle 4-81: Überblick der Ergebnisse der Interaktionstests aus Subgruppenanalysen der Studie STELLAR für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT)

STELLAR (MK-7962-003) ^a	p-Werte der Interaktionstests für Behandlung nach Subgruppe ^b				
	Geschlecht	Alter (Jahre)	Krankheitsschwere	Region	PAH-Hintergrundtherapie
Unerwünschte Ereignisse gesamt (SOC und PT)^c					
Erkrankungen des Blutes und des Lymphsystems	0,013^d	0,732	0,555	0,622	0,028^d
Thrombozytopenie	0,325	0,363	0,337	0,476	0,483
Augenerkrankungen	0,148	0,948	0,237	0,095	0,202
Untersuchungen	0,441	0,087	0,454	0,258	0,285
Erhöhter Hämoglobin-Wert	n.c.	n.c.	n.c.	n.c.	n.c.
Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen	0,932	0,609	0,084	0,045^d	0,021^d
Erkrankungen des Nervensystems	0,327	0,402	0,450	0,685	0,416
Schwindel	0,290	0,025^d	0,409	n.a.	0,520
Erkrankungen der Atemwege, des Brustraums und Mediastinums	0,234	0,663	0,797	0,244	0,779
Dyspnoe	0,620	0,412	0,068	n.a.	0,466
Nasenbluten	0,356	0,451	0,909	0,024^d	0,303
Nasale Kongestion	n.c.	n.c.	n.c.	n.a.	n.c.
Erkrankungen der Haut und des Unterhautzellgewebes	0,631	0,888	0,073	0,957	0,173
Teleangiektasie	0,483	0,631	0,875	0,611	0,171
a: Datenschnitt: 06. Dezember 2022					
b: Basierend auf einem Breslow-Day Test mit Subgruppen als Stratifizierungsfaktor. Falls in einer Subgruppenkategorie kein Ereignis auftrat oder alle Patient:innen einer Subgruppenkategorie Ereignisse aufweisen wird die Subgruppenkategorie vom Interaktionstest ausgeschlossen, wenn nur noch eine Subgruppenkategorie verbleibt wird kein Interaktionstest durchgeführt.					
c: Eine Systemorganklasse wird dargestellt, wenn Ereignisse bei mindestens 10% der Patient:innen in einem Studienarm ODER Ereignisse bei mindestens 10 Patient:innen und bei mindestens 1 % der Patient:innen in einem Studienarm auftreten und ein signifikanter Unterschied bei Betrachtung der Zielpopulation vorliegt (p < 0,05)					
d: p-Wert des Interaktionstests kleiner als 0,05					
n.a.: nicht anwendbar; n.c.: not calculated (nicht berechnet; mindestens 10 Patient:innen in jeder Subgruppe und mindestens 10 Ereignisse in einer der Subgruppen notwendig); PAH: Pulmonale arterielle Hypertonie ; PT: Preferred Terms; SOC: Systemorganklasse					

Schwerwiegende unerwünschte Ereignisse (gegliedert nach SOC und PT)

Es wurden keine Subgruppenanalysen durchgeführt, da auf Gesamtstudienebene keine statistisch signifikanten Effekte vorlagen.

Schwere unerwünschte Ereignisse (gegliedert nach SOC und PT)

Es wurden keine Subgruppenanalysen durchgeführt, da auf Gesamtstudienebene keine statistisch signifikanten Effekte vorlagen.

4.3.1.3.2.2 Ergebnisse für Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$)

Stellen Sie schließlich alle Subgruppenergebnisse dar.

Sofern eine Effektmodifikation für mehr als ein Subgruppenmerkmal vorliegt, kann eine Untersuchung auf eine Wechselwirkung höherer Ordnung sinnvoll sein. Dies gilt insbesondere dann, wenn diese Effektmodifikation konsistent über mehrere Endpunkte besteht. Zur Interpretation der Ergebnisse sollte dann für diese Endpunkte zusätzlich eine Subgruppenanalyse durchgeführt werden, die die Merkmale mit Effektmodifikation kombiniert. Beispiel: Für die Endpunkte Mortalität, gesundheitsbezogene Lebensqualität und schwere unerwünschte Ereignisse liegt sowohl für das Merkmal Geschlecht (mit den Ausprägungen „weiblich“ und „männlich“) als auch für das Merkmal Schweregrad (mit den Ausprägungen „niedrig“ und „hoch“) eine Effektmodifikation vor. Die zusätzliche Subgruppenanalyse erfolgt dann für die 3 genannten Endpunkte für das kombinierte Merkmal Geschlecht/Schweregrad mit den 4 Ausprägungen weiblich/niedrig, weiblich/hoch, männlich/niedrig und männlich/hoch.

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.

Die folgenden Abschnitte enthalten die ausführliche Ergebnisdarstellung von Subgruppenanalysen, für die ein positiver Interaktionstest ($p < 0,05$) vorliegt. Entsprechend dem in Abschnitt 4.2.5.5 beschriebenen strukturierten, mehrstufigen Vorgehen wird im nächsten Schritt anhand der Lage des Effektschätzers und des KI geprüft, ob eine abweichende Interpretation des Zusatznutzens und damit eine bewertungsrelevante Effektmodifikation grundsätzlich vorliegen könnte. Ist dies der Fall, wird abschließend anhand der in Abschnitt 4.2.5.5 beschriebenen Detailkriterien untersucht, ob es sich um eine tatsächliche Effektmodifikation oder etwa um ein falsch-positives Ergebnis infolge des multiplen Testens handelt.

Gesamtmortalität

Mortalität zu Woche 24

Es wurden keine Subgruppenanalysen durchgeführt, da die Anzahl an Ereignissen unterhalb des Schwellenwerts liegt.

Morbidität**6MWD Veränderung gegenüber Baseline zu Woche 24**

Tabelle 4-82: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt 6MWD Veränderung gegenüber Baseline zu Woche 24 aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept N ^b =163	Placebo N ^b =160	Sotatercept vs. Placebo	p-Wert für Interaktionstest ^e
Veränderung gegenüber Baseline der 6-Minuten Gehstrecke (in Metern) zu Woche 24 ^c	Anzahl Patient:innen n (%)	Anzahl Patient:innen n (%)	Hodges-Lehmann Location Shift (ASE) [95 %-KI] ^d	
Krankheitsschwere				
WHO FK II	79 (48,5)	78 (48,8)	21,6 (7,64) [6,67; 36,60]	0,002
WHO FK III	84 (51,5)	82 (51,3)	60,9 (10,43) [40,46; 81,35]	

a: Datenschnitt: 06. Dezember 2022
b: Anzahl der Patient:innen: Full-Analysis-Set
c: Für Patient:innen mit fehlenden Werten für den 6MWD zu Woche 24 wegen sonstigen Gründen außer aufgrund von Tod oder einem nicht-tödlichen klinischem Ereignis wird eine Standard Multiple Imputation-Methode verwendet. Für Patient:innen mit fehlendem Wert aufgrund von Tod wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -2000 Metern zugeordnet um den schlechtestmöglichen Rang zu erhalten. Für Patient:innen mit fehlendem Wert aufgrund eines nicht-tödlichen klinischen Ereignisses wurde als Änderung gegenüber Baseline zu Woche 24 ein Wert von -1000 Metern zugeordnet um den nächstniedrigeren Rang zu erhalten.
d: Hodges-Lehman Location Shift im Vergleich zum Placebo-Schätzer, Median von allen gepaarten Differenzen zwischen den beiden Behandlungsarmen
e: Cochran's Q (Chi-Quadrat) zweiseitiger p-Wert für die Heterogenität zwischen den Subgruppen
6MWD: 6-Minuten Gehstrecke; ASE: Asymptotischer Standardfehler; FK: Funktionsklasse; KI: Konfidenzintervall; WHO: World Health Organization

Für die in der Studie STELLAR präspezifizierte Analyse der Veränderung der 6MWD lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Krankheitsschwere (WHO-FK II vs. WHO-FK III) feststellen ($p = 0,002$). Aufgrund des gleichgerichteten Effekts in beiden Subgruppenmerkmalen und der Übereinstimmung der Effektrichtung zugunsten der Behandlung mit Sotatercept mit der Hauptanalyse, wird von keiner relevanten Effektmodifikation ausgegangen. Diese Einschätzung deckt sich mit dem Ergebnis der Responderanalyse, welche keine Effektmodifikation nahelegt (siehe Tabelle 4-70).

Insgesamt wird daher die beobachtete Effektmodifikation als irrelevant erachtet.

Zeit bis zur klinischen Verschlechterung oder Tod

Es traten keine positiven Interaktionstests auf.

Verbesserung der Dyspnoe gemäß Borg CR10-Skala

Es traten keine positiven Interaktionstests auf.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Krankheitssymptomatik und Gesundheitszustand

Tabelle 4-83: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Krankheitssymptomatik und Gesundheitszustand (EQ-5D VAS) aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo			p-Wert für Interaktions- ^g Test
	Patient:innen mit Ereignis		Patient:innen mit Ereignis		Relatives Risiko/ Peto-Odds Ratio ^d		Adjustierte Risikodifferenz ^f	
Verbesserung des EQ-5D VAS ^b	N ^c	n (%)	N ^c	n (%)	[95 %-KI]	p-Wert ^e	[95 %-KI]	
Krankheitsschwere								
WHO FK II	65	16 (24,6)	65	5 (7,7)	3,20 [1,24; 8,26]	0,016	17,04 [4,53; 29,94]	0,026
WHO FK III	59	13 (22,0)	61	15 (24,6)	0,91 [0,48; 1,73]	0,764	-2,34 [-17,60; 13,14]	
PAH Hintergrundtherapie								
Mono-/Zweifachtherapie	49	6 (12,2)	45	9 (20,0)	0,60 [0,24; 1,55]	0,296	-8,01 [-23,93; 7,11]	0,020
Dreifachtherapie	75	23 (30,7)	81	11 (13,6)	2,29 [1,19; 4,43]	0,014	17,22 [4,33; 29,99]	

a: Datenschnitt: 06. Dezember 2022
 b: Erfüllt, sofern ein:e Patientin eine Verbesserung gegenüber Baseline für den EQ-5D VAS zu Woche 24 erreicht (Verbesserung um $\geq 15\%$)
 c: Anzahl Patient:innen:
 • mit Erhebung sowohl zu Baseline als auch zu Woche 24
 • sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19.
 • Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt.
 Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.
 d: Peto-Odds Ratio anstelle des Relativen Risikos wenn die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 e: Zweiseitiger p-Wert (Wald-Test)
 f: Miettinen und Nurminen Methode stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 g: Basierend auf einem linearen Modell mit Behandlung und Subgruppe als Kovariate sowie Behandlung nach Subgruppen Interaktion (p-Wert des Likelihood Ratio Tests) stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)
 COVID-19: Coronavirus Disease of 2019; EQ-5D: European Quality of Life 5 Dimensions; FK: Funktionsklasse; KI: Konfidenzintervall; PAH: Pulmonale arterielle Hypertonie ; VAS: Visuelle Analogskala; WHO: World Health Organization

Für den Endpunkt Krankheitssymptomatik und Gesundheitszustand für die VAS des EQ-5D lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Krankheitsschwere (WHO-FK II vs. WHO-FK III) ($p = 0,026$) sowie nach PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) ($p = 0,020$) feststellen.

Für die Krankheitsschwere zeigt sich ein statistisch signifikanter Unterschied in der WHO-FK II (RR [95 %-KI]: 3,20 [1,24; 8,26]; $p = 0,016$) zugunsten der Behandlung mit Sotatercept gegenüber Placebo. In der WHO-FK III zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm. Dies entspricht dem Ergebnis der Gesamtpopulation.

Für die PAH-Hintergrundtherapie zeigt sich ein statistisch signifikanter Unterschied für die Patient:innen unter Dreifachtherapie (RR [95 %-KI]: 2,29 [1,19; 4,43]; $p = 0,014$) zugunsten der Behandlung mit Sotatercept gegenüber Placebo. Für die Patient:innen unter Mono/Zweifachtherapie zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm. Dies entspricht dem Ergebnis der Gesamtpopulation.

Die beiden vorliegenden positiven Interaktionstests zeigen sich in keinen weiteren Endpunkten zur Symptomatik oder Lebensqualität. Daher wird von keiner relevanten Effektmodifikation ausgegangen.

Gesundheitsbezogene Lebensqualität

Gesundheitsbezogene Lebensqualität

Es traten keine positiven Interaktionstests auf.

Nebenwirkungen

Unerwünschte Ereignisse Gesamtraten

Es traten keine positiven Interaktionstests auf.

Unerwünschte Ereignisse (gegliedert nach SOC und PT)

Tabelle 4-84: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) für SOC aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		p-Wert für Interaktionstest ^e
	N ^b	Patient:innen n mit Ereignis n (%)	N ^b	Patient:innen n mit Ereignis n (%)	Relatives Risiko [95 %-KI] ^c	p-Wert ^d	
SOC ^f : Erkrankungen des Blutes und des Lymphsystems							
Geschlecht							
Männlich	34	8 (23,53)	33	1 (3,03)	7,76 [1,03; 58,70]	0,015	0,013

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		p-Wert für Interaktionstest ^e
	N ^b	Patient:innen mit Ereignis n (%)	N ^b	Patient:innen mit Ereignis n (%)	Relatives Risiko [95 %-KI] ^c	p-Wert ^d	
Weiblich	129	21 (16,28)	127	24 (18,90)	0,86 [0,51; 1,47]	0,583	
PAH-Hintergrundtherapie							
Mono/Zweifach	65	7 (10,77)	60	12 (20,00)	0,54 [0,23; 1,28]	0,153	0,028
Dreifach	98	22 (22,45)	100	13 (13,00)	1,73 [0,92; 3,23]	0,082	
SOC^f: Skelettmuskulatur-, Bindegewebs- und Knochenkrankungen							
Region ^g							
WHO Stratum A	135	49 (36,30)	127	26 (20,47)	1,77 [1,18; 2,67]	0,005	0,045
RoW	28	1 (3,57)	33	4 (12,12)	0,29 [0,03; 2,49]	0,229	
PAH-Hintergrundtherapie							
Mono/Zweifach	65	15 (23,08)	60	15 (25,00)	0,92 [0,49; 1,72]	0,802	0,021
Dreifach	98	35 (35,71)	100	15 (15,00)	2,38 [1,39; 4,07]	0,001	
a: Datenschnitt: 06. Dezember 2022							
b: Anzahl der Patient:innen: Safety Set Population							
c: Basierend auf einer Vierfeldertafel mit Wald Konfidenzintervall. Falls in einem der Behandlungsarme keine Patient:innen oder in beiden Behandlungsarmen kein Ereignis auftritt, wird "n.a." berichtet							
d: Basierend auf einem Cochran-Mantel-Haenszel Test. Falls in einem der Behandlungsarme keine Patient:innen ein Ereignis aufweist oder in beiden Behandlungsarmen kein Ereignis auftritt, wird "n.a." berichtet							
e: Basierend auf einem Breslow-Day Test mit Subgruppen als Stratifizierungsfaktor. Falls in einer Subgruppenkategorie kein Ereignis auftrat oder alle Patient:innen einer Subgruppenkategorie Ereignisse aufweisen wird die Subgruppenkategorie vom Interaktionstest ausgeschlossen, wenn nur noch eine Subgruppenkategorie verbleibt wird kein Interaktionstest durchgeführt.							
f: Nicht übersetzen							
g: WHO Stratum A umfasst folgende Länder: Andorra, Australien, Österreich, Belgien, Brunei Darussalam, Kanada, Kroatien, Kuba, Zypern, Tschechische Republik, Dänemark, Finnland, Frankreich, Deutschland, Griechenland, Island, Israel, Italien, Japan, Luxemburg, Malta, Monaco, Niederlande, Neuseeland, Norwegen, Portugal, San Marino, Singapur, Slowenien, Spanien, Schweden, Schweiz, Vereinigtes Königreich und Vereinigte Staaten von Amerika							
FK: Funktionsklasse; KI: Konfidenzintervall; n.a.: nicht anwendbar; PAH: Pulmonale arterielle Hypertonie; RoW: Rest der Welt; SOC: Systemorganklasse; WHO: World Health Organization							

Für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für die SOC Erkrankungen des Blutes und des Lymphsystems lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Geschlecht (männlich vs. weiblich) ($p = 0,013$) sowie nach PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) ($p = 0,028$) feststellen.

Für das Geschlecht zeigt sich ein statistisch signifikanter Unterschied für die männlichen Patienten (RR [95 %-KI]: 7,76 [1,03; 58,70]; $p = 0,015$) zuungunsten der Behandlung mit Sotatercept gegenüber Placebo. Dieser ist jedoch nicht fazitrelevant. Für die Patientinnen zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und dem Kontrollarm.

Für die PAH-Hintergrundtherapie zeigt sich in beiden Gruppen kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm.

Daher wird insgesamt in beiden Subgruppenanalysen die beobachtete Effektmodifikation als irrelevant betrachtet.

Für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für die SOC Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Region (WHO Stratum A vs. RoW ($p = 0,045$) sowie nach PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie) ($p = 0,021$) feststellen.

Für die Region zeigt sich ein statistisch signifikanter Unterschied für die Patient:innen in WHO-Stratum A (RR [95 %-KI]: 1,77 [1,18; 2,67]; $p = 0,005$) zuungunsten der Behandlung mit Sotatercept gegenüber Placebo. Für die Patient:innen der restlichen Welt zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm.

Für die PAH-Hintergrundtherapie zeigt sich ein statistisch signifikanter Unterschied für die Patient:innen unter Dreifachtherapie (RR [95 %-KI]: 2,38 [1,39; 4,07]; $p = 0,001$) zuungunsten der Behandlung mit Sotatercept gegenüber Placebo. Für die Patient:innen unter Mono/Zweifachtherapie zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm.

Aufgrund der fehlenden medizinischen Rationale wird in beiden Subgruppenanalysen die beobachtete Effektmodifikation als irrelevant betrachtet.

Tabelle 4-85: Subgruppenanalysen mit positivem Interaktionstest ($p < 0,05$) für den Endpunkt Unerwünschte Ereignisse gesamt (SOC und PT) für PT aus RCT mit dem zu bewertenden Arzneimittel (STELLAR)

Studie: STELLAR (MK-7962-003) ^a	Sotatercept		Placebo		Sotatercept vs. Placebo		p-Wert für Interaktionstest ^e
	N ^b	Patient:innen mit Ereignis n (%)	N ^b	Patient:innen mit Ereignis n (%)	Relatives Risiko [95 %-KI] ^c	p-Wert ^d	
SOC: Erkrankungen des Nervensystems - PT^f: Schwindelgefühl							
Alter (Jahre)							
< 65	138	17 (12,32)	131	10 (7,63)	1,61 [0,77; 3,39]	0,202	0,025
≥ 65	25	7 (28,00)	29	0	n.a. [n.a.; n.a.]	n.a.	
SOC: Erkrankungen der Atemwege, des Brustraums und des Mediastinums - PT^f: Nasenbluten							
Region ^g							
WHO Stratum A	135	35 (25,93)	127	2 (1,57)	16,46 [4,04; 67,04]	<0,001	0,024
RoW	28	1 (3,57)	33	1 (3,03)	1,18 [0,08; 17,99]	0,907	
a: Datenschnitt: 06. Dezember 2022							
b: Anzahl der Patient:innen: Safety Set Population							
c: Basierend auf einer Vierfeldertafel mit Wald Konfidenzintervall. Falls in einem der Behandlungsarme keine Patient:innen oder in beiden Behandlungsarmen kein Ereignis auftritt, wird "n.a." berichtet							
d: Basierend auf einem Cochran-Mantel-Haenszel Test. Falls in einem der Behandlungsarme kein:e Patient:innen ein Ereignis aufweist oder in beiden Behandlungsarmen kein Ereignis auftritt, wird "n.a." berichtet							
e: Basierend auf einem Breslow-Day Test mit Subgruppen als Stratifizierungsfaktor. Falls in einer Subgruppenkategorie kein Ereignis auftrat oder alle Patient:innen einer Subgruppenkategorie Ereignisse aufweisen wird die Subgruppenkategorie vom Interaktionstest ausgeschlossen, wenn nur noch eine Subgruppenkategorie verbleibt wird kein Interaktionstest durchgeführt.							
f: Nicht übersetzen							
g: WHO Stratum A umfasst folgende Länder: Andorra, Australien, Österreich, Belgien, Brunei Darussalam, Kanada, Kroatien, Kuba, Zypern, Tschechische Republik, Dänemark, Finnland, Frankreich, Deutschland, Griechenland, Island, Israel, Italien, Japan, Luxemburg, Malta, Monaco, Niederlande, Neuseeland, Norwegen, Portugal, San Marino, Singapur, Slowenien, Spanien, Schweden, Schweiz, Vereinigtes Königreich und Vereinigte Staaten von Amerika							
KI: Konfidenzintervall; n.a.: nicht anwendbar; PT: Preferred Terms; RoW: Rest der Welt; SOC: Systemorganklasse; WHO: World Health Organization							

Für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für den PT Schwindelgefühl innerhalb der SOC Erkrankungen des Nervensystems lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Alter (< 65 Jahre vs. ≥ 65 Jahre) ($p = 0,025$) feststellen.

Für das Alter zeigt sich in beiden Gruppen kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm.

Daher wird die beobachtete Effektmodifikation als irrelevant betrachtet.

Für den Endpunkt Unerwünschte Ereignisse (gegliedert nach SOC und PT) für den PT Nasenbluten innerhalb der SOC Erkrankungen der Atemwege, des Brustraums und Mediastinums lässt sich ein statistisch signifikanter Interaktionstest in der Subgruppe nach Region (WHO Stratum A vs. RoW) ($p = 0,024$) feststellen.

Für die Region zeigt sich ein statistisch signifikanter Unterschied für die Patient:innen in WHO Stratum A (RR [95 %-KI]: 16,46 [4,04; 67,04]; $p = < 0,001$) zuungunsten der Behandlung mit Sotatercept gegenüber Placebo. Für die Patient:innen der restlichen Welt zeigt sich kein statistisch signifikanter Unterschied zwischen dem Interventions- und Kontrollarm.

Aufgrund der fehlenden medizinischen Rationale wird in beiden Subgruppenanalysen die beobachtete Effektmodifikation als irrelevant betrachtet.

Schwerwiegende unerwünschte Ereignisse (gegliedert nach SOC und PT)

Es wurden keine Subgruppenanalysen durchgeführt, da auf Gesamtstudienoberfläche keine statistisch signifikanten Effekte vorlagen.

Schwere unerwünschte Ereignisse (gegliedert nach SOC und PT)

Es wurden keine Subgruppenanalysen durchgeführt, da auf Gesamtstudienoberfläche keine statistisch signifikanten Effekte vorlagen.

4.3.1.4 Liste der eingeschlossenen Studien – RCT

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).

Studie	Studienbericht	Einträge in Studienregistern	Publikationen
STELLAR	(36)	ClinicalTrials.gov: NCT04576988 (32) EU-CTR: 2020-004142-11 (33) ICTRP: NCT04576988 (34) EUCTR2020-004142-11 (35)	(31)
EU-CTR: European Union Clinical Trials Register; ICTRP: International Clinical Trials Registry Platform			

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 sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der Internetseite des G-BA*
- *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 identifiziert wurden und bewerten Sie darüber hinaus deren Ähnlichkeit. Begründen Sie darauf basierend den Ein- bzw. Ausschluss von Studien für die von Ihnen durchgeführten indirekten Vergleiche. Bewerten

*Sie das Verzerrungspotenzial der für indirekte Vergleiche herangezogenen Studien. **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 4-86: 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 drei 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 4-87: 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 4-88: 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 4-89: 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 4-90: 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 der üblichen Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Stellen Sie die Ergebnisse der paarweisen Meta-Analysen dar. 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 inkonsistente Ergebnisse.*

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 4.3.1.3.2.

Nicht zutreffend.

4.3.2.1.4 Liste der eingeschlossenen Studien – indirekte Vergleiche aus RCT

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).

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/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der G-BA Internetseite*
- *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 4-91: 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

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

Tabelle 4-92: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien

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

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 4-93: 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 4-94: 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 4.3.1.3.2.***

Nicht zutreffend.

4.3.2.2.4 Liste der eingeschlossenen Studien – nicht randomisierte vergleichende 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).

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 sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). Benennen Sie für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie***

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der G-BA Internetseite*

- *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

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

Tabelle 4-95: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen

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

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 4-96: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt. 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.

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.

4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen

*Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. **Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.***

Nicht zutreffend.

4.3.2.3.4 Liste der eingeschlossenen Studien – weitere Untersuchungen

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).

Nicht zutreffend.

4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens

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.

Die Ermittlung des Ausmaßes des medizinischen Zusatznutzens von Sotatercept erfolgt auf Grundlage der doppelblinden, multizentrischen, randomisierten und Placebo-kontrollierten Zulassungsstudie STELLAR.

Die Studie STELLAR untersucht die Wirksamkeit und Sicherheit von Sotatercept gegenüber Placebo in Kombination mit anderen Therapien gegen PAH, für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit. Neben der 6-Minuten-Gehstrecke als primärer Endpunkt der Studie stehen auch weitere patientenrelevante Endpunkte zur Verfügung und ermöglichen somit eine valide Beurteilung des medizinischen Zusatznutzens von Sotatercept.

MSD folgt der Entscheidung des G-BA und wählt eine patientenindividuell optimierte medikamentöse Therapie für die unter Anwendungsgebiet A betrachtete Patientenpopulation als zVT (siehe auch Abschnitt 3.1.2 im Modul 3A). Damit erbringt MSD für das Anwendungsgebiet den Nachweis des medizinischen Nutzens und medizinischen Zusatznutzens gegenüber der vom G-BA definierten zVT anhand eines direkten Vergleichs (siehe Abschnitt 4.3.1.3).

Gemäß der VerfO des G-BA entspricht die Studie STELLAR der Evidenzstufe Ib (2). Das Verzerrungspotenzial der Studie wurde als niedrig eingeschätzt, sodass von einer hohen Studienqualität ausgegangen werden kann (siehe Ausführungen in Abschnitt 4.3.1.2.2).

Insgesamt erlaubt die Studie STELLAR einen aussagekräftigen Nachweis über den Zusatznutzen von Sotatercept und die Aussagesicherheit der Ergebnisse ist als hoch einzustufen. Demzufolge lässt sich auf deren Basis ein **Hinweis** für den Zusatznutzen von Sotatercept ableiten.

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 4.4.1 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.

Tabelle 4-97: Übersicht über die zur Ableitung des Zusatznutzens herangezogenen Endpunkte der Studie STELLAR

Studie: STELLAR ^a	Sotatercept			Placebo			Sotatercept vs. Placebo		Zusatznutzen
	N	Mediane Patienten mit Ereignis n (%)	Mediane Ereigniszeit [95 %-KI]	N	Mediane Patienten mit Ereignis n (%)	Mediane Ereigniszeit [95 %-KI]	Effekt-schätzer [95 %-KI]	p-Wert	
Mortalität									
Mortalität zu Woche 24	163 ^b	0 (0,0)	n.a.	160 ^b	6 (3,8)	n.a.	Peto-OR ^c : 0,13 [0,03; 0,64]	0,013 ^d	Hinweis auf einen erheblichen Zusatznutzen
Morbidität									
Verbesserung der 6MWD	163 ^c	72 (44,2)	n.a.	159 ^e	29 (18,2)	n.a.	RR ^c : 2,40 [1,66; 3,48]	< 0,001 ^d	Hinweis auf einen erheblichen Zusatznutzen
Zeit bis zur klinischen Verschlechterung	163 ^b	11 (6,7)	Nicht erreicht [-; -] ^e	160 ^b	42 (26,3)	78,00 [-; -] ^e	HR ^g : 0,23 [0,12; 0,44]	< 0,001 ^{g,h}	Hinweis auf einen erheblichen Zusatznutzen
Verbesserung der Dyspnoe gemäß Borg CR10-Skala	160 ^f	38 (23,8)	n.a.	159 ^f	37 (23,3)	n.a.	RR ^c : 1,02 [0,69; 1,51]	0,918 ^d	Zusatznutzen nicht belegt
Krankheitssymptomatik und Gesundheitszustand									
Verbesserung der PAH-SYMPACT Domäne: kardiopulmonale Symptome	115 ^f	47 (40,9)	n.a.	117 ^f	35 (29,9)	n.a.	RR ^c : 1,35 [0,95; 1,93]	0,095 ^d	Zusatznutzen nicht belegt
Verbesserung der PAH-SYMPACT Domäne: kardiovaskuläre Symptome	115 ^f	49 (42,6)	n.a.	117 ^f	34 (29,1)	n.a.	RR ^c : 1,48 [1,04; 2,11]	0,030 ^d	Hinweis auf einen geringen Zusatznutzen
Verbesserung des EQ-5D VAS	124 ^f	29 (23,4)	n.a.	126 ^f	20 (15,9)	n.a.	RR ^c : 1,49 [0,89; 2,49]	0,131 ^d	Zusatznutzen nicht belegt
Gesundheitsbezogene Lebensqualität									
Verbesserung der PAH-SYMPACT Domäne: Auswirkungen auf die körperliche Funktion	117 ^f	39 (33,3)	n.a.	123 ^f	31 (25,2)	n.a.	RR ^c : 1,31 [0,87; 1,96]	0,193 ^d	Zusatznutzen nicht belegt
Verbesserung der PAH-SYMPACT Domäne: kognitive/emotionale Auswirkungen	117 ^f	30 (25,6)	n.a.	123 ^f	30 (24,4)	n.a.	RR ^c : 1,04 [0,67; 1,60]	0,866 ^d	Zusatznutzen nicht belegt
Nebenwirkungen									
Unerwünschte Ereignisse Gesamtraten									
Schwerwiegende unerwünschte Ereignisse	163 ⁱ	40 (24,54)	n.a.	160 ⁱ	47 (29,38)	n.a.	RR ⁱ : 0,84 [0,58; 1,20]	0,3281 ^k	Zusatznutzen nicht belegt
Schwere unerwünschte Ereignisse	163 ⁱ	24 (14,72)	n.a.	160 ⁱ	33 (20,63)	n.a.	RR ⁱ : 0,71 [0,44; 1,15]	0,1649 ^k	Zusatznutzen nicht belegt

Studie: STELLAR ^a	Sotatercept			Placebo			Sotatercept vs. Placebo	Zusatznutzen
Therapieabbruch wegen unerwünschter Ereignisse	163 ⁱ	6 (3,68)	n.a.	160 ⁱ	11 (6,88)	n.a.	RR ^j : 0,1994 ^k 0,54 [0,20; 1,41]	
<p>a: Datenschnitt: 06. Dezember 2022</p> <p>b: Anzahl der Patient:innen: Full-Analysis-Set</p> <p>c: Peto-Odds Ratio anstelle des Mantel-Haenszel relativen Risikos, sofern die Inzidenz $\leq 1\%$ oder $\geq 99\%$ in mindestens einer Zelle beträgt, stratifiziert nach WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie)</p> <p>d: Zweiseitiger p-Wert (Wald-Test)</p> <p>e: Anzahl der Patient:innen: Full Analysis Set Population. Patient:innen mit fehlender Erhebung zu Woche 24 und nicht aufgrund von COVID-19 werden als Non-Responder gewertet. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden ausgeschlossen</p> <p>f: Anzahl Patient:innen:</p> <ul style="list-style-type: none"> mit Erhebung sowohl zu Baseline als auch zu Woche 24 sowie Patient:innen mit Erhebung zu Baseline und fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19. Patient:innen mit fehlender Erhebung zu Woche 24 aufgrund von COVID-19 werden nicht berücksichtigt. <p>Die Patient:innen mit fehlender Erhebung zu Woche 24 nicht aufgrund von COVID-19 werden als Non-Responder gewertet.</p> <p>g: Basierend auf einem Cox Regressionsmodell mit Behandlung als Kovariate, stratifiziert nach folgenden Kovariaten: WHO-Funktionsklasse (II vs. III) und PAH-Hintergrundtherapie (Mono/Zweifach- vs. Dreifachtherapie)</p> <p>h: Zweiseitiger p-Wert (Wald-Test; Score-Test im Fall von null Ereignissen in einem der Studienarme)</p> <p>i: Anzahl der Patient:innen: Safety Set Population</p> <p>j: 2x2 Kontingenztabelle mit Konfidenzintervall nach Wald. Bei keinem:keiner Patient:Patientin mit Ereignis in einer der Behandlungsgruppen wird eine Korrektur auf alle Zellen der 2x2 Kontingenztabelle, um den Faktor +0,5 und ein Konfidenzintervall nach Wald angewendet. Wenn in beiden Behandlungsgruppe kein Ereignis oder bei allen Patient:innen ein Ereignis auftritt, wird n.a. berichtet</p> <p>k: Basierend auf einen Cochran-Mantel-Haenszel Test. Für den Fall, dass entweder keine Ereignisse aufgetreten sind oder alle Patient:innen ein Ereignis in beiden Behandlungsarmen aufweisen, wird n.a. berichtet</p> <p>KI: Konfidenzintervall; n.a.: nicht anwendbar; PAH: Pulmonale arterielle Hypertonie ; WHO: World Health Organization; COVID-19: Coronavirus Disease of 2019; 6MWD: 6-Minuten Gehstrecke; G-BA: Gemeinsamer Bundesausschuss; PAH-SYMPACT: Pulmonale arterielle Hypertonie-Symptome und Auswirkungen; EQ-5D: European Quality of Life 5 Dimensions; VAS: Visuelle Analogskala</p>								

Mortalität

Gesamtüberleben

Für den Endpunkt Mortalität zu Woche 24 zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (Peto OR [95 %-KI]: 0,13 [0,03; 0,64]; $p = 0,013$), dies ist eine erhebliche Reduktion des Mortalitätsrisikos.

Für den Endpunkt Mortalität zu Woche 24 ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Morbidität

Verbesserung der 6MWD

Für den Endpunkt Verbesserung der 6MWD zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (RR [95 %-KI]: 2,40 [1,66; 3,48]; $p = < 0,001$). 44,2 % der Patient:innen im Sotatercept-Arm konnten ihre innerhalb von 6 Minuten zurückgelegte Gehdistanz um mindestens 40 Meter gegenüber ihrem Baseline-Wert verbessern, im Placebo-Arm lediglich 18,2 %. Es ist anzumerken, dass diese bisher nicht erreichte, erhebliche

Verbesserung der körperlichen Belastbarkeit unter Verwendung einer MID erreicht wird, welche die in der Fachliteratur als klinisch relevant erachtete MID von 33 Metern weit überschreitet. Darüber hinaus ist festzuhalten, dass es sich in der Studie STELLAR durch das Add-on Design um bereits therapierte Patient:innen handelt.

Für den Endpunkt Verbesserung der 6MWD ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Zeit bis zur klinischen Verschlechterung oder Tod

Für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (HR [95 %-KI]: 0,23 [0,12; 0,44]; $p = < 0,001$). Das Risiko zu versterben oder eine klinische Verschlechterung zu erfahren, beträgt etwa ein Viertel für die Patient:innen im Sotatercept-Arm im Vergleich zum Placebo-Arm.

Für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test

Für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala pre 6MWD-Test (Verbesserung der Dyspnoe gemäß Borg CR10-Skala um 15 % der Skalenspannweite) zeigt sich kein Unterschied zwischen Sotatercept und Placebo.

Für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala ist der **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Krankheitssymptomatik und Gesundheitszustand

Für die Krankheitssymptomatik anhand des Symptom-Items kardiopulmonale Symptome des PAH-SYMPACT und Gesundheitszustand (anhand der EQ-5D VAS) sind positive Effekte der Behandlung mit Sotatercept zu beobachten, die in den Responderanalysen jedoch keine statistische Signifikanz erreichen. Für die Krankheitssymptomatik anhand des Symptom-Items kardiovaskuläre Symptome des PAH-SYMPACT zeigt sich ein statistisch signifikanter Unterschied zugunsten der Behandlung mit Sotatercept (RR [95 %-KI]: 1,48 [1,04; 2,11]; $p = 0,030$).

Für den Endpunkt Krankheitssymptomatik und Gesundheitszustand ist das Verzerrungspotenzial als niedrig zu bewerten. In der Gesamtschau liegen positive Effekte und ein Zusatznutzen hinsichtlich der kardiovaskulären Symptome vor.

Somit ist für die patientenberichtete Krankheitssymptomatik und den Gesundheitszustand anhand des Symptom-Items kardiopulmonale Symptome des PAH-SYMPACT und Gesundheitszustand (anhand der EQ-5D VAS) ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Für den Endpunkt Krankheitssymptomatik anhand des Symptom-Items kardiovaskuläre Symptome des PAH-SYMPACT ist das Verzerrungspotenzial als niedrig zu bewerten und es ergibt sich somit ein **Hinweis auf einen geringen Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie.

Gesundheitsbezogene Lebensqualität

Gesundheitsbezogene Lebensqualität

In Bezug auf die Auswirkungen auf die körperliche Funktion in Alltagssituationen berichten Patient:innen unter Sotatercept häufiger eine Verbesserung als im Placebo-Arm. Diese erreicht jedoch keine statistische Signifikanz. Für das Impact-Item „Kognitive/Emotionale Auswirkungen“ lässt sich kein Unterschied zwischen der Behandlung mit Sotatercept und Placebo feststellen.

Für den Endpunkt Gesundheitsbezogene Lebensqualität (anhand der Impact-Items des PAH-SYMPACT) ist das Verzerrungspotenzial als niedrig zu bewerten und in der Gesamtschau liegt kein fazitrelevanter Zusatznutzen vor. Somit ist für die Gesundheitsbezogene Lebensqualität ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Nebenwirkungen

Unerwünschte Ereignisse Gesamtraten

In der Studie STELLAR zeigt sich bei den Unerwünschten Ereignissen Gesamtraten kein Unterschied zwischen Sotatercept und Placebo.

Für den Endpunkt Unerwünschte Ereignisse Gesamtraten ist das Verzerrungspotenzial als niedrig zu bewerten. Somit ist für den Endpunkt Unerwünschte Ereignisse Gesamtraten ein **Zusatznutzen** von Sotatercept gegenüber Placebo, jeweils kombiniert mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie, **nicht belegt**.

Fazit

Die Ableitung des Zusatznutzens für Sotatercept, in Kombination mit anderen Therapien gegen PAH, für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit, basiert auf den Ergebnissen der randomisierten, doppelblinden und Placebo-kontrollierten Studie STELLAR. Diese Studie umfasst Patient:innen der WHO-FK II und III, behandelt mit stabiler, patientenindividuell optimierten PAH-Hintergrundtherapie als Mono- (4,0 %), Zweifachkombinations- (34,7 %)

oder Dreifachkombinationstherapie (61,3 %). Etwa 40 % der Studienteilnehmer:innen erhielten aufgrund des bereits fortgeschrittenen Krankheitsstadiums eine parenterale Prostazyklin-Therapie. Bei der Behandlung dieser Patient:innen wurden bereits alle weiteren verfügbaren Wirkansätze ausgeschöpft.

Die PAH ist trotz der aktuell verfügbaren Behandlungsoptionen nach wie vor eine rasch fortschreitende, lebenszeitverkürzende und letztlich tödlich verlaufende Erkrankung, die für betroffene Patient:innen mit einem hohen Leidensdruck einhergeht. Im Unterschied zu den aktuell verfügbaren Substanzklassen steht mit Sotatercept nun erstmals ein Wirkstoff zur Verfügung, der über die Behandlung der Symptome der PAH hinaus auch kausal in die Pathophysiologie der PAH eingreift. Dabei steuert Sotatercept einen neuen, bisher nicht verfolgten, 4. Signalweg an und ermöglicht sogar bei Patient:innen, die bisher verfügbare Therapieoptionen ausgeschöpft haben, umfangreiche klinisch relevante und für die Patient:innen unmittelbar spürbare Verbesserungen.

Durch eine Behandlung mit Sotatercept wird eine nachhaltige und bisher nicht erreichte große Verbesserung des therapielevanten Nutzens gegenüber der zVT erzielt. Hervorzuheben ist hier einerseits die erhebliche Verbesserung der körperlichen Belastbarkeit der Patient:innen, welche sich unmittelbar in alltäglichen Funktionen und Aktivitäten widerspiegelt und andererseits die nachhaltige, erhebliche Reduktion des Mortalitätsrisikos. Mit Sotatercept kann nicht nur eine Verbesserung, sondern auch eine Verzögerung dieser progredienten chronischen Erkrankung erreicht werden. Die Patient:innen erleiden erheblich seltener eine klinische Verschlechterung.

Zusammenfassend handelt es sich bei Sotatercept, in Kombination mit einer patientenindividuell optimierten medikamentösen PAH-Hintergrundtherapie., für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-FK II bis III zur Verbesserung der körperlichen Leistungsfähigkeit, um eine hochwirksame und gut verträgliche Therapieoption. In der Gesamtschau ergibt sich somit ein **Hinweis auf einen erheblichen Zusatznutzen** von Sotatercept im Vergleich zur zVT.

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 4-98: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
WINREVAIR® ist, in Kombination mit anderen Therapien gegen pulmonale arterielle Hypertonie (PAH), für die Behandlung von PAH bei erwachsenen Patienten mit der WHO-Funktionsklasse (FK) II bis III zur Verbesserung der körperlichen Leistungsfähigkeit angezeigt (siehe Abschnitt 5.1 der Fachinformation).	Hinweis auf einen erheblichen Zusatznutzen
FK: Funktionsklasse; PAH: ; WHO: World Health Organization (Weltgesundheitsorganisation)	

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 nicht randomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nicht randomisierte 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 2005¹⁷, Molenberghs 2010¹⁸). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrundeliegende 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 2006¹⁹) 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 2006²⁰) 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

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.

Nicht zutreffend.

¹⁷ Burzykowski T (Ed.): The evaluation of surrogate endpoints. New York: Springer; 2005.

¹⁸ 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.

¹⁹ Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.

²⁰ Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

4.6 Referenzliste

Listen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregister-einträge), 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. Gemeinsamer Bundesausschuss (G-BA). Niederschrift zum Beratungsgespräch gemäß § 8 AM-NutzenV. Beratungsanforderung 2022-B-349. Sotatercept zur Behandlung der Pulmonal Arteriellen Hypertonie (PAH) bei erwachsenen Patienten mit WHO Funktionsklasse II und III. 2023.
2. Gemeinsamer Bundesausschuss (G-BA). Verfahrensordnung des Gemeinsamen Bundesausschusses. 2024.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden Version 7.02023 11.07.2024. Available from: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.
4. Merck Sharp & Dohme B.V. NL-Haarlem. Fachinformation Winrevair® (Sotatercept) 45 mg/- 60 mg Pulver und Lösungsmittel zur Herstellung einer Injektionslösung. Stand der Information: August 2024. 2024.
5. Committee for Medicinal Products for Human Use (CHMP). Guideline on the clinical investigations of medicinal products for the treatment of pulmonary arterial hypertension 2009 11.07.2024. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigations-medicinal-products-treatment-pulmonary-arterial-hypertension_en.pdf.
6. Wong SS-L, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. Journal of the Medical Library Association. 2006;94(4):451.
7. Gemeinsamer Bundesausschuss (G-BA). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Macitentan. 2014.
8. Gemeinsamer Bundesausschuss (G-BA). Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Riociguat. 2014.
9. Gilbert C, Brown MC, Cappelleri JC, Carlsson M, McKenna SP. Estimating a minimally important difference in pulmonary arterial hypertension following treatment with sildenafil. Chest. 2009;135(1):137-42.
10. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. American journal of respiratory and critical care medicine. 1997;155(4):1278-82.
11. Mathai SC, Puhan MA, Lam D, Wise RA. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. American journal of respiratory and critical care medicine. 2012;186(5):428-33.
12. Gabler NB, French B, Strom BL, Palevsky HI, Taichman DB, Kawut SM, et al. Validation of 6-minute walk distance as a surrogate end point in pulmonary arterial hypertension trials. Circulation. 2012;126(3):349-56.
13. Moutchia J, McClelland RL, Al-Naamani N, Appleby DH, Blank K, Grinnan D, et al. Minimal Clinically Important Difference in the 6-minute-walk Distance for Patients with

Pulmonary Arterial Hypertension. American journal of respiratory and critical care medicine. 2023;207(8):1070-9.

14. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden Version 6.1 2022 11.07.2024. Available from: https://www.iqwig.de/methoden/allgemeine-methoden-v6-1_abgeloest.pdf.

15. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Respir J. 2022;61(1).

16. Hoeper MM, Kramer T, Pan Z, Eichstaedt CA, Spiesshoefer J, Benjamin N, et al. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. Eur Respir J. 2017;50(2).

17. Opitz C, Blindt R, Blumberg F, Borst M, Bruch L, Leuchte H, et al. Pulmonale Hypertonie: invasive Diagnostik. DMW-Deutsche Medizinische Wochenschrift. 2010;135(S 03):S78-S86.

18. Chin KM, Gomberg-Maitland M, Channick RN, Cuttica MJ, Fischer A, Frantz RP, et al. Psychometric validation of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT) questionnaire: results of the SYMPHONY trial. Chest. 2018;154(4):848-61.

19. Rabin R, Charro Fd. EQ-SD: a measure of health status from the EuroQol Group. Annals of medicine. 2001;33(5):337-43.

20. McCollister D, Shaffer S, Badesch DB, Filusch A, Hunsche E, Schüler R, et al. Development of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT®) questionnaire: a new patient-reported outcome instrument for PAH. Respir Res. 2016;17(1):72.

21. Frantz RP, Chin KM, Zhao C, Flynn M, Badesch D. Pulmonary Arterial Hypertension—Symptoms and Impact Questionnaire: feasibility of utilizing one-day versus seven-day symptom reporting. Pulmonary Circulation. 2020;10(2):2045894020923957.

22. EuroQoL. EQ-5D Terminology 2024 11.07.2024. Available from: <https://euroqol.org/support/terminology/>.

23. Grünig E, Benjamin N, Krüger U, Kaemmerer H, Harutyunova S, Olsson K, et al. Allgemeine und supportive Therapie der pulmonal arteriellen Hypertonie: Empfehlungen der Kölner Konsensus Konferenz 2016. DMW-Deutsche Medizinische Wochenschrift. 2016;141(S 01):S26-S32.

24. Miettinen O, Nurminen M. Comparative analysis of two rates. Statistics in medicine. 1985;4(2):213-26.

25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995;57(1):289-300.

26. Mehrotra DV, Adewale AJ. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. Statistics in Medicine. 2012;31(18):1918-30.

27. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). IQWiG-Berichte – Nr. 678. Treatment Switching in onkologischen Studien. Version 1.02018 19.07.2024. Available from: https://www.iqwig.de/download/ga14-04_treatment-switching-in-onkologischen-studien_arbeitspapier_v1-0.pdf?rev=117386.

28. Medizinischer Dienst der Spitzenverbände der Krankenkassen e.V. (MDS). G-2 Gutachten: Aussagekraft von Subgruppenanalysen2004 11.07.2024. Available from: <https://md->

bund.de/fileadmin/dokumente/Publikationen/GKV/Begutachtungsgrundlagen_GKV/subgruppen-gutachten.pdf.

29. European Medicines Agency (EMA). Guideline on the investigation of subgroups in confirmatory clinical trials 2019 11.07.2024. Available from: <https://www.ema.europa.eu/en/investigation-subgroups-confirmatory-clinical-trials-scientific-guideline>.
30. Acceleron Pharma Inc. Clinical Study Protocol A011-12 (SOTERIA). An Open-label Long-term Follow-up Study to Evaluate the Effects of Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH. Version 6.0. 2024.
31. Hoepfer MM, Badesch DB, Ghofrani HA, Gibbs JSR, Gombert-Maitland M, McLaughlin VV, et al. Phase 3 Trial of Sotatercept for Treatment of Pulmonary Arterial Hypertension. *N Engl J Med*. 2023;388(16):1478-90.
32. ClinicalTrials.gov. NCT04576988: A Study of Sotatercept for the Treatment of Pulmonary Arterial Hypertension (MK-7962-003/A011-11) (STELLAR)2024 11.07.2024. Available from: <https://clinicaltrials.gov/study/NCT04576988?term=NCT04576988%20&rank=1>.
33. EU Clinical Trials Register. A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Compare the Efficacy and Safety of Sotatercept Versus Placebo When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH (EU-CTR 2020-004142-11)2021 11.07.2024. Available from: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-004142-11/DE>.
34. International Clinical Trials Registry Platform (ICTRP). NCT04576988: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Compare the Efficacy and Safety of Sotatercept Versus Placebo When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH2024 11.07.2024. Available from: <https://trialsearch.who.int/?TrialID=NCT04576988>.
35. International Clinical Trials Registry Platform (ICTRP). EUCTR2020-004142-11-SE: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Compare the Efficacy and Safety of Sotatercept Versus Placebo When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH - A Phase 3 Study of Sotatercept for the Treatment of PAH2021 11.07.2024. Available from: <https://trialsearch.who.int/?TrialID=EUCTR2020-004142-11-SE>.
36. Merck Sharp & Dohme LLC. Rahway, NJ, USA. Clinical Study Report: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Compare the Efficacy and Safety of Sotatercept Versus Placebo When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH 2023.
37. Gemeinsamer Bundesausschuss (G-BA). Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Verfahrensordnung: Änderung der Modulvorlage in der Anlage II zum 5. Kapitel. 2021.

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	07.11.2016	
Zeitsegment	1974 to 2016 November 04	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle ²¹] – Strategy minimizing difference between sensitivity and specificity	
#	Suchbegriffe	Ergebnis
1	Diabetes Mellitus/	552986
2	Non Insulin Dependent Diabetes Mellitus/	195234
3	(diabet* or niddm or t2dm).ab,ti.	714228
4	or/1-3	847068
5	linagliptin*.mp.	1562
6	(random* or double-blind*).tw.	1193849
7	placebo*.mp.	388057
8	or/6-7	1382838
9	and/4,5,8	633

²¹ 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.

Anhang 4-A1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Datenbankname	Embase Classic+Embase	
Suchoberfläche	Ovid	
Datum der Suche	09.07.2024	
Zeitsegment	1947 to 2024 July 08	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 – Strategy minimizing difference between sensitivity and specificity	
#	Suchbegriffe	Ergebnis
1	exp Hypertension, Pulmonary/	130601
2	(arterial or primary).mp.	3784847
3	1 and 2	53914
4	(pulmonary arterial hypertension or primary pulmonary hypertension or PPH or PHN or PAH or IPAH or HPAH or FPAH or DPAH).mp.	73202
5	3 or 4	95152
6	(sotatercept or ace-011 or ActRIIA-IgG1 or MK-7962).mp.	513
7	random*.tw. or placebo*.mp. or double-blind*.tw.	2398202
8	5 and 6 and 7	44

Datenbankname	Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions 1946 to July 08, 2024	
Suchoberfläche	Ovid	
Datum der Suche	09.07.2024	
Zeitsegment	1946 to July 08, 2024;	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 – Strategy minimizing difference between sensitivity and specificity	
#	Suchbegriffe	Ergebnis
1	exp Hypertension, Pulmonary/	43609
2	(arterial or primary).mp.	2545226
3	1 and 2	20278
4	(pulmonary arterial hypertension or primary pulmonary hypertension or PPH or PHN or PAH or IPAH or HPAH or FPAH or DPAH).mp.	50031
5	3 or 4	56232
6	(sotatercept or ace-011 or ActRIIA-IgG1 or MK-7962).mp.	137
7	randomi#ed controlled trial.pt. or randomi#ed.mp. or placebo.mp.	1196088
8	5 and 6 and 7	14

Datenbankname	EBM Reviews - Cochrane Central Register of Controlled Trials June 2024	
Suchoberfläche	Ovid	
Datum der Suche	09.07.2024	
Zeitsegment	June 2024;	
Suchfilter	kein Suchfilter verwendet	
#	Suchbegriffe	Ergebnis
1	exp Hypertension, Pulmonary/	1679
2	(arterial or primary).mp.	650511
3	1 and 2	1133
4	(pulmonary arterial hypertension or primary pulmonary hypertension or PPH or PHN or PAH or IPAH or HPAH or FPAH or DPAH).mp.	3972
5	3 or 4	4330
6	(sotatercept or ace-011 or ActRIIA-IgG1 or MK-7962).mp.	63
7	5 and 6	31

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/ Studienergebnisdatenbanken

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern/ Studienergebnisdatenbanken 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 jede/s durchsuchte Studienregister/ Studienergebnisdatenbank ist eine separate Strategie darzustellen. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters/ Studienergebnisdatenbank (z. B. [clinicaltrials.gov](http://www.clinicaltrials.gov)), die Internetadresse, unter der das/die Studienregister/ Studienergebnisdatenbank 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/ Studienergebnisdatenbank	International Clinical Trials Registry Platform Search Portal
Internetadresse	http://apps.who.int/trialsearch/
Datum der Suche	07.11.2016
Eingabeoberfläche	Standard Search
Suchstrategie	linagliptin OR BI 1356
Treffer	169

Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Studienregister	ClinicalTrials.gov	
Internetadresse	https://clinicaltrials.gov/ct2/search/advanced	
Datum der Suche	09.07.2024	
Eingabeoberfläche	Advanced Search	
Suchstrategie	Condition or disease:	
	Intervention/treatment:	sotatercept OR ace-011 OR ActRIIA-IgG1 OR MK-7962
	Other terms	
Treffer	25	

Studienregister	EU Clinical Trials Register
Internetadresse	https://www.clinicaltrialsregister.eu/ctr-search/search
Datum der Suche	09.07.2024
Suchstrategie	sotatercept OR ace-011 OR ActRIIA-IgG1 OR MK-7962
Treffer	10

Studienregister	WHO International Clinical Trial Registry Platform
Internetadresse	https://trialsearch.who.int/
Datum der Suche	09.07.2024
Eingabeoberfläche	Standard Search
Suchstrategie	sotatercept OR ace-011 OR ActRIIA-IgG1 OR MK-7962
Treffer	32

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

Nr.	Quelle	Ausschlussgrund
1.	Olsson K.M, Fuge J, Park D.-H, Kamp J.C, Brod T, Harrigfeld B, Schupp J.C, Hoeper M.M, (et al.), 2023, Effects of sotatercept on lung diffusion capacity and blood gases in patients with pulmonary arterial hypertension, European Respiratory Journal. 62(2) (no pagination), , (no pagination)	A4 andere Endpunkte

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/ Studienergebnisdatenbanken)

Listen Sie nachfolgend die durch die Studienregistersuche(n)/ Studienergebnisdatenbanksuche(n) identifizierten, aber ausgeschlossenen Registereinträge 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

Liste der ausgeschlossenen Registereinträge

Register	Trefferzahl	Ausgeschlossene Registereinträge	Eingeschlossene Registereinträge
CT.gov	25	24 (Nr. 1 – 24)	1
EU-CTR	10	9 (Nr. 25 – 33)	1
ICTRP	32	30 (Nr. 34 – 63)	2
Summe	67	63	4

Nr.	Studien-ID	Titel	Ausschlussgrund
ClinicalTrials.gov (CT.gov)			
1.	NCT00709540	A Randomized, Double-Blind, Placebo-Controlled, Multiple-Dose, Dose-Escalation Study to Evaluate the Safety, Tolerability and Pharmacodynamics of ACE-011 (ActRIIA-IgG1) in Healthy Postmenopausal Women. ClinicalTrials.gov. 8-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT00709540	A1 andere Population
2.	NCT00747123	A Phase 2a, Multi-Center, Randomized, Multiple-Dose Study to Evaluate the Safety, Tolerability and Efficacy of ACE-011 (hActRIIA-IgG1) in Patients With Osteolytic Lesions of Multiple Myeloma. ClinicalTrials.gov. 9-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT00747123	A1 andere Population
3.	NCT00931606	A Phase 2, Double-blind, Randomized, Placebo-Controlled Study of ACE-011 for the Treatment of Chemotherapy Induced Anemia in Patients With Metastatic Breast Cancer. ClinicalTrials.gov. 6-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT00931606	A1 andere Population
4.	NCT01146574	A Phase 2a, Multi-center, Randomized, Single Dose, Double-blind, Placebo-controlled Followed by a Multiple-dose, Single-blind, Placebo-controlled, Dose Escalation Study to Evaluate the Pharmacokinetics, Safety, Efficacy,	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
		Tolerability, and Pharmacodynamics of Sotatercept (ACE-011) for the Correction of Anemia in Subjects With End-stage Renal Disease (ESRD) on Hemodialysis (HD).. ClinicalTrials.gov. 6-30. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01146574	
5.	NCT01190644	A Phase 2, Open-Label, Pharmacodynamic Study to Evaluate the Effect of Sotatercept (ACE-011) on Red Blood Cell Mass and Plasma Volume in Subjects With Solid Tumors. ClinicalTrials.gov. 6-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01190644	A1 andere Population
6.	NCT01284348	An Open-Label Randomized, Phase 2A, Dose-Ranging Study of Sotatercept (ACE-011) for Chemotherapy-Induced Anemia in Subjects With Advanced or Metastatic Solid Tumors Treated With Platinum-Based Chemotherapeutic Regimens. ClinicalTrials.gov. 3-25. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01284348	A1 andere Population
7.	NCT01464164	Phase I/II, Open-Label Study to Determine Safety and Efficacy of Sotatercept (ACE-011) in Adults With Red Blood Cell Transfusion- Dependent Diamond Blackfan Anemia. ClinicalTrials.gov. 2-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01464164	A1 andere Population
8.	NCT01562405	Phase 1, Multicenter, Open-label, Dose-escalation Study of Sotatercept (ACE-011) in Combination With Lenalidomide or Pomalidomide and Dexamethasone in Patients With Relapsed and/or Refractory Multiple Myeloma. ClinicalTrials.gov. 2-05. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01562405	A1 andere Population
9.	NCT01571635	A Phase 2A, Open-label Dose Finding Study to Determine the Safety and Tolerability of Sotatercept (ACE-011) in Adults With BETA(b)-THALASSEMIA.. ClinicalTrials.gov. 0-10. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01571635	A1 andere Population
10.	NCT01712308	A Phase-2, Prospective, Open-Label Study to Determine the Safety and Efficacy of Sotatercept (ACE-011) in Subjects With Myeloproliferative Neoplasm (MPN) - Associated Myelofibrosis and Anemia. ClinicalTrials.gov. 2-21. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01712308	A1 andere Population
11.	NCT01736683	An Open-label, Randomized, Phase 2, Parallel, Dose-Ranging, Multicenter Study of Sotatercept for the Treatment of Patients With Anemia and Low or Intermediate-1 Risk Myelodysplastic Syndromes or Non-	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
		proliferative Chronic Myelomonocytic Leukemia (CMML). ClinicalTrials.gov. 1-28. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01736683	
12.	NCT01999582	A Phase 2 Multicenter, Randomized, Open-Label , Multiple-Dose Study of Intravenous and Subcutaneous Administration of Sotatercept (ACE-011) in Subjects With End-Stage Kidney Disease on Hemodialysis Switched From Erythropoiesis Stimulating Agents With Staggered Dose Group Escalation in Part 1 Followed by a Parallel Group, Active Controlled Study of Selected Dose(s) and Regimen(s) in Part 2: To Evaluate the Pharmacokinetics, Safety, Tolerability, Efficacy, Dosing Regimen, and Pharmacodynamics of Sotatercept. ClinicalTrials.gov. 1-30. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT01999582	A1 andere Population
13.	NCT02230917	A Phase IIa Study of Sotatercept on Bone Mass and Turnover in Patients With Multiple Myeloma. ClinicalTrials.gov. 4-10. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT02230917	A1 andere Population
14.	NCT03496207	A Phase 2, Double-Blind, Placebo-Controlled, Randomized Study to Compare the Efficacy and Safety of Sotatercept (ACE-011) Versus Placebo When Added to Standard of Care for the Treatment of Pulmonary Arterial Hypertension (PAH). ClinicalTrials.gov. 6-13. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT03496207	A2 andere Intervention
15.	NCT03724227	Expanded Access for ACE-011. ClinicalTrials.gov. 7-17. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT03724227	A1 andere Population
16.	NCT03738150	A Phase 2a Single-Arm, Open-Label, Multicenter Exploratory Study to Assess the Effects of Sotatercept (ACE-011) for the Treatment of Pulmonary Arterial Hypertension. ClinicalTrials.gov. 4-19. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT03738150	A5 anderer Studientyp
17.	NCT04796337	An Open-label Long-term Follow-up Study to Evaluate the Effects of Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH. ClinicalTrials.gov. 5-12. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT04796337	A7 anderer Publikationstyp
18.	NCT04811092	A Phase 3, Randomized, Double-blind, Placebo-controlled Study to Evaluate Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy in Newly	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
		Diagnosed Intermediate- and High-risk PAH Patients. ClinicalTrials.gov. 3-18. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT04811092	
19.	NCT04896008	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate Sotatercept When Added to Maximum Tolerated Background Therapy in Participants With Pulmonary Arterial Hypertension (PAH) World Health Organization (WHO) Functional Class (FC) III or FC IV at High Risk Mortality. ClinicalTrials.gov. 2-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT04896008	A1 andere Population
20.	NCT04945460	A Phase 2, Double-blind, Randomized, Placebo-controlled Study to Evaluate the Effects of Sotatercept Versus Placebo for the Treatment of Combined Postcapillary and Precapillary Pulmonary Hypertension (Cpc-PH) Due to Heart Failure With Preserved Ejection Fraction (HFpEF). ClinicalTrials.gov. 2-29. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT04945460	A1 andere Population
21.	NCT05587712	A Phase 2 Open-label Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of Sotatercept (MK-7962) in Children From 1 to Less Than 18 Years of Age With PAH on Standard of Care. ClinicalTrials.gov. 1-19. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT05587712	A1 andere Population
22.	NCT05818137	A Phase 3 Non-randomized, Non-controlled, Open Label Clinical Study to Evaluate the Efficacy and Safety of MK-7962 (Sotatercept) add-on to Background Therapy in Japanese Participants With Pulmonary Arterial Hypertension (PAH). ClinicalTrials.gov. 5-10. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT05818137	A5 anderer Studientyp
23.	NCT06351345	129Xenon MR Imaging and Spectroscopy Response to Sotatercept in Pulmonary Arterial Hypertension. ClinicalTrials.gov. 6-01. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT06351345	A5 anderer Studientyp
24.	NCT06409026	Effect of Sotatercept on Central Cardiopulmonary Performance and Peripheral Oxygen Transport During Exercise in Pulmonary Arterial Hypertension. ClinicalTrials.gov. 4-07. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrials.gov/study/NCT06409026	A5 anderer Studientyp
EU-Clinical Trials Register (EU-CTR)			
25.	2011-005659-15	A Phase 2a, open-label, dose finding study to determine the safety and tolerability of sotatercept (ACE-011) in adults with beta (β)-thalassemia. EU-CTR. 2012. [Zugriffsdatum: 09.07.2024].	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
		Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2011-005659-15	
26.	2012-002601-22	AN OPEN-LABEL, RANDOMIZED, PHASE 2, PARALLEL, DOSE-RANGING, MULTICENTER STUDY OF SOTATERCEPT FOR THE TREATMENT OF PATIENTS WITH ANEMIA AND LOW- OR INTERMEDIATE-1 RISK MYELODYSPLASTIC SYNDROMES OR NON-PROLIFERATIVE CHRONIC MYELOMONOCYTIC LEUKEMIA (CMML). EU-CTR. 2013. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2012-002601-22	A1 andere Population
27.	2012-003788-23	A Phase 2 Multicenter, Randomized, Open Label, Multiple Dose Study of Intravenous and Subcutaneous Administration of Sotatercept (ACE-011) in Subjects with End-Stage Kidney Disease on Hemodialysis Switched from Erythropoiesis Stimulating Agents with Staggered Dose Group Escalation in Part 1 Followed by a Parallel Group, Active Controlled Study of Selected Dose(s) and Regimen(s) in Part 2: to Evaluate the Pharmacokinetics, Safety, Tolerability, Efficacy, Dosing Regimen, and Pharmacodynamics of Sotatercept. EU-CTR. 2013. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2012-003788-23	A1 andere Population
28.	2017-004738-27	A Phase 2, Double-Blind, Placebo-Controlled, Randomized Study to Compare the Efficacy and Safety of Sotatercept (ACE-011) Versus Placebo When Added to Standard of Care for the Treatment of Pulmonary Arterial Hypertension (PAH). EU-CTR. 2018. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2017-004738-27	A2 andere Intervention
29.	2020-005061-13	An Open-Label Long-term Follow-up Study to Evaluate the Effects of Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH. EU-CTR. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2020-005061-13	A7 anderer Publikationstyp
30.	2021-000199-12	A Phase 3, Randomized, Double-blind, Placebo-controlled Study to Evaluate Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy in Newly Diagnosed Intermediate- and High-risk PAH Patients.. EU-CTR. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2021-000199-12	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
31.	2021-001498-21	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate Sotatercept When Added to Maximum Tolerated Background Therapy in Participants With Pulmonary Arterial Hypertension (PAH) World Health Organization (WHO) Functional Class (FC) III or FC IV at High Risk of Mortality. EU-CTR. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2021-001498-21	A1 andere Population
32.	2021-003020-32	Phase 2, Double-blind, Randomized, Placebo-controlled Study to Evaluate the Effects of Sotatercept versus Placebo-Controlled for the Treatment of Combined Postcapillary and Precapillary Pulmonary Hypertension (Cpc-PH) due to Heart Failure with Preserved Ejection Fraction (HFpEF). EU-CTR. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2021-003020-32	A1 andere Population
33.	2022-000478-25	A Phase 2 Open-label Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of Sotatercept (MK-7962) in Children from 1 to Less Than 18 Years of Age With PAH on Standard of Care. EU-CTR. 2022. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2022-000478-25	A1 andere Population
WHO International Clinical Trial Registry Platform (ICTRP)			
34.	NCT00709540	A Randomized, Double-Blind, Placebo-Controlled, Multiple-Dose, Dose-Escalation Study to Evaluate the Safety, Tolerability and Pharmacodynamics of ACE-011 (ActRIIA-IgG1) in Healthy Postmenopausal Women. ICTRP. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT00709540	A1 andere Population
35.	NCT00747123	A Phase 2a, Multi-Center, Randomized, Multiple-Dose Study to Evaluate the Safety, Tolerability and Efficacy of ACE-011 (hActRIIA-IgG1) in Patients With Osteolytic Lesions of Multiple Myeloma. ICTRP. 2020. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT00747123	A1 andere Population
36.	NCT00931606	A Phase 2, Double-blind, Randomized, Placebo-Controlled Study of ACE-011 for the Treatment of Chemotherapy Induced Anemia in Patients With Metastatic Breast Cancer. ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT00931606	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
37.	NCT01146574	A Phase 2a, Multi-center, Randomized, Single Dose, Double-blind, Placebo-controlled Followed by a Multiple-dose, Single-blind, Placebo-controlled, Dose Escalation Study to Evaluate the Pharmacokinetics, Safety, Efficacy, Tolerability, and Pharmacodynamics of Sotatercept (ACE-011) for the Correction of Anemia in Subjects With End-stage Renal Disease (ESRD) on Hemodialysis (HD).. ICTRP. 2017. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01146574	A1 andere Population
38.	NCT01190644	A Phase 2, Open-Label, Pharmacodynamic Study to Evaluate the Effect of Sotatercept (ACE-011) on Red Blood Cell Mass and Plasma Volume in Subjects With Solid Tumors. ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01190644	A1 andere Population
39.	NCT01284348	An Open-Label Randomized, Phase 2A, Dose-ranging Study of Sotatercept (ACE-011) for Chemotherapy-Induced Anemia in Subjects With Advanced or Metastatic Solid Tumors Treated With Platinum-Based Chemotherapeutic Regimens. ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01284348	A1 andere Population
40.	NCT01464164	Phase I/II, Open-Label Study to Determine Safety and Efficacy of Sotatercept (ACE-011) in Adults With Red Blood Cell Transfusion- Dependent Diamond Blackfan Anemia. ICTRP. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01464164	A1 andere Population
41.	NCT01562405	Phase 1, Multicenter, Open-label, Dose-escalation Study of Sotatercept (ACE-011) in Combination With Lenalidomide or Pomalidomide and Dexamethasone in Patients With Relapsed and/or Refractory Multiple Myeloma. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01562405	A1 andere Population
42.	NCT01571635	A Phase 2A, Open-label Dose Finding Study to Determine the Safety and Tolerability of Sotatercept (ACE-011) in Adults With BETA(b)-THALASSEMIA.. ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01571635	A1 andere Population
43.	NCT01712308	A Phase-2, Prospective, Open-Label Study to Determine the Safety and Efficacy of Sotatercept (ACE-011) in Subjects With Myeloproliferative Neoplasm (MPN) - Associated Myelofibrosis and Anemia. ICTRP. 2023.	A1 andere Population

Nr.	Studien-ID	Titel	Ausschlussgrund
		[Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01712308	
44.	NCT01736683	An Open-label, Randomized, Phase 2, Parallel, Dose-Ranging, Multicenter Study of Sotatercept for the Treatment of Patients With Anemia and Low or Intermediate-1 Risk Myelodysplastic Syndromes or Non-proliferative Chronic Myelomonocytic Leukemia (CMML). ICTRP. 2022. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01736683	A1 andere Population
45.	EUCTR2012-002601-22-FR	AN OPEN-LABEL, RANDOMIZED, PHASE 2, PARALLEL, DOSE-RANGING, MULTICENTER STUDY OF SOTATERCEPT FOR THE TREATMENT OF PATIENTS WITH ANEMIA AND LOW- OR INTERMEDIATE-1 RISK MYELODYSPLASTIC SYNDROMES OR NON-PROLIFERATIVE CHRONIC MYELOMONOCYTIC LEUKEMIA (CMML). ICTRP. 2018. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2012-002601-22-FR	A1 andere Population
46.	EUCTR2012-003788-23-BE	A Phase 2 Multicenter, Randomized, Open Label, Multiple Dose Study of Intravenous and Subcutaneous Administration of Sotatercept (ACE-011) in Subjects with End-Stage Kidney Disease on Hemodialysis Switched from Erythropoiesis Stimulating Agents with Staggered Dose Group Escalation in Part 1 Followed by a Parallel Group, Active Controlled Study of Selected Dose(s) and Regimen(s) in Part 2: to Evaluate the Pharmacokinetics, Safety, Tolerability, Efficacy, Dosing Regimen, and Pharmacodynamics of Sotatercept. ICTRP. 2016. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2012-003788-23-BE	A1 andere Population
47.	EUCTR2012-003788-23-GB	A Phase 2 Multicenter, Randomized, Open Label, Multiple Dose Study of Intravenous and Subcutaneous Administration of Sotatercept (ACE-011) in Subjects with End-Stage Kidney Disease on Hemodialysis Switched from Erythropoiesis Stimulating Agents with Staggered Dose Group Escalation in Part 1 Followed by a Parallel Group, Active Controlled Study of Selected Dose(s) and Regimen(s) in Part 2: to Evaluate the Pharmacokinetics, Safety, Tolerability, Efficacy, Dosing Regimen, and Pharmacodynamics of Sotatercept. ICTRP. 2019. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2012-003788-23-GB	A1 andere Population
48.	NCT02230917	A Phase IIa Study of Sotatercept on Bone Mass and	A1 andere

Nr.	Studien-ID	Titel	Ausschlussgrund
		Turnover in Patients With Multiple Myeloma. ICTRP. 2016. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02230917	Population
49.	NCT03496207	A Phase 2, Double-Blind, Placebo-Controlled, Randomized Study to Compare the Efficacy and Safety of Sotatercept (ACE-011) Versus Placebo When Added to Standard of Care for the Treatment of Pulmonary Arterial Hypertension (PAH). ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT03496207	A2 andere Intervention
50.	NCT03724227	Expanded Access for ACE-011. ICTRP. 2020. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT03724227	A1 andere Population
51.	NCT03738150	A Phase 2a Single-Arm, Open-Label, Multicenter Exploratory Study to Assess the Effects of Sotatercept (ACE-011) for the Treatment of Pulmonary Arterial Hypertension. ICTRP. 2022. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT03738150	A5 anderer Studientyp
52.	NCT04796337	An Open-label Long-term Follow-up Study to Evaluate the Effects of Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT04796337	A7 anderer Publikationstyp
53.	NCT04811092	A Phase 3, Randomized, Double-blind, Placebo-controlled Study to Evaluate Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy in Newly Diagnosed Intermediate- and High-risk PAH Patients. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT04811092	A1 andere Population
54.	EUCTR2021-000199-12-BE	A Phase 3, Randomized, Double-blind, Placebo-controlled Study to Evaluate Sotatercept When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy in Newly Diagnosed Intermediate- and High-risk PAH Patients. - A Phase 3 Study of Sotatercept in Newly Diagnosed Intermediate- and High-risk PAH Patients. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2021-000199-12-BE	A1 andere Population
55.	EUCTR2021-	A Phase 3, Randomized, Double-Blind, Placebo-Controlled	A1 andere

Nr.	Studien-ID	Titel	Ausschlussgrund
	001498-21-ES	Study to Evaluate Sotatercept When Added to Maximum Tolerated Background Therapy in Participants With Pulmonary Arterial Hypertension (PAH) World Health Organization (WHO) Functional Class (FC) III or FC IV at High Risk of Mortality - A Phase3 Study of Sotatercept in Participants with PAH WHO FC III or FC IV at High Risk of Mortality. ICTRP. 2021. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2021-001498-21-ES	Population
56.	NCT04945460	A Phase 2, Double-blind, Randomized, Placebo-controlled Study to Evaluate the Effects of Sotatercept Versus Placebo for the Treatment of Combined Postcapillary and Precapillary Pulmonary Hypertension (Cpc-PH) Due to Heart Failure With Preserved Ejection Fraction (HFpEF). ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT04945460	A1 andere Population
57.	EUCTR2021-003020-32-BE	Phase 2, Double-blind, Randomized, Placebo-controlled Study to Evaluate the Effects of Sotatercept versus Placebo-Controlled for the Treatment of Combined Postcapillary and Precapillary Pulmonary Hypertension (Cpc-PH) due to Heart Failure with Preserved Ejection Fraction (HFpEF) - A Phase 2 Study of Sotatercept for Cpc-PH Treatment. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2021-003020-32-BE	A1 andere Population
58.	NCT05587712	A Phase 2 Open-label Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of Sotatercept (MK-7962) in Children From 1 to Less Than 18 Years of Age With PAH on Standard of Care. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT05587712	A1 andere Population
59.	EUCTR2022-000478-25-DE	A Phase 2 Open-label Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of Sotatercept (MK-7962) in Children from 1 to Less Than 18 Years of Age With PAH on Standard of Care - Sotatercept in Children with PAH on SoC. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2022-000478-25-DE	A1 andere Population
60.	NCT05818137	A Phase 3 Non-randomized, Non-controlled, Open Label Clinical Study to Evaluate the Efficacy and Safety of MK-7962 (Sotatercept) add-on to Background Therapy in Japanese Participants With Pulmonary Arterial Hypertension (PAH). ICTRP. 2024. [Zugriffsdatum: 09.07.2024].	A5 anderer Studientyp

Nr.	Studien-ID	Titel	Ausschlussgrund
		Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT05818137	
61.	NCT06351345	129Xenon MR Imaging and Spectroscopy Response to Sotatercept in Pulmonary Arterial Hypertension. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT06351345	A5 anderer Studientyp
62.	NCT06409026	Effect of Sotatercept on Central Cardiopulmonary Performance and Peripheral Oxygen Transport During Exercise in Pulmonary Arterial Hypertension. ICTRP. 2024. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT06409026	A5 anderer Studientyp
63.	JPRN-jRCT2051220078	A Single-Administration Clinical Study to Evaluate the Safety, Tolerability and Pharmacokinetics of Sotatercept (MK-7962) in Healthy Japanese Participants. ICTRP. 2023. [Zugriffsdatum: 09.07.2024]. Verfügbar unter: https://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-jRCT2051220078	A1 andere Population

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 4.3.1.1.5 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-99 inklusive eines Flow-Charts für den Patientenfluss.

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 4-99 inklusive eines Flow-Charts für den Patientenfluss.

Tabelle 4-99 (Anhang): Studiendesign und -methodik für Studie STELLAR

Item ^a	Charakteristikum	Studieninformation
Studienziel		
2 b	Genaue Ziele, Fragestellung und Hypothesen	Bewertung der Wirksamkeit und Sicherheit von Sotatercept (inklusive PAH-Hintergrundtherapie) im Vergleich zu Placebo (inklusive PAH-Hintergrundtherapie) zu Woche 24 bei Patient:innen mit PAH
Methoden		
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Randomisierte, multizentrische, doppel-blinde, Placebo-kontrollierte, Phase-III-Studie Geeignete Patient:innen wurden im Zuteilungsverhältnis 1:1 in folgende Behandlungsgruppen randomisiert: <ul style="list-style-type: none"> • Sotatercept • Placebo
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<u>Amendment 1 vom 06.10.2021:</u> <ul style="list-style-type: none"> • Aktualisierter Text bezüglich des Nachweises von Antikörpern gegen Arzneimittel, und Text über das Risiko einer Teleangiektasie hinzugefügt • Teleangiektasie als einziges AESI hinzugefügt und Bereitstellung von Überwachungsparameter • Dosismodifikationen benötigt für unerwünschte Ereignisse von Teleangiektasie hinzugefügt • Neues Ausschlusskriterium zur Anzahl der Blutplättchen hinzugefügt • Zusätzlicher Leitfaden für 6MWD • Sekundäre und explorative Endpunkte aktualisiert • Änderung der Kriterien für den Abbruch der Studie aufgrund einer klinischen Verschlechterung und die Anweisungen für weitere Studienbesuche geändert • Bewertung von Hämatologie aus den Visiten 10-12, 14-16, 18-20, 22-24, 26-28 und 30-32 entfernt
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<u>Wichtigste Einschlusskriterien:</u> <ul style="list-style-type: none"> • Männliche und weibliche Teilnehmer:innen im Alter von ≥ 18 Jahren mit dokumentierter diagnostischer Rechtsherzkatheterisierung (RHK) und bestätigter Diagnose von PAH der WHO-Gruppe 1 im Subtyp der

Item ^a	Charakteristikum	Studieninformation
		<p>idiopathischen PAH, hereditären PAH, arzneimittel-/toxininduzierten PAH, PAH in Verbindung mit CTD oder PAH in Verbindung mit einfachen, kongenitalen systemisch-pulmonalen Shunts</p> <ul style="list-style-type: none"> • Symptomatische PAH klassifiziert als WHO-FK II oder III • Baseline RHK mit dokumentierter $PVR \geq 5$ WU und ein pulmonaler kapillärer Verschlussdruck oder linksventrikulärer diastolischer Druck von ≤ 15 mmHg beim Screening • Mindestens 90 Tage vor der Untersuchung auf stabiler PAH-Hintergrundtherapie und Diuretika • 6MWD ≥ 150 und ≤ 500 m in zwei wiederholten Messungen beim Screening <p><u>Wichtigste Ausschlusskriterien:</u></p> <ul style="list-style-type: none"> • Diagnose von PH WHO-Gruppen 2,3,4 oder 5 • Diagnose der folgenden PAH-Subtypen der Gruppe 1: HIV-assoziierte PAH und PAH in Verbindung mit portaler Hypertension. Zu den Ausschlüssen in der PAH-Gruppe 1 sollten auch gehören Schistosomiasis-assoziierte PAH und pulmonale veno-okklusive Erkrankung • Hämoglobinwert beim Screening über der geschlechtsspezifischen ULN • Thrombozytenzahl zu Baseline $< 50,000/mm^3$ ($< 50.0 * 10^9/L$) • Unkontrollierte systemische Hypertonie nachgewiesen durch systolischen Blutdruck im Sitzen > 160 mmHg oder diastolischen Blutdruck im Sitzen > 100 mmHg während des Screenings nach einer Ruhephase • Systolischer Blutdruck < 90 mmHg beim Screening
4b	Studienorganisation und Ort der Studiendurchführung	<p>Die Studie wurde an 91 Studienzentren in 21 Ländern durchgeführt.:</p> <p>Argentinien, Australien, Belgien, Brasilien, Deutschland, Frankreich, Israel, Italien, Kanada, Korea, Mexiko, Neuseeland, Niederlande, Polen, Schweden, Schweiz, Serbien, Spanien, Tschechische Republik, UK, USA</p>
5	Interventionen	<p><u>Interventions-Arm:</u> Initialdosis 0,3 mg/kg Körpergewicht als subkutane Injektion, ab Woche 3 0,7 mg/kg Körpergewicht als</p>

Item ^a	Charakteristikum	Studieninformation
	Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<p>subkutane Injektion alle 21 Tage + PAH-Hintergrundtherapie</p> <p><u>Kontroll-Arm:</u> Subkutane Placebo-Injektion alle 21 Tage + PAH-Hintergrundtherapie</p> <p><u>PAH-Hintergrundtherapie:</u> Mono- oder Kombinationstherapie bestehend aus Endothelin-Rezeptor-Antagonisten, Phosphodiesterase-Typ-5-Inhibitoren, Stimulator der löslichen Guanylatzyklase, Prostazyklin-Analoga, Selektive Prostazyklin-Rezeptor-Agonisten</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><u>Primäre Endpunkte:</u></p> <ul style="list-style-type: none"> • 6MWD (Veränderung gegenüber Baseline zu Woche 24) <p><u>Sekundäre Endpunkte:</u></p> <ul style="list-style-type: none"> • Mehrkomponentenverbesserung • Pulmonary Vascular Resistance (PVR) • NT-proBNP • WHO-Funktionsklasse • Zeit bis zur klinischen Verschlechterung oder Tod (TTCW) • French Risk Score • Symptom- und Impact-Items des PAH-SYMPACT <p><u>Explorative patientenrelevante Endpunkte:</u></p> <ul style="list-style-type: none"> • Dyspnoe gemäß Borg CR10-Skala • EQ-5D VAS
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Siehe Item 3b
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<p>Die Bestimmung der Fallzahl erfolgte anhand des primären Endpunktes (Veränderung der 6MWD zu Woche 24) und des sekundären Endpunktes (Verbesserung in WHO-FK). Die Annahmen bezüglich des Behandlungseffekts und die Variabilität (6MWD-Endpunkt) basieren auf der PULSAR-Studie und einer publizierten klinischen Studie an PAH-Patienten.</p> <p>Für die 6MWD ergibt sich, unter den Annahmen einer 1:1-Randomisierung, einer 2-seitigen 0,05 Typ-I-Fehlerrate, einer 25 m-Verbesserung der Sotatercept-</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Behandlung im Vergleich zu Placebo, einer gemeinsamen Standardabweichung von 50 m und 121 Patient:innen pro Arm, eine statistische Power von etwa 96 % unter Verwendung des Wilcoxon-Rangsummentest von N-Query.</p> <p>Für den sekundären Endpunkt Anteil der Patient:innen, mit Verbesserung der WHO-FK zu Woche 24 wird angenommen, dass der Anteil der mit Placebo und Sotatercept behandelten Teilnehmer:innen 0,11 bzw. 0,25 beträgt. Die statistische Power auf der Grundlage des 2-Stichproben-Chi-Quadrat-Tests beträgt etwa 80 %.</p> <p>Unter der Annahme einer Abbrecherquote von 15 % beträgt die Fallzahl n = 284 (n = 142 Teilnehmer:innen pro Arm).</p>
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Keine Zwischenanalyse geplant
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die Zuteilung zur Randomisierung erfolgte über ein computergestütztes System durch IRT.
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Die Randomisierung erfolgte stratifiziert nach Baseline WHO-Funktionsklasse (Klasse II vs. III) und PAH-Hintergrundtherapie (Mono-/Zweifachtherapie vs. Dreifachtherapie).
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>	Die Zuteilung zu den Behandlungsgruppen/ Randomisierung erfolgte zentral über IRT im Zuteilungsverhältnis 1:1 in die Behandlungsarme.
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?	Der Sponsor generierte die randomisierte Zuteilungssequenz für die Zuordnung der Behandlung.
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?	<p>a) Ja</p> <p>b) Ja</p> <p>c) Ja</p>

Item ^a	Charakteristikum	Studieninformation
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Nicht zutreffend.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p>Der primäre Endpunkt, die Veränderung der 6MWD zu Woche 24 im Vergleich zu Baseline wurde anhand eines stratifizierten Wilcoxon-Tests mit WHO-Funktionsklasse (Klasse II vs. Klasse III) und PAH-Hintergrundtherapie (Mono/Doppel- vs. Dreifachtherapie) als Strata analysiert. Als Schätzer für den medianen Unterschied zwischen den Behandlungsarmen wird der Hodges-Lehman Location Shift inklusive zugehörigem 95 %-KI mit zweiseitigem p-Wert dargestellt.</p> <p>Die sekundären Endpunkte Mehrkomponentenverbesserung, Verbesserung der WHO-FK und French Risk Score wurden mittels der Stratum-adjustierten Mantel-Haenszel-Methode ausgewertet, mit den Stratifizierungsfaktoren WHO-Funktionsklasse und PAH-Hintergrundtherapie.</p> <p>Die sekundären Endpunkte PVR, NT-proBNP und Symptom- und Impact-Items des PAH-SYMPACT wurden mittels stratifizierten Wilcoxon-Tests und stratifizierter ANCOVA analysiert.</p> <p>Der sekundäre Endpunkt TTCW wurde mittels Ereigniszeitanalysen ausgewertet. Die Berechnung des Hazard Ratio (HR) inklusive zugehörigem 95 %-KI und zweiseitigem p-Wert (stratifizierter Log-Rank-Test) basiert auf der Anwendung eines Cox-Regressionsmodells mit Behandlung als Kovariate, stratifiziert nach WHO-Funktionsklasse und PAH-Hintergrundtherapie.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p>Subgruppenanalysen wurden für den primären Endpunkt und folgende sekundäre Endpunkte dargestellt: PVR, NT-proBNP wenn für ein Merkmal der Stichprobenumfang in jeder Subgruppe ≥ 10 ist und auf den Analysen für die Gesamtpopulation beruht.</p> <ul style="list-style-type: none"> • Geschlecht (weiblich vs. männlich) • PAH-Ätiologische Untergruppen (idiopathische PAH vs. hereditäre PAH vs. Medikamenten-/Toxin-induzierte PAH vs. PAH assoziiert mit Bindegewebskrankheit vs. PAH assoziiert mit angeborener Herzkrankheit mit Shunt-Wiederherstellung) • PAH-Hintergrundtherapie (Mono vs. Zweifach- vs. Dreifachtherapie) • Prostazyklin-Infusion bei Studienbeginn (ja vs. nein) • Baseline WHO-Funktionsklasse (II vs. III)

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> Baseline PVR ($\leq 800 \text{ dynes} \cdot \text{sec} / \text{cm}^5$ vs. $> 800 \text{ dynes} \cdot \text{sec} / \text{cm}^5$)
Resultate		
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
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	<u>Sotatercept:</u> a) 163 b) 163 c) 163 <u>Placebo:</u> a) 160 b) 160 a) 160
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Siehe Flow-Chart
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Studienstart: 25.01.2021
14b	Informationen, warum die Studie endete oder beendet wurde	Studienende: 06.12.2023
<p>a: nach CONSORT 2010. 6MWD: 6-minutes-walking-distance (6-Minuten-Gehstrecke); AESI: Adverse Event of Special Interest (Unerwünschtes Ereignis von besonderem Interesse); ANCOVA: Analysis of Covariance (Kovarianzanalyse); CONSORT: Consolidated Standards of Reporting Trials; CTD: Common Technical Document (Zulassungsdossier); EQ-5D: EuroQoL-5 Dimensions; HIV: Humanes Immundefizienz-Virus; HR: Hazard Ratio; IRT: Interactive Response Technology; KI: Konfidenzintervall; kg: Kilogramm; L: Liter; mg: Milligramm; NT-proBNP: N-Terminal Pro-B-Type Natriuretic Peptide; PAH: Pulmonale arterielle Hypertonie; PAH-SYMPACT: Pulmonary Arterial Hypertension – Symptoms and Impact; PVR: Pulmonary Vascular Resistance; RHK: Rechtsherzkatheterisierung; TTCW: Time to Clinical Worsening (Zeit bis zur klinischen Verschlechterung oder Tod); ULN: Upper Limit of Normal (obere Normgrenze); VAS: Visual Analog Scale (visuelle Analogskala); WHO: World Health Organization (Weltgesundheitsorganisation); WU: Wood Unit (Wood-Einheit)</p>		

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.

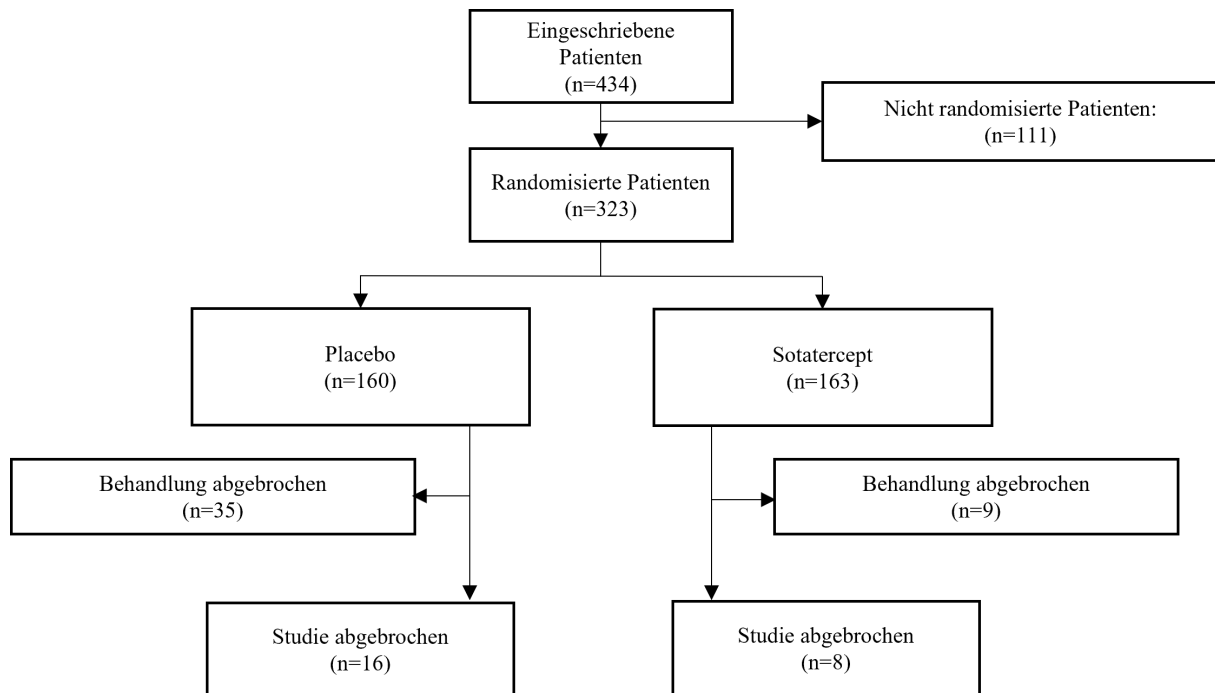


Abbildung 10: Flow-Chart der Studie STELLAR

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 4-100 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie STELLAR

Studie: STELLAR

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Klinischer Studienbericht A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Compare the Efficacy and Safety of Sotatercept Versus Placebo When Added to Background Pulmonary Arterial Hypertension (PAH) Therapy for the Treatment of PAH	CSR
Statistischer Analyseplan	SAP
Klinisches Studienprotokoll	CSP

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:

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** **nein**

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 unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und behandelnden Personen

Patient:

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Es handelt sich um eine doppelblinde Studie.

behandelnde bzw. weiterbehandelnde Personen:

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Es handelt sich um eine doppelblinde Studie.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

ja unklar nein

Angaben 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:

Die Erzeugung der Randomisierungssequenz wurde in der Studie STELLAR adäquat durchgeführt, die Gruppenzuteilung fand verdeckt statt. In dieser doppelblinden Studie waren Patient:innen und behandelnde Personen verblindet. Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige, das Verzerrungspotenzial beeinflussende Faktoren wurden nicht identifiziert. Somit wird das Verzerrungspotenzial für die Studie STELLAR auf Studienebene als niedrig eingestuft.

B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

Endpunkt: Gesamtmortalität

1. Verblindung der Endpunkterheber

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß Intention-To-Treat (ITT)-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Mortalität zu Woche 24 für die Studie STELLAR ist somit als niedrig einzustufen.

Endpunkt: Verbesserung der 6MWD**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Verbesserung der 6MWD in der Studie STELLAR ist somit als niedrig einzustufen.

Endpunkt: Zeit bis zur klinischen Verschlechterung oder Tod**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Zeit bis zur klinischen Verschlechterung oder Tod für die Studie STELLAR ist somit als niedrig einzustufen.

Endpunkt: Verbesserung der Dyspnoe gemäß Borg CR10-Skala**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Verbesserung der Dyspnoe gemäß Borg CR10-Skala für die Studie STELLAR ist somit als niedrig einzustufen.

Endpunkt: Mehrkomponentenverbesserung**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Mehrkomponentenverbesserung für die Studie STELLAR ist somit als niedrig einzustufen.

Endpunkt: Krankheitssymptomatik und Gesundheitszustand**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Krankheitssymptomatik und Gesundheitszustand für die Studie STELLAR ist somit als niedrig einzustufen

Endpunkt: Gesundheitsbezogene Lebensqualität**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der FAS-Population, also gemäß ITT-Prinzip. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren, liegen nicht vor und es gibt keine Hinweise auf eine ergebnis-gesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Gesundheitsbezogene Lebensqualität für die Studie STELLAR ist somit als niedrig einzustufen

Endpunkt: Nebenwirkungen**1. Verblindung der Endpunkterheber** **ja** **unklar** **nein**Angaben zum Kriterium; obligate Begründung für die Einstufung:**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 Erhebung des Endpunkts erfolgte verblindet. Die Auswertung erfolgte auf Basis der Safety-Analysis-Set Population. Diese umfasst alle Patient:innen, die mindestens eine Dosis der Studienmedikation erhalten haben. Sonstige, das Verzerrungspotenzial beeinflussende Faktoren liegen nicht vor und es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. Das Verzerrungspotenzial für den Endpunkt Nebenwirkungen für die Studie STELLAR 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: endpunktübergreifend; 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 statistisch nicht 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.
 - 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:

Anhang 4-G: Weitere Ergebnisse

Siehe separates Dokument.